

**Black hole and de Sitter microstructures from a semiclassical perspective**Chitraang Murdia<sup>1,2</sup>, Yasunori Nomura<sup>1,2,3</sup> and Kyle Ritchie<sup>1,2</sup><sup>1</sup>*Berkeley Center for Theoretical Physics, Department of Physics, University of California, Berkeley, California 94720, USA*<sup>2</sup>*Theoretical Physics Group, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*<sup>3</sup>*Kavli Institute for the Physics and Mathematics of the Universe (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*

(Received 22 July 2022; accepted 7 December 2022; published 20 January 2023)

We describe two different, but equivalent semiclassical views of black hole physics in which the equivalence principle and unitarity are both accommodated. In one, unitarity is built-in, while the black hole interior emerges only effectively as a collective phenomenon involving horizon (and possibly other) degrees of freedom. In the other, more widely studied approach, the existence of the interior is manifest, while the unitarity of the underlying dynamics can be captured only indirectly by incorporating certain nonperturbative effects of gravity. These two pictures correspond to a distant description and the description based on entanglement islands/replica wormholes, respectively. We also present a holographic description of de Sitter spacetime based on the former approach, in which the holographic theory is located on the stretched horizon of a static patch. We argue that the existence of these two approaches is rooted in the two formulations of quantum mechanics: the canonical and path integral formalisms.

DOI: [10.1103/PhysRevD.107.026016](https://doi.org/10.1103/PhysRevD.107.026016)**I. INTRODUCTION AND SUMMARY**

For a long time, the puzzle of black hole information loss has caused confusion about how gravity works at the quantum level [1–3]. This confusion arises mostly due to the following features of the semiclassical description of black holes:

- (i) Despite the fact that spacetime is fundamentally quantum mechanical, it is described as a classical object.
- (ii) While the fundamental theory has a preferred class of time foliations for spacetimes with a horizon, general relativity seems to treat all the coordinates equally.

The purpose of this paper is to elucidate these points and present a coherent picture in which the results of semiclassical theory are consistently interpreted to address issues related to the information puzzle. This picture builds on recent developments of our understanding of quantum gravity.

The first hint that spacetime consists of quantum degrees of freedom came from the discovery that black holes have entropy [4,5]. In the standard statistical mechanical

interpretation, this implies that the horizon, a region which general relativity describes as empty space, contains a quantum mechanical substance. This interpretation is indeed supported by the anti-de Sitter (AdS)/conformal field theory (CFT) correspondence [6]—a concrete realization of holography [6–10].

Recent progress in understanding the AdS/CFT correspondence, and holography more generally, has shown that while general relativity is diffeomorphism invariant, there appear to be a preferred class of coordinates at the quantum level. From the viewpoint of the boundary theory, these coordinates correspond to descriptions based on quantum operators constructed by a simple procedure or operators which are not exponentially complex in fundamental degrees of freedom [11–16], and they cover only a portion of spacetime when there is a horizon. In a system with a black hole, these coordinates are associated with Schwarzschild time slicing, or any other time foliation associated with an observer located outside the horizon. This conforms to the earlier idea that black hole evolution obeys the standard rules of quantum mechanics—and hence is unitary—when viewed by an external observer [17–20]. Indeed, we see that there is nothing unusual with the results of semiclassical theory if a black hole is described using only external frames.

The issue, then, is how to interpret the existence of the black hole interior which arises when the external coordinates are analytically extended in general relativity. We take the view that the picture of the black hole interior arises only effectively at the semiclassical level. In particular, we adopt the construction in Refs. [21–25] in which the interior

---

*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 SCOAP<sup>3</sup>.*

emerges because of the special, chaotic nature of the horizon dynamics.<sup>1</sup> A key idea is that when the black hole is described in an external quasistatic reference frame, its large acceleration with respect to the free falling frame makes the dynamics at the horizon—more precisely the stretched horizon [19]—appear string theoretic, which is chaotic across all low energy species. This makes the black hole vacuum microstate generic in the relevant microcanonical ensemble, allowing us to erect the effective theory of the interior.

The erected effective theory is defined only up to errors of order  $e^{-(S_{\text{BH}}+S_{\text{rad}})/2}$ , where  $S_{\text{BH}}$  and  $S_{\text{rad}}$  are the Bekenstein-Hawking entropy and the coarse-grained entropy of the degrees of freedom entangled with the black hole, respectively. Operators describing the interior are state dependent [26,31,32], though only weakly in the sense of Refs. [25,30]. Reflecting the fact that the Hilbert space associated with the black hole system is finite dimensional, the effective theory can be used only for a finite time interval; the existence of the black hole singularity is consistent with this [21]. Similarly, weak cosmic censorship can be viewed as a statement that a distant description, which corresponds to a simple boundary description in holography, can be used for an arbitrarily long time, to the extent that the theory is well defined in the infrared.

In this paper, we review the construction described above and refine it, including subtle evolutionary effects for a dynamically formed black hole. We present the whole framework in a coherent manner and address various issues associated with it, including its realization in the boundary description of holography. In doing so, we also discuss the relation of the present construction to other recent related works. In particular, we discuss the relation between the picture described here and that based on quantum extremal surfaces and entanglement wedge reconstruction [33–37]. We argue that the latter emerges through coarse graining necessary to describe a semiclassical black hole without specifying its microscopic structure [24]. We also see that the two are closely tied, respectively, to the canonical and path integral formulations of quantum mechanics [38].

We expect that at the macroscopic level, generic (quasi-)static horizons, including cosmic horizons, have locally the same statistical features. Indeed, the horizon of de Sitter spacetime has the same entropy per area as that of a black hole [39], and so is the temperature of the Hawking cloud at the stretched horizon. In fact, building on the analyses in Refs. [22,40], we see that the holographic description of a static patch of de Sitter spacetime is very much an “inside-out” version of that of a black hole. We present this description in the context of more general holography for cosmological spacetimes [41,42]. We also discuss the

relationship of the present description with other recent proposals [43–48].

## A. Overall picture and outline of the paper

In the rest of this section, we present an overview of the picture presented in this paper, pointing to where the details of each subject are covered. The assumptions about the setup which we adopt throughout the paper will be summarized at the end of this section.

### 1. No puzzle for a black hole when viewed from the exterior

Consider a nonrotating, noncharged black hole in a 4-dimensional asymptotically flat spacetime. At the classical level, the black hole is uniquely specified by one continuous parameter: its mass  $M$ . We view this system from a distance, i.e., we describe it using Schwarzschild time slicing or something related to it in a simple manner.

When quantum effects are included, the black hole has a finite entropy

$$S_{\text{BH}}(M) = 4\pi \frac{G_{\text{N}} M^2}{\hbar c}, \quad (1.1)$$

where  $G_{\text{N}}$  is Newton’s constant. This means that at the quantum level the black hole is characterized by discrete—albeit exponentially dense—states, rather than a continuous, classical number.<sup>2</sup> Specifically, the number of independent states in the energy interval between  $M - \delta M/2$  and  $M + \delta M/2$  is given by

$$\mathcal{N}(M) \sim e^{S_{\text{BH}}(M)} \frac{\delta M}{M}. \quad (1.2)$$

Note that this is a general phenomenon in quantum mechanics. Like other physical systems, the entropy in Eq. (1.1) diverges in the limit  $\hbar \rightarrow 0$ .

The semiclassical theory treats the black hole as a classical object while retaining the finite nature of the entropy. This is not inconsistent: the concept of entropy can be defined at the level of thermodynamics without explicitly taking into account the fundamental discreteness. The disadvantage of this treatment, however, is that one can no longer resolve each microstate, hence requiring a statistical, or thermodynamic, treatment of the system [4,5,49,50].

At the semiclassical level, a black hole is described as having a definite mass  $M$  but with the Hawking cloud around it, which is in a thermal mixed state of temperature

$$T_{\text{H}} = \frac{\hbar c^3}{8\pi k G_{\text{N}} M}, \quad (1.3)$$

<sup>1</sup>A detailed description of relations of this construction to earlier work [26–30] is given in Ref. [25] and throughout this paper.

<sup>2</sup>Of course, quantum mechanics allows for a superposition of these independent states, so that the expectation value of the energy, or  $M$ , can take continuous values.

where  $k$  is the Boltzmann constant. This is a proxy for an ensemble of black hole microstates with energies between  $M - \delta M/2$  and  $M + \delta M/2$ , which dominates the corresponding microcanonical ensemble of states associated with the spacetime region near the black hole. Below, we adopt natural units  $\hbar = c = k = 1$ .

The formation and evaporation of a black hole described in an external frame is a process in which quantum information in the initial collapsing matter is dispersed among spacetime degrees of freedom, represented by the Bekenstein-Hawking entropy; Hawking emission then transfers it back to matter degrees of freedom in the semiclassical description. For an external observer, this entire process is unitary if all the microscopic degrees of freedom are accounted for. An important point is that with quantum effects, the instantaneously defined apparent horizon is stretched, at which the local (Tolman) Hawking temperature is the string scale [19].<sup>3</sup> The trajectory of this stretched horizon is timelike, so an object falling into a black hole reaches there within finite time. This object is then absorbed by the black hole, whose information will eventually be sent back to ambient space as Hawking radiation.

This implies that in the external frame description, the stretched horizon behaves as a regular material surface as far as the flow of quantum information is concerned. The semiclassical theory, however, treats the horizon degrees of freedom appearing in the middle of the process to be classical, or at best an ensemble of quantum states represented by the Hawking cloud. Therefore, in any semiclassical calculation, the quantum information in the initial matter *must* appear to be lost in the final state; there is no way to completely describe the microscopic information on the horizon while staying in the semiclassical regime. In other words, Hawking's calculation [5] must have led to information loss, and indeed it did [1].

## 2. Emergence of the interior

A puzzling feature of the picture described above is that general relativity seems to imply that the black hole horizon is smooth. Namely, when an object freely falls into the horizon, it does not experience anything special there. This is obviously not the case when an object falls into a regular material surface.

As discussed in Refs. [21–25], the stretched horizon is distinguished from other, regular material surfaces by its chaotic [51], fast-scrambling [52,53] dynamics across *all* low energy species. Here, by all low energy species, we mean all quantum fields appearing in the low energy effective theory below the string scale. These features arise

<sup>3</sup>By the string scale, we mean the scale at which the low energy effective field theory description breaks down. The appearance of this scale is associated with the nonzero Newton's constant, which controls quantum corrections to the system since it always appears with  $\hbar$ .

because of the large relative acceleration, of order the string scale, between the quasistatic frame (a natural frame in holography) and the free falling frame at the stretched horizon. It is this aspect of the horizon dynamics that allows us to erect a description in which an object falls freely through the stretched horizon. This is done by making the state take a fully generic form in the relevant microcanonical ensemble.<sup>4</sup>

Specifically, when described in a quasistatic reference frame, quantum degrees of freedom of a black hole consist of modes in a spatial region near the stretched horizon (called the zone) as well as those at the stretched horizon. We call them zone and horizon modes, respectively. While the dynamics of the former is described by the low energy effective field theory, that of the latter is not. Note that these modes are defined for each time interval in which the black hole can be viewed as quasistatic.

Let us now focus on a (small) subset of the zone modes which is relevant for describing an infalling physical object. In particular, the object will be described by excitations of these modes over the black hole vacuum. Let us call these modes hard modes and all other black hole (zone and horizon) modes soft modes. From the assumption that the horizon degrees of freedom obey chaotic dynamics, we can then conclude that a black hole vacuum microstate takes the form

$$|\Psi(M)\rangle = \sum_n e^{S_{\text{bh}}(M-E_n)} \sum_{i_n=1} c_{ni_n} |\{n_\alpha\}\rangle |\psi_{i_n}^{(n)}\rangle, \quad (1.4)$$

where  $|\{n_\alpha\}\rangle$  represent states of the hard modes, specified by the occupation numbers  $n_\alpha$  for each mode  $\alpha$ , and  $|\psi_{i_n}^{(n)}\rangle$  are the states of the soft modes that have energy  $M - E_n$ , where  $E_n$  is the energy carried by  $|\{n_\alpha\}\rangle$ . Note that since the total energy of the black hole system is constrained to be  $M$ , the number of independent horizon-mode states  $|\psi_{i_n}^{(n)}\rangle$  coupling to  $|\{n_\alpha\}\rangle$  in Eq. (1.4) is given by  $S_{\text{bh}}(M - E_n)$ , the density of states at energy  $M - E_n$ . Here,

$$S_{\text{bh}}(E) = 4\pi E^2 l_p^2 \quad (1.5)$$

is the Bekenstein-Hawking entropy density at energy  $E$  with  $l_p$  being the Planck length. (The contribution from the hard modes to the black hole entropy is negligible.) Also note that because of the chaotic nature of the horizon dynamics, the coefficients  $c_{ni_n}$  take random values across all low energy species. In fact, the quasistatic form of Eq. (1.4) is achieved quickly after any disturbance, i.e., within the scrambling timescale of order  $M l_p^2 \ln(M l_p)$  [52,53].

<sup>4</sup>For a charged or rotating black hole, the relevant ensemble consists of microstates with charge or angular momentum constrained to lie within a small window dictated by quantum uncertainties.

The universality of the form of the state in Eq. (1.4) allows us to erect the effective theory of the interior. Specifically, we can define the normalized state coupling to  $|\{n_\alpha\}\rangle$  in Eq. (1.4)

$$|\{\!|n_\alpha\}\!\rangle = \varsigma_n \sum_{i_n=1}^{e^{S_{\text{bh}}(M-E_n)}} c_{ni_n} |\psi_{i_n}^{(n)}\rangle, \quad (1.6)$$

where

$$\varsigma_n = \frac{1}{\sqrt{\sum_{i_n=1}^{e^{S_{\text{bh}}(M-E_n)}} |c_{ni_n}|^2}} = e^{\frac{E_n}{2T_{\text{H}}}} \sqrt{\sum_m e^{-\frac{E_m}{T_{\text{H}}}} [1 + O(e^{-\frac{1}{2}S_{\text{bh}}(M)})]}. \quad (1.7)$$

Plugging this into Eq. (1.4), we obtain the standard thermofield double form [54,55]

$$|\Psi(M)\rangle = \frac{1}{\sqrt{\sum_m e^{-\frac{E_m}{T_{\text{H}}}}} \sum_n e^{-\frac{E_n}{2T_{\text{H}}}} |\{n_\alpha\}\rangle |\{\!|n_\alpha\}\!\rangle, \quad (1.8)$$

up to exponentially small corrections of order  $e^{-S_{\text{bh}}/2}$ . We can thus evolve an object in the zone (generated by acting creation/annihilation operators on  $|\{n_\alpha\}\rangle$ 's) using the time evolution operator associated with the proper time of the falling object, which is different from the original time evolution operator in the boundary theory but still local in bulk spacetime [21–25]. This implies the existence of a description in which the falling object passes the horizon and enters the black hole interior smoothly.

We emphasize that the criterion for a state to take the universal form in Eq. (1.4) is much stronger than that for regular thermalization; in particular, it must exhibit universal thermalization across all low energy species. It is reasonable to expect that such strong thermalization is achieved within a reasonable timescale only by the string scale dynamics, which singles out the stretched horizon. We also stress that the prescription of obtaining interior spacetime described here does not require a detailed knowledge about microscopic dynamics of quantum gravity; only some basic assumptions are sufficient.

As the evaporation of a black hole progresses, information stored in the horizon degrees of freedom is gradually transferred to Hawking radiation. After the Page time [20], entanglement between hard modes and early Hawking radiation becomes nonnegligible, so that the interior description must involve this Hawking radiation [29]. The way this occurs is that the black hole interior, more precisely  $|\{\!|n_\alpha\}\!\rangle$ 's in Eq. (1.8), must involve early Hawking radiation *in addition to* the black hole degrees of freedom.

This last statement may seem to contradict the recent claim about entanglement wedge reconstruction that the interior of an old black hole can be reconstructed using only

early Hawking radiation [33–35]. This is, however, not the case [24]. A key point is that such entanglement wedge reconstruction makes crucial use of boundary time evolution, which allows for reconstructing states of the black hole and radiation modes at an earlier time. On the other hand, the effective theory of the interior uses these modes directly at the time relevant for describing an infalling object. This understanding resolves an apparent puzzle associated with causality in entanglement wedge reconstruction.

In this paper, we first perform a detailed analysis of the near horizon mode structure in Secs. II and III, and study its relation to the holographic boundary picture in Sec. IV. The construction of the effective theory of the interior is presented in Secs. III and V, where we also discuss the refinement of the construction needed to accommodate the effects of evaporation on the instantaneous state of the zone modes. The relationship of this picture to entanglement wedge reconstruction is discussed in Sec. V.

### 3. de Sitter holography

The similarity between the static patch description of de Sitter spacetime and the external description of black hole spacetime has been studied for a long time [39]. Based on this similarity, it was suggested that de Sitter spacetime may admit a holographic description with finite-dimensional Hilbert space [56,57]. In this paper, building on the earlier analysis in Refs. [22,40], we develop a holographic description of de Sitter spacetime based on the (quasi)static picture.

First, we note that the analogue of a collapse formed, single-sided black hole is cosmological de Sitter spacetime, which arises approximately in the middle of a cosmological history, e.g., at late times in a bubble universe with positive cosmological constant. In this case, there is a single static patch as viewed by an observer (timelike geodesic), which is the inside-out analog of the exterior of the black hole. We define the zone and horizon modes as those inside (on the observer side of) the stretched horizon and those on the stretched horizon, respectively. As in the case of a black hole, a microstate for the de Sitter vacuum then takes the form

$$|\Psi(E)\rangle = \sum_n e^{S_{\text{dS}}(E-E_n)} \sum_{i_n=1} c_{ni_n} |\{n_\alpha\}\rangle |\psi_{i_n}^{(n)}\rangle, \quad (1.9)$$

where  $|\{n_\alpha\}\rangle$  represent states of the hard modes and  $|\psi_{i_n}^{(n)}\rangle$  are the states of the soft modes that have energy  $E - E_n$ , where  $E_n$  is the energy of  $|\{n_\alpha\}\rangle$  measured at the location of the observer. Here,

$$S_{\text{dS}}(E) = \pi E^2 l_p^2 \quad (1.10)$$

represents the de Sitter entropy, with  $E$  being the ‘‘energy’’ of the de Sitter vacuum related to the Hubble radius  $\alpha$  by



$$E = \frac{\alpha}{l_p}. \quad (1.11)$$

Again, the coefficients  $c_{n_i}$  take random values across all low energy species because of universally chaotic dynamics at the horizon. This randomization is achieved rather quickly, within the scrambling timescale of order  $\alpha \ln(\alpha/l_p)$  [58].

Using a microstate in Eq. (1.9), the effective theory describing a region outside the horizon (as viewed from the observer) can be constructed analogously to the black hole case. By identifying the normalized state  $|\{\{n_\alpha\}\}\rangle$  using Eq. (1.9), we can rewrite the full state as

$$|\Psi(E)\rangle = \frac{1}{\sqrt{\sum_m e^{-\frac{E_m}{T_H}}} \sum_n e^{-\frac{E_n}{2T_H}} |\{n_\alpha\}\rangle |\{\{n_\alpha\}\}\rangle, \quad (1.12)$$

where  $T_H$  is the de Sitter temperature. This state represents the semiclassical vacuum state of global de Sitter spacetime at time when the effective theory is erected.

We thus find that the effective global de Sitter picture emerges from a cosmological de Sitter spacetime as a collective phenomenon involving horizon modes. Like the black hole case, the effective theory of global de Sitter spacetime is intrinsically semiclassical in that there is an intrinsic ambiguity of order  $e^{-S_{\text{GH}}/2}$  in the definition of the theory, where  $S_{\text{GH}}$  is the Gibbons-Hawking entropy. This, therefore, addresses the issue identified as a puzzling feature in Ref. [59] that symmetries of classical (global) de Sitter spacetime cannot be implemented exactly in a finite-dimensional Hilbert space.

The holographic theory of de Sitter spacetime described here is presented in Secs. III and V. Its relation to holography in more general spacetimes is discussed in Sec. IV, where we will see how static patch de Sitter holography arises naturally from holography in more general cosmological spacetimes. In Sec. V, we will describe the construction of the effective theory and see that this theory is sufficient to provide a semiclassical description of future measurements of the observer.

#### 4. Intrinsically two-sided systems

While we find that the analytically extended Schwarzschild and global de Sitter spacetimes emerge from the more physical, single-sided spacetimes via collective dynamics involving horizon modes, there is nothing theoretically wrong with considering an ‘‘intrinsically two-sided’’ system, which comprises two copies of the holographic theory for the single-sided system.

In Sec. VI, we analyze such two-sided systems using our framework. We will see that these two-sided systems lead to physics similar to the corresponding single-sided systems at the semiclassical level, despite the fact that the microscopic structures of relevant states are significantly

different. We also comment on possible relations of our framework to other proposals for holographic theories developed in the context of intrinsically two-sided systems. In particular, we discuss the relationship of our description of de Sitter spacetime with the DS/dS correspondence [43,44] and the Shaghoulian-Susskind proposal [45–48].

#### 5. Gravitational path integral: Ensemble from coarse graining

The approach described so far is formulated most naturally in the canonical formalism of quantum mechanics. For example, we have introduced the concept of horizon modes and their associated states. However, quantum mechanics must also be formulable using path integral. What does the picture look like in this case?

The path integral formalism has a very different starting point. In the context of quantum gravity, it is the collection of classical field configurations on classical geometries, which are then integrated over to obtain physical results. This implies, however, that a black hole (or de Sitter) spacetime appearing in the path integral must be treated as a (semi)classical object, so its detailed microscopic structure cannot be discriminated. In fact, since the energy differences between different black hole microstates are suppressed by  $e^{-S_{\text{BH}}(M)}$ , these states cannot be discriminated by direct measurement.

From the microscopic point of view, this means that a black hole appearing in gravitational path integral is a coarse-grained object. This mandatory coarse graining introduces the concept of ensemble averaging in interpreting the result of gravitational path integral. This interpretation, in fact, reproduces many features which are attributed to the ensemble nature of holographic theories in lower dimensional quantum gravity [36] using an ensemble of microscopic states in a single theory [24]. However, if the classical black hole spacetime necessarily represents an ensemble of microstates, how can the underlying unitarity of a black hole evolution be captured by the quantum extremal surface method [33–35] which uses such classical spacetime?.

A key observation is that while gravitational path integral can only calculate ensemble averages, it can still do so for many different quantities. In particular, adopting the replica method, the gravitational path integral can calculate the traces of powers of the density matrix of Hawking radiation  $\overline{\text{Tr}\rho_R^n}$ . This is the replica wormhole calculation of Refs. [36,37]. From this, the ensemble average of the von Neumann entropy of the radiation can be obtained

$$\overline{S_R} = -\lim_{n \rightarrow 1} \frac{\partial}{\partial n} \overline{\text{Tr}\rho_R^n} \sim \min\{S_{\text{rad}}, S_{\text{bh}}\}, \quad (1.13)$$

which follows the Page curve. The reason why we could obtain the Page curve here is because we have calculated

the ensemble average of the microscopic von Neumann entropy, which obeys the Page curve for all members of the ensemble. This is unlike Hawking’s calculation which gives the von Neumann entropy of the averaged state.

This replica method prescription involving gravitational path integrals is virtually equivalent to the entanglement island prescription for calculating entropies, adopted in Refs. [33–35]. This, therefore, provides the following interpretation of the results in Refs. [33–35]: while any formalism treating a black hole as a classical object, including the gravitational path integral, must involve an ensemble average over microstates, microscopic information about some quantities can still be deduced by computing such ensemble averages. The entanglement island prescription (implicitly) adopts this for von Neumann entropies.

The issue described here is discussed in Sec. VII. It is based on the picture outlined in Refs. [24,38]. The idea that the semiclassical description involves an ensemble of microstates was also discussed in Refs. [60–64], and the understanding of the Page curve presented there is based on the developments in Refs. [65–71].

### 6. *Quantum mechanics vs general relativity in quantum gravity*

It is important that the extremization procedure involved in the calculation of an entanglement island is performed on global spacetime of general relativity. In particular, for a black hole spacetime, it must be performed on the whole spacetime including the interior of the black hole.

This picture, therefore, is complementary to that based on the external view of the black hole. Here the existence of the black hole interior is manifest, while understanding the unitarity of black hole evolution requires a method incorporating nonperturbative effects of quantum gravity, such as replica wormholes. On the other hand, in the framework based on the external view, the unitarity of the evolution is built-in, and the interior emerges only effectively as a collective phenomenon involving horizon (and possibly other, entangled) degrees of freedom.

Despite the fact that the two pictures appear very different, they give the same physical conclusions. In particular, a black hole evolves unitarily and has a smooth horizon. The origin of historical confusions about black hole physics come from the fact that only one of these features is manifest in a given low energy description; the other appears in a highly nontrivial manner. It is interesting that the description in which unitarity of quantum mechanics is manifest naturally comes with the canonical/Hamiltonian formulation of quantum mechanics, while the one in which the interior predicted by general relativity is manifest is naturally associated with the path integral/Lagrangian formulation. It is an interesting question if a microscopic formulation of quantum gravity can make both these features manifest.

## B. List of assumptions

Here we list out the assumptions that we use throughout the paper. We focus on spacetimes which are spherically symmetric at the semiclassical level, although excitations on them are not restricted to be spherically symmetric. We take the number of spacetime dimensions to be  $3 + 1$ , although our arguments can be generalized straightforwardly to  $d + 1$  dimensions with  $d \geq 3$ .

When discussing black holes, we mostly consider a spherically symmetric, noncharged black hole in an asymptotically flat spacetime (or an analogous object such as a spherically symmetric, noncharged small black hole in an asymptotically AdS spacetime). It is straightforward to include the effect of a charge or rotation unless the black hole is extremal or near extremal. An extension to (near) extremal and lower dimensional black holes is expected to be nontrivial, since these black holes have different structures for the densities of states than “generic” black holes considered in this paper; see, e.g., Refs. [72,73].

## II. QUANTUM FIELD ON BACKGROUND SPACETIME

In this section, we discuss the behavior of quantum fields in spherically symmetric background spacetimes. The calculation presented here is elementary. The results, however, are used in later sections, so we include them for completeness.

We consider a 4-dimensional spacetime with the metric

$$ds^2 = -f(r)dt^2 + \frac{1}{f(r)}dr^2 + r^2d\Omega^2, \quad (2.1)$$

where  $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ . For simplicity, we consider a minimally coupled real scalar field  $\Phi$  in this spacetime, whose action is given by

$$I = \frac{1}{2} \int d^4x \sqrt{-g} (-g^{\mu\nu} \partial_\mu \Phi \partial_\nu \Phi - m_\Phi^2 \Phi^2). \quad (2.2)$$

After changing the radial coordinate to the tortoise coordinate  $r_*$  defined by

$$dr_* = \frac{dr}{f(r)}, \quad (2.3)$$

the metric becomes

$$ds^2 = f(r)(-dt^2 + dr_*^2) + r^2d\Omega^2. \quad (2.4)$$

Here,  $r$  should be regarded as a function of  $r_*$ . The equation of motion derived from Eq. (2.2) is then given by

$$\left(-\partial_t^2 + \partial_{r_*}^2 - \frac{f(r)f'(r)}{r} + \frac{f(r)}{r^2}\partial_\Omega^2 - m_\Phi^2 f(r)\right)(r\Phi) = 0, \quad (2.5)$$

where  $\partial_\Omega^2$  is defined by  $\partial_\Omega^2 \chi = (1/\sin\theta)\partial_\theta(\sin\theta\partial_\theta\chi) + (1/\sin^2\theta)\partial_\phi^2\chi$ .

We look for positive frequency solutions of the form

$$\Phi(t, r_*, \theta, \phi) = e^{-i\omega t} \frac{\varphi_{\ell m}(r_*)}{r} Y_{\ell m}(\theta, \phi) \quad (2.6)$$

with  $\omega \geq 0$ , where  $Y_{\ell m}(\theta, \phi)$  represent real spherical harmonics, satisfying  $\partial_\Omega^2 Y_{\ell m} = -\ell(\ell+1)Y_{\ell m}$ . This results in a linear, second-order differential equation for  $\varphi_{\ell m}(r)$

$$-\frac{d^2}{(dr_*)^2} \varphi_{\ell m}(r_*) + \{V_\ell(r_*) - \omega^2\} \varphi_{\ell m}(r_*) = 0, \quad (2.7)$$

where the effective potential  $V_\ell(r_*)$  is given by

$$V_\ell(r) = f(r) \left( \frac{f'(r)}{r} + \frac{\ell(\ell+1)}{r^2} + m_\Phi^2 \right). \quad (2.8)$$

The corresponding equations for higher spin fields are generally more complicated, but they can also be derived using a semiclassical method [74–77].

We are interested in spacetime that has a horizon at radius  $r_+$  determined by

$$f(r_+) = 0. \quad (2.9)$$

The location,  $r_s$ , of the stretched horizon [19] is determined by the condition that the proper distance between  $r = r_+$  and  $r_s$  is the string length  $l_s$ :

$$\left| \int_{r_+}^{r_s} \frac{dr}{\sqrt{f(r)}} \right| \approx l_s, \quad (2.10)$$

giving

$$|r_s - r_+| \approx \frac{|f'(r_+)| l_s^2}{4}. \quad (2.11)$$

Here, we have assumed that  $|f'(r_+)|$  is not much suppressed compared with the natural size determined by dimensional analysis,  $|f'(r_+)| \sim 1/r_+$ , which is indeed the case for the spacetimes that we consider in this paper. The Hawking temperature, as measured at  $r$  satisfying  $f(r) = 1$ , is given by

$$T_H = \frac{|f'(r_+)|}{4\pi}, \quad (2.12)$$

so that the location of the stretched horizon also coincides with the place where the local (Tolman) Hawking

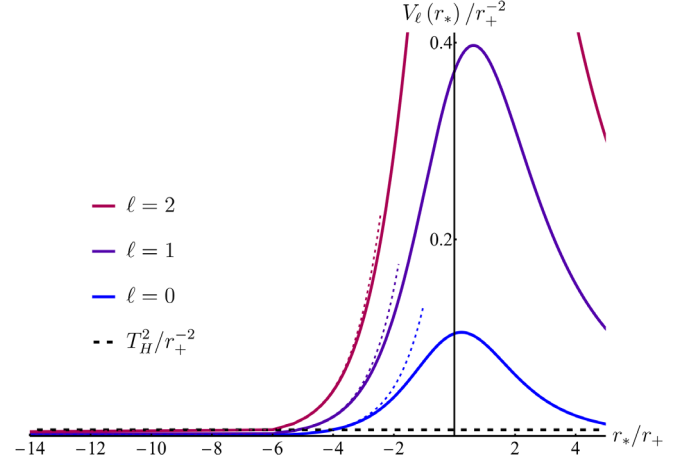


FIG. 1. The effective potential  $V_\ell$  of a Schwarzschild black hole (solid lines) for a minimally coupled massless scalar field with  $\ell = 0, 1, 2$  as a function of the tortoise coordinate  $r_*$ . The Hawking temperature  $T_H = 1/4\pi r_+$  is indicated by the horizontal dashed line. The approximated curves in Eq. (2.20) are also plotted (dotted lines) in the region where they are relevant.

temperature  $T_{\text{loc}}(r) = T_H/\sqrt{f(r)}$  becomes the string scale,  $\approx 1/2\pi l_s$  [24].

### A. Schwarzschild black hole

Let us consider a Schwarzschild black hole of mass  $M$ . The metric is given by

$$f(r) = 1 - \frac{r_+}{r}, \quad (2.13)$$

where  $r_+ = 2Ml_p^2$ . The tortoise coordinate is given by

$$r_* = r + r_+ \ln \frac{r - r_+}{r_+}, \quad (2.14)$$

which maps  $r: (r_+, \infty)$  to  $r_*: (-\infty, \infty)$ , and the effective potential is

$$V_\ell(r_*) = \left(1 - \frac{r_+}{r}\right) \left(\frac{r_+}{r^3} + \frac{\ell(\ell+1)}{r^2} + m_\Phi^2\right), \quad (2.15)$$

which is plotted in Fig. 1 for  $m_\Phi = 0$ .<sup>5</sup> The stretched horizon is located at

<sup>5</sup>For a general, not necessarily a scalar, field of  $m_\Phi = 0$ , we have

$$V_\ell(r_*) = \left(1 - \frac{r_+}{r}\right) \left(\frac{(1-s^2)r_+}{r^3} + \frac{\ell(\ell+1)}{r^2}\right), \quad (2.16)$$

where  $s$  is the spin-weight parameter and  $\ell \geq |s|$ . In the near horizon limit, this leads to the approximate potential of Eq. (2.20) but with  $\lambda_\ell = (\ell^2 + \ell + 1 - s^2)/r_+^2$ , giving a higher potential barrier for larger  $|s|$  for a fixed value of  $\ell - |s|$ .

$$r_s - r_+ \approx \frac{l_s^2}{4r_+} \iff r_{*s} - r_+ \approx -2r_+ \ln \frac{2r_+}{l_s}, \quad (2.17)$$

and the Hawking temperature is given by

$$T_H = \frac{1}{4\pi r_+}. \quad (2.18)$$

While we cannot solve Eq. (2.7) analytically, we can study the behavior of the field in the near horizon and far regions by making appropriate approximations. We are mostly interested in the near horizon “zone” region

$$r_s < r < r_z, \quad r_z \approx \frac{3}{2}r_+, \quad (2.19)$$

where  $r_z$  indicates the location of the potential barrier. The near horizon limit corresponds to  $r - r_+ \ll O(r_+)$ , where  $r_* \approx r_+ + r_+ \ln[(r - r_+)/r_+] \iff r \approx r_+ + r_+ e^{(r_* - r_+)/r_+}$ . The effective potential in this region is given by

$$V_\ell(r_*) \approx \lambda_\ell e^{\frac{r_* - r_+}{r_+}}, \quad \lambda_\ell = \frac{\ell^2 + \ell + 1}{r_+^2} + m_\Phi^2. \quad (2.20)$$

This approximation is valid if  $r_*$  is negative with  $|r_*|$  sufficiently larger than  $r_+$ ; see Fig. 1. This implies that we can trust solutions obtained using Eq. (2.20) if  $\omega^2 \ll V_\ell(r_* = -r_+) \sim \lambda_\ell$ .

Two independent real solutions of Eq. (2.7) with Eq. (2.20) can be taken as

$$\text{Re}[I_{2ir_+\omega}(2\sqrt{\lambda_\ell}r_+e^{\frac{r_* - r_+}{2r_+}})], \quad \text{Im}[I_{2ir_+\omega}(2\sqrt{\lambda_\ell}r_+e^{\frac{r_* - r_+}{2r_+}})], \quad (2.21)$$

where  $I_\nu(x)$  is the modified Bessel function of the first kind. The first solution is exponentially increasing in  $r_*$  at large  $r_*$ , so it does not correspond to modes that are localized in the zone region; it corresponds to decaying modes for signals sent from the far region. We thus focus on the second solution, which is exponentially damped at large  $r_*$ . This solution is approximated by a trigonometric function in the near horizon region

$$\begin{aligned} \varphi_{\ell m}(r_*) &\propto \text{Im}[I_{2ir_+\omega}(2\sqrt{\lambda_\ell}r_+e^{\frac{r_* - r_+}{2r_+}})] \\ &\xrightarrow[r_* < 0]{|r_*| \gg r_+} \sin[\omega r_* + \omega r_+(\ln(\lambda_\ell r_+^2) - 1)]. \end{aligned} \quad (2.22)$$

Given that the effective potential at the stretched horizon has the value  $V_\ell(r_{*s}) = \lambda_\ell l_s^2/4r_+^2$ , we find

$$\omega \gtrsim \frac{\sqrt{\lambda_\ell} l_s}{2r_+}. \quad (2.23)$$

We also find that imposing a boundary condition at the stretched horizon makes the spectrum discrete, with the gap between adjacent levels given by

$$\Delta\omega \equiv \omega_n - \omega_{n-1} \approx \frac{\pi}{2r_+ |\ln \frac{2}{\sqrt{\lambda_\ell} l_s}|}, \quad (2.24)$$

where we have imposed a simple Dirichlet boundary condition  $\varphi_{\ell m}(r_{*s}) = 0$  for illustrative purposes. Note that for a given  $\ell$ , each level has  $(2\ell + 1)$ -fold degeneracy corresponding to  $m = -\ell, -\ell + 1, \dots, \ell$ . While the details of the spectrum do depend on the boundary condition, its basic structure—the discreteness and the scale characterizing the gaps—does not.

## B. de Sitter spacetime

A similar analysis can be performed for de Sitter spacetime in static coordinates

$$f(r) = 1 - \frac{r^2}{r_+^2}, \quad (2.25)$$

leading to the same basic conclusion. Here,  $r_+ = \alpha$  with  $\alpha$  being the Hubble radius. The tortoise coordinate is given by

$$r_* = \frac{r_+}{2} \ln \frac{1 + \frac{r}{r_+}}{1 - \frac{r}{r_+}}, \quad (2.26)$$

which maps  $r: (0, r_+)$  to  $r_*: (0, \infty)$ , and the effective potential is

$$V_\ell(r_*) = \left(1 - \frac{r^2}{r_+^2}\right) \left(-\frac{2}{r_+^2} + \frac{\ell(\ell + 1)}{r^2} + m_\Phi^2\right), \quad (2.27)$$

which is plotted in Fig. 2 for  $m_\Phi = 0$ . The stretched horizon is located at

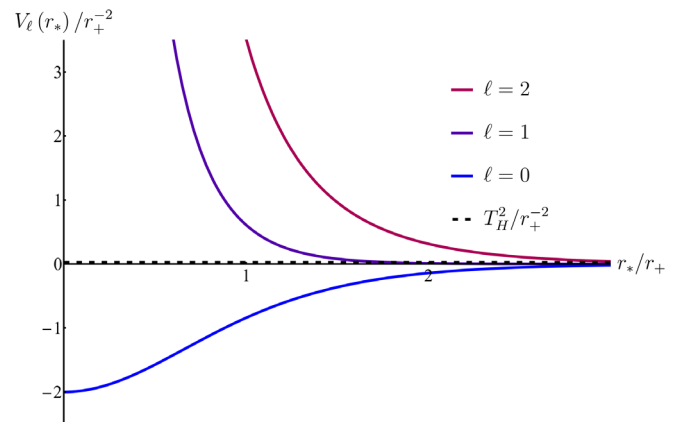


FIG. 2. The effective potential  $V_\ell$  of de Sitter spacetime for a minimally coupled massless scalar field with  $\ell = 0, 1, 2$  as a function of the tortoise coordinate  $r_*$ . The Hawking temperature  $T_H = 1/2\pi r_+$  is indicated by the horizontal dashed line.



$$r_+ - r_s \approx \frac{l_s^2}{2r_+} \iff r_{*s} \approx r_+ \ln \frac{2r_+}{l_s}, \quad (2.28)$$

and the Hawking temperature is given by

$$T_H = \frac{1}{2\pi r_+}. \quad (2.29)$$

For  $\ell \neq 0$ , the effective potential blows up at small values of  $r_*$ . Therefore, if  $\omega$  is small, we can restrict to  $r_* > r_+$  which corresponds to the region near the cosmological horizon. For  $\ell = 0$ , the effective potential blows up at small  $r_*$  only when  $m_\Phi > \sqrt{2}/r_+$ ; otherwise, the potential is confining, leading to a small number of bound states. Below, we focus on the case  $m_\Phi > \sqrt{2}/r_+$  for  $\ell = 0$ .

With this restriction, the effective potential in the near horizon region is given by

$$V_\ell(r_*) \approx 4\lambda_\ell e^{-2\frac{r_*}{r_+}}, \quad \lambda_\ell = \frac{\ell^2 + \ell - 2}{r_+^2} + m_\Phi^2, \quad (2.30)$$

except for  $\ell = 1$  with  $m_\Phi = 0$ , in which case

$$V_{\ell=1}(r_*) \approx \frac{32}{r_+^2} e^{-4\frac{r_*}{r_+}} \quad \text{for } m_\Phi = 0. \quad (2.31)$$

This approximation is valid only for  $r_* \gtrsim r_+$ . We can thus trust the solutions of Eq. (2.7) obtained using these  $V_\ell(r_*)$ 's only if  $\omega^2 \ll V(r_* = r_+) \sim \lambda_\ell$  (or  $\ll 1/r_+^2$  for  $\ell = 1, m_\Phi = 0$ ).

The most general solution corresponding to the approximate potential in Eq. (2.30) is given by

$$\varphi_{\ell m}(r_*) = \text{Re}[A I_{ir_+\omega}(2\sqrt{\lambda_\ell} r_+ e^{-\frac{r_*}{r_+}})], \quad (2.32)$$

where  $A \in \mathbb{C}$  is an arbitrary constant. In order for the original field  $\Phi(t, r_*, \theta, \phi)$  in Eq. (2.6) to be well defined, the exact solution of  $\varphi_{\ell m}(r_*)$  must vanish at least as fast as  $r_*$  when  $r_* \approx r \rightarrow 0$ . This condition fixes the phase of  $A$  in the approximate solution, leading to a single solution for  $\varphi_{\ell m}(r_*)$  in the near horizon region:

$$\varphi_{\ell m}(r_*) \xrightarrow{r_* \gg r_+} \sin \left[ \omega r_* - \frac{1}{2} \omega r_+ \ln(\lambda_\ell r_+^2) + \delta \right], \quad (2.33)$$

where  $\delta$  is determined by the phase of  $A$ . For  $\ell = 1$  and  $m_\Phi = 0$ , a similar analysis gives

$$\begin{aligned} \varphi_{\ell m}(r_*) &\propto \text{Re} \left[ A I_{\frac{1}{2}r_+\omega}(2\sqrt{2}e^{-\frac{r_*}{r_+}}) \right] \\ &\xrightarrow{r_* \gg r_+} \sin \left[ \omega r_* - \frac{\ln 2}{4} \omega r_+ + \delta \right]. \end{aligned} \quad (2.34)$$

Given that the effective potential in Eq. (2.30) has the value  $V_\ell(r_{*s}) = \lambda_\ell l_s^2 / r_+^2$ , we find

$$\omega \gtrsim \frac{\sqrt{\lambda_\ell} l_s}{r_+} \quad \text{for } (\ell, m_\Phi) \neq (1, 0), \quad (2.35)$$

and by imposing the boundary condition  $\varphi_{\ell m}(r_{*s}) = 0$ , we obtain a discrete spectrum with

$$\Delta\omega \approx \frac{\pi}{r_+ \left| \ln \frac{2}{\sqrt{\lambda_\ell} l_s} \right|} \quad \text{for } (\ell, m_\Phi) \neq (1, 0). \quad (2.36)$$

For  $(\ell, m_\Phi) = (1, 0)$ , the corresponding quantities are given instead by  $V_\ell(r_{*s}) = 2l_s^4 / r_+^6$  and

$$\omega \gtrsim \frac{\sqrt{2} l_s^2}{r_+^3}, \quad \Delta\omega \approx \frac{\pi}{r_+ \ln \frac{2^{3/4} r_+}{l_s}} \quad \text{for } (\ell, m_\Phi) = (1, 0). \quad (2.37)$$

Once again, the basic structure of the spectrum found here does not depend on the boundary condition at the stretched horizon.

### C. General near horizon limit

We now see that the structure of the spectrum for

$$\omega \ll \sqrt{\frac{\ell^2}{r_+^2} + m_\Phi^2}, \quad \ell \gg 1 \quad (2.38)$$

in the near horizon region is universal for generic horizons, beyond the black hole and de Sitter spacetimes discussed so far. This reflects the universality of the near horizon limit, giving Rindler spacetime.

To see this, let us consider a spacetime with the metric in Eq. (2.1) which has a horizon at  $r = r_+$ . We assume that  $f(r) > 1$  in the region  $r > r_+$ , which we call the allowed region. For a given horizon, this can always be arranged. Specifically, if  $f(r) > 1$  in  $r < r_+$ , as in de Sitter spacetime, we redefine  $r \rightarrow -r$  (and  $r_+ \rightarrow -r_+$ ) to make the allowed region  $r > r_+$ . Note that this makes  $r$  and  $r_+$  negative.

In the near horizon region, we then have  $f(r) \approx f'(r_+)(r - r_+)$  with  $f'(r_+) > 0$ . Here, we have assumed that  $f'(r_+)$  is not too suppressed compared with its natural size  $f'(r_+) \sim 1/|r_+|$ . (This excludes the horizon of a near extremal black hole from our consideration.) The tortoise coordinate in the near horizon region is then

$$r_* \approx \frac{1}{f'(r_+)} \ln \frac{r - r_+}{c|r_+|}, \quad (2.39)$$

where  $c > 0$  is an unimportant  $O(1)$  number defining the origin of  $r_*$ ,<sup>6</sup> and the effective potential in this region is given by

$$V_\ell(r_*) \approx c|r_+|f'(r_+)\lambda_\ell e^{f'(r_+)r_*} \approx \lambda_\ell e^{f'(r_+)r_*},$$

$$\lambda_\ell = \frac{f'(r_+)}{r_+} + \frac{\ell(\ell+1)}{r_+^2} + m_\Phi^2. \quad (2.40)$$

We want  $\lambda_\ell > 0$  for this potential to trap modes in the near horizon region. (Here we exclude the nongeneric case of  $\lambda_\ell = 0$  from consideration.) If  $r_+ > 0$ , this condition is satisfied for any values of  $\ell$  and  $m_\Phi$ . If  $r_+ < 0$ , we restrict our treatment to modes with large enough  $\ell$  or  $m_\Phi$  such that  $\lambda_\ell > 0$ .

For the above approximation to hold, we need  $\omega^2 \ll V(r_* \sim -1/f'(r_+)) \sim \lambda_\ell$ , which for  $\ell \gg 1$  gives the condition in Eq. (2.38). Assuming that  $\omega$  is in this range, we can obtain the general solution to Eq. (2.7) as

$$\varphi_{\ell m}(r_*) = \text{Re} \left[ AI \frac{2i\omega}{f'(r_+)} \left( 2\sqrt{\frac{c|r_+|\lambda_\ell}{f'(r_+)}} e^{\frac{f'(r_+)r_*}{2}} \right) \right], \quad (2.41)$$

where  $A \in \mathbb{C}$  is an arbitrary constant. The relevant solution is selected by a boundary condition at  $r_* \approx O(r_+)$ , which depends on spacetime under consideration. For Schwarzschild black hole and de Sitter spacetimes, these were given at  $r \approx r_z$ , and  $r = 0$ , respectively. This fixes the phase of  $A$  and gives us a single solution for each  $\omega$ . In the near horizon limit, it takes the form

$$\varphi_{\ell m}(r_*) \approx \sin \left( \omega r_* + \frac{\omega}{f'(r_+)} \ln \frac{c|r_+|\lambda_\ell}{f'(r_+)} + \delta \right)$$

for  $r_* < 0$ ,  $|r_*| \gg |r_+|$ , (2.42)

where  $\delta$  is an  $O(1)$  phase determined by the phase of  $A$ .

The existence of the stretched horizon, at  $r_{*s} \approx -(1/f'(r_+)) \ln(4c|r_+|/l_s^2 f'(r_+))$ , has two effects. First, since  $V_\ell(r_{*s}) \approx \lambda_\ell l_s^2 / |r_+|^2$ , frequency  $\omega$  for given  $\lambda_\ell$  is bounded from below:

$$\omega \gtrsim \frac{\sqrt{\lambda_\ell} l_s}{|r_+|}. \quad (2.43)$$

Second, imposing a boundary condition at  $r_* = r_{*s}$  quantizes the spectrum:

<sup>6</sup>In Secs. II A and II B,  $c$  was taken to be  $1/e$  and  $2$ , respectively.

$$\Delta\omega \sim \frac{\pi f'(r_+)}{2 \left| \ln \frac{2}{\sqrt{\lambda_\ell} l_s} \right|}. \quad (2.44)$$

We find that these are universal features of the spectrum in the regime of Eq. (2.38).

### III. THE ABSENCE OF SPACETIME BELOW THE STRING LENGTH

A key element of the analysis in the previous section is the hypothesis that spacetime, as we usually perceive, does not exist below the string length. In normal circumstances, this hardly affects low energy physics because of Wilsonian decoupling. In the presence of a horizon, however, large gravitational red/blue shift makes this fact relevant for low energy physics. Its significance can be best seen in the tortoise coordinate, in which the wavelength of a massless mode is preserved while propagating; in terms of this coordinate, the removal of the spacetime region within the string length from the horizon leads to excising a half line.

The proper distance, used in defining the stretched horizon in Eq. (2.10), is associated with the particular time foliation that leads to the static form of the metric, Eq. (2.1). In the case of an evaporating black hole, we apply our treatment to a sufficiently small time window, e.g.,  $\Delta t \lesssim r_+$ , in which the system can be viewed as approximately static. The issue of time slicing will be discussed further in Sec. IV.

In this section, we demonstrate that the Bekenstein-Hawking entropy,  $S_{\text{BH}}$ , as well as Hawking temperature,  $T_{\text{H}}$ , follow only from two inputs from the ultraviolet (UV) physics (or two assumptions from the point of view of low energy theory):

- (i) Spacetime does not exist below the string length, which introduces the “boundary” of space—the stretched horizon—at a proper distance  $l_s$  away from the horizon.
- (ii) Physics associated with the boundary is (maximally) quantum chaotic [51] across all low energy species [22], so the black hole and de Sitter vacuum microstates (or the microstates associated with any horizon) are typical in a suitable microcanonical ensemble.

In particular, these reproduce  $S_{\text{BH}}$  and  $T_{\text{H}}$  up to incalculable  $O(1)$  coefficients. In the situation where the state evolves in time at the semiclassical level, for example when the black hole is evaporating or the system is perturbed by excitations, we assume that the typicality described above is reached quickly. In fact, it is believed that the dynamics associated with horizons is fast scrambling [52,53].<sup>7</sup>

<sup>7</sup>As discussed in Ref. [22], this assumption implies the absence of fundamental global symmetries (see, e.g., Refs. [78–81]); specifically, any linearly realized global symmetry is explicitly broken by an  $O(1)$ —or not exponentially suppressed—amount at the string scale.

Next we discuss the UV sensitivities of the various modes, classifying them into modes that can be reliably described in the semiclassical theory and those that are intrinsically quantum gravitational. This provides a refinement of the concept of hard and soft modes introduced in Refs. [21,22] to describe an analytic extension of spacetime at the microscopic level; the meaning of this analytic extension in quantum gravity will be discussed in more detail in Sec. V. Finally, we will discuss how the vacuum and excited states of semiclassical theory are related to the microscopic description given here.

### A. Entropy, temperature, and microstates for a semiclassical vacuum

As stated above, we assume that the unknown UV physics of quantum gravity appears in low energy physics as a lack of space below the proper length of order  $l_s$ . As we have seen in Sec. II, this makes the spectrum discrete:

$$\omega_n^{(\ell)} = \omega_0^{(\ell)} + n\Delta\omega^{(\ell)} \quad n = 0, 1, 2, \dots, \quad (3.1)$$

where

$$\omega_0^{(\ell)} \approx \frac{\sqrt{\lambda_\ell} l_s}{|r_+|}, \quad \Delta\omega^{(\ell)} \sim \left| \frac{\pi f'(r_+)}{2 \ln \frac{2}{\sqrt{\lambda_\ell} l_s}} \right|. \quad (3.2)$$

Following Refs. [21,22], we take the view that the energy of the system, which is usually attributed to the background, is carried by the quanta that fill these energy levels.

For concreteness, let us consider a Schwarzschild black hole. The case of de Sitter spacetime can be analyzed similarly, which we will discuss at the end of this subsection. Suppose that there are  $N_n^{(\ell)}$  quanta of field  $\Phi$  at the  $n$ th level of angular momentum  $\ell$ , which has  $(2\ell + 1)$ -fold degeneracy. The total energy carried by  $\Phi$  is then

$$E_\Phi = g_\Phi \sum_{\ell=0}^{\infty} \sum_{n=0}^{\infty} \omega_n^{(\ell)} N_n^{(\ell)}, \quad (3.3)$$

where  $g_\Phi$  is the number of degrees of freedom for field  $\Phi$ . ( $g_\Phi = 1$  for a real scalar field.) The picture of Refs. [21,22] is that the sum of this energy for all low energy fields represents the total energy of the system (as measured at  $r$  satisfying  $f(r) = 1$ ), which is determined self-consistently by the spacetime background used in calculating  $\omega_n^{(\ell)}$ . In the present case

$$E_{\text{bh}} = \sum_{\Phi} E_\Phi = M, \quad (3.4)$$

where the sum runs over all low energy fields that can be viewed as elementary in the effective field theory at a scale slightly below  $1/l_s$ .<sup>8</sup>

While the expressions for  $\omega_0^{(\ell)}$  and  $\Delta\omega^{(\ell)}$  are different for a field with spin, their values are of the same order as those of a scalar field. In particular, using Eq. (2.16) in footnote 5, one finds that  $\lambda_\ell$  in Eq. (3.2) is simply replaced as

$$\lambda_\ell \rightarrow \lambda_{\ell,s} = \frac{\ell^2 + \ell + 1 - s^2}{r_+^2} \quad (\ell \geq |s|) \quad (3.5)$$

for  $m_\Phi = 0$ . This does not change the values of  $\omega_0^{(\ell)}$  and  $\Delta\omega^{(\ell)}$  much, as long as  $|s| \approx O(1)$  which we assume here. In any case, the precise numbers for these quantities are not important, or trustable, as we will discuss in Sec. III B.

Because of the assumption of chaotic and fast scrambling dynamics at the stretched horizon, especially those across all low energy species, the distribution of energy among various species and levels is determined purely by the content of low energy fields. In particular, the distribution of quanta in each degree of freedom is given by maximizing the combinatorial numbers

$$C_\Phi(\{N_n^{(\ell)}\}) = \begin{cases} \prod_{\ell=0}^{\infty} \prod_{n=0}^{\infty} \frac{(N_n^{(\ell)} + 2\ell)!}{N_n^{(\ell)}!(2\ell)!} & (N_n^{(\ell)} \geq 0) & \text{for } \Phi: \text{ boson} \\ \prod_{\ell=0}^{\infty} \prod_{n=0}^{\infty} \frac{(2\ell+1)!}{N_n^{(\ell)}!(2\ell+1-N_n^{(\ell)})!} & (0 \leq N_n^{(\ell)} \leq 2\ell+1) & \text{for } \Phi: \text{ fermion} \end{cases} \quad (3.6)$$

under the constraint of Eq. (3.4).<sup>9</sup> Following the standard analysis in statistical mechanics, we find

$$N_n^{(\ell)} = \frac{2\ell + 1}{e^{\beta\omega_n^{(\ell)}} \mp 1}, \quad (3.7)$$

<sup>8</sup>We ignore the kinetic energy, which is not important for a black hole that is spherically symmetric at the classical level.

<sup>9</sup>As in the standard statistical mechanics, this gives the most probable configuration of quanta. The probability of finding other configurations satisfying the constraint is not zero but exponentially suppressed.

where  $\mp$  takes the minus and plus signs for bosonic and fermionic degrees of freedom, respectively, and  $\beta$  is determined by the condition

$$E_{\text{bh}} = \sum_{\Phi} g_{\Phi} \sum_{\ell=0}^{\infty} \sum_{n=0}^{\infty} \frac{(2\ell+1)\omega_n^{(\ell)}}{e^{\beta\omega_n^{(\ell)}} \mp 1}. \quad (3.8)$$

Note that  $\beta$  in Eq. (3.8) does not depend on  $\Phi$ , and there is no ‘‘chemical potential’’ for any  $\Phi$ , since the dynamics at the stretched horizon is chaotic across all low energy species.

We can calculate the right-hand side of Eq. (3.8) by using Eq. (3.1) and replacing the sum over  $\ell$  and  $n$  with the corresponding integrals. Assuming that

$$\ell \gg 1, \quad \frac{\ell^2}{r_+^2} \gg m_{\Phi}^2, \quad (3.9)$$

which is justified *a posteriori* as the integral is dominated by  $\ell \sim r_+/l_s$ , we may use

$$\omega_0^{(\ell)} \approx \frac{\ell l_s}{r_+^2}, \quad \Delta\omega^{(\ell)} \sim \frac{1}{|r_+ \ln \frac{r_+}{\ell l_s}|}. \quad (3.10)$$

This gives

$$E_{\text{bh}} \approx N_{\text{dof}} \frac{r_+^5}{\beta^4 l_s^2} \left( \left| \ln \frac{\beta}{r_+} \right| + O(1) \right), \quad (3.11)$$

where  $N_{\text{dof}} = \sum_{\Phi} g_{\Phi}$  is the total number of degrees of freedom of low energy fields. Using

$$\frac{1}{l_p^2} \approx \frac{N_{\text{dof}}}{l_s^2} \quad (3.12)$$

(see, e.g., Ref. [82]) as well as  $E_{\text{bh}} = M$  and  $r_+ = 2Ml_p^2$ , we find

$$\beta \approx Ml_p^2, \quad (3.13)$$

which is the inverse Hawking temperature. We also find that

$$S = \sum_{\Phi} g_{\Phi} \ln C_{\Phi} \approx M^2 l_p^2, \quad (3.14)$$

as indicated by the Bekenstein-Hawking entropy. Note that since the contribution to

$$\ln C_{\Phi} \approx \sum_{\ell=0}^{\infty} 2\ell \sum_{n=0}^{\infty} \beta \omega_n^{(\ell)} e^{-\beta \omega_n^{(\ell)}} \quad (3.15)$$

comes predominantly from the  $\ell \approx O(r_+/l_s)$  modes, the entropy is given primarily by the number of different independent states that these large  $\ell$  modes can take.

In the present method based on a low energy description of the system, the coefficients in Eqs. (3.13) and (3.14) cannot be obtained. This is because the quantities are dominated by the contributions from the  $\ell \sim r_+/l_s$  modes that are localized near the stretched horizon, where the effect of unknown UV physics dominates. While this implies that the calculation is UV sensitive, its agreement with the results of Bekenstein and Hawking still gives us information about the UV physics; in particular, this unknown physics does not drastically increase the degrees of freedom compared to what is suggested by naively cutting off the spacetime at  $r_* \approx r_{*s}$ . Note that this calculation is equivalent to that in Refs. [21,22], in which the mass and entropy of a black hole are obtained by integrating appropriate powers of the local Hawking temperature. The logic, however, is reversed here; once the geometry is given (within a time window of order  $r_+$ ) by Eqs. (2.1) and (2.13), and correspondingly the energy of the system by  $M$ , then a typical state represents a black hole vacuum microstate with the temperature and entropy of the black hole given by Eqs. (3.13) and (3.14). Microstates of the black hole vacuum correspond to different ways in which the energy levels in Eq. (3.1) are occupied under the energy constraint.

As in the standard thermal system, the microscopic state of a black hole changes generically in a timescale of order the inverse temperature  $\beta \approx Ml_p^2$ . This implies that the energy of the system, i.e., the mass of the black hole, can be specified only up to the precision of  $\Delta E \sim 1/\beta$ ; the state of a black hole comprises a superposition of energy eigenstates with the spread of eigenvalues of order  $1/\beta$  or larger. In the following, we always assume that the state of the system of interest, e.g., a black hole or de Sitter spacetime, is specified with this maximal precision. A similar comment also applies to other quantities, such as the momentum of a black hole, whose minimal uncertainty is of order  $\Delta p \sim \Delta(\sqrt{2ME}) \sim 1/Ml_p^2$ . The number of independent states consistent with this specification is given by the Bekenstein-Hawking entropy of Eq. (3.14). If the state involves superpositions of wider ranges of energy, momentum, and so on, e.g., as a result of backreaction of Hawking emission [83,84], then our discussion below applies to each branch specified with the maximal precision for these quantities.

Note that a typical state in the space spanned by the independent microstates specified by  $E = M$  and  $p = 0$  within the minimal uncertainties of  $\Delta E$  and  $\Delta p$  has angular momentum of order the uncertainty  $\Delta J \sim \sqrt{N_{\text{dof}}} r_+/l_s \sim r_+/l_p$ . This is consistent with the maximal precision with which angles can be specified consistently with the UV cutoff:  $\Delta\theta \sim l_s/r_+$  for each species. We can thus regard these microstates as those of a nonrotating black hole in semiclassical theory.

We finally mention that we can discuss de Sitter spacetime in a similar manner. One difference is that we



do not have a well-established notion of energy attributed to the spacetime in this case. However, by taking

$$E_{\text{ds}} = \frac{\alpha}{l_{\text{p}}^2}, \quad (3.16)$$

as implied by the Bekenstein-Hawking (or Gibbons-Hawking [39]) entropy  $S_{\text{GH}} = \pi\alpha^2/l_{\text{p}}^2$  and Hawking temperature  $T_{\text{H}} = 1/2\pi\alpha$ , we reproduce all the properties associated with a static patch of the de Sitter spacetime [22]. (Physically, this energy can be specified only up to the precision of order  $T_{\text{H}}$ .) While the relationship of this energy to more conventionally defined energies is not clear, it represents some “energy” defined at  $r = 0$ , at which  $f(r) = 1$ , rather than at asymptotic infinity.<sup>10</sup>

### B. UV (in)sensitivity: Zone and horizon modes

Even though the density of states given by Eqs. (3.1) and (3.2) reproduces the entropy and temperature associated with the horizon, we do not expect that the precise spectrum is given by these expressions. This is because interactions near the stretched horizon, with the effective coupling given by  $T_{\text{loc}}(r_{\text{s}})^2 l_{\text{s}}^2 \sim O(1)$ , deform the spectrum significantly from that of free theories (though the density of states will not change when coarse grained at a scale of order  $r_+$  in the  $r_*$  coordinate).

To analyze this issue in more detail, let us consider the effective potential  $V_{\ell}(r_*)$  for a fixed  $\ell$  and negligible  $m_{\Phi}$ . From Eq. (3.2), the average gap between adjacent energy levels is given by

$$\Delta\omega^{(\ell)} \sim \frac{1}{r_+ \ln \frac{r_+}{l_{\text{s}}}}. \quad (3.17)$$

For a black hole, the height of the barrier is given by

$$\omega_{\text{barrier}}^{(\ell)} \sim \frac{\ell}{r_+}. \quad (3.18)$$

As we have seen, the temperature of the system is  $\sim 1/r_+$ , and the modes with

$$\omega \lesssim \frac{1}{r_+} \quad \text{and} \quad \ell \lesssim \frac{r_+}{l_{\text{s}}} \quad (3.19)$$

are significantly occupied. For each of these modes, the wave function is oscillatory between  $r_* = r_{*\text{s}}$  and

<sup>10</sup>It is interesting to note that if one considers the quantity  $E = \int_{\Sigma} \sqrt{-g} \rho d^3x$ , along the lines of Ref. [85], then one would get  $E = \alpha/l_{\text{p}}^2$ . Here,  $\Sigma$  is the  $t = 0$  surface of *global* de Sitter spacetime comprising two static patches,  $g = -r^4 \sin^2 \theta$  is the determinant of the *spacetime* metric (in the static coordinates),  $\rho = 3/8\pi l_{\text{p}}^2 \alpha^2$  is the energy density, and  $d^3x = dr d\theta d\phi$ .

$$r_* \approx -\frac{2}{f'(r_+)} \ln \ell \equiv r_{*\text{b}}^{(\ell)}, \quad (3.20)$$

outside of which it is exponentially damped. Thus the size of the region supporting these modes is given, in units of the wavelength of Hawking radiation, as

$$\frac{|r_{*\text{b}}^{(\ell)} - r_{*\text{s}}|}{r_+} \approx \left| \frac{1}{r_+ f'(r_+)} \ln \frac{r_+}{\ell^2 l_{\text{s}}^2 |f'(r_+)|} \right| \sim \left| \ln \frac{r_+}{\ell l_{\text{s}}} \right|, \quad (3.21)$$

where we have used  $|f'(r_+)| \sim 1/r_+$  in the last expression.

For illustration, let us first consider the two extreme cases of  $\ell \approx O(1)$  and  $O(r_+/l_{\text{s}})$ . For  $\ell \approx O(1)$ , there are  $1/\Delta\omega^{(\ell)} r_+ \sim \ln(r_+/l_{\text{s}})$  independent modes that have  $\omega \approx O(1/r_+)$  for each orbital and magnetic quantum numbers  $\ell$  and  $m$ . We may take them to be wave packets of width  $\approx r_+$  in  $r_*$  distributed uniformly between  $r_{*\text{s}}$  and  $r_{*\text{b}}^{(\ell)}$  without having a significant overlap with each other; see Fig. 3. In order for these to form a basis, we need to prepare two sets of wave packets, moving toward larger and smaller values of  $r_*$ . Among these (approximately orthogonal) wave packets, the ones closest to the stretched horizon, i.e., those located within  $\sim r_+$  in  $r_*$  from the stretched horizon, are special in that their dynamics cannot be described by a semiclassical theory. This is because the interaction strength of the unknown UV dynamics is strong there, which can also be seen from the fact that the unknown  $O(1)$  coefficient in the definition of the stretched horizon in Eq. (2.10) translates into the ambiguity of the location of the stretched horizon of order  $r_+$  in  $r_*$ . We call modes corresponding to these wave packets, i.e., the wave packets “next to” the stretched horizon, *horizon modes*. On the other hand, the dynamics of other  $\sim \ln(r_+/l_{\text{s}})$  wave packets can be described by a semiclassical theory (at least) in the relevant timescale of order  $1/T_{\text{H}} \sim r_+$ . Modes associated with these semiclassically describable wave packets are called *zone modes*.

For  $\ell \approx O(r_+/l_{\text{s}})$ , the situation is different. In this case, there are only  $1/\Delta\omega^{(\ell)} r_+ \approx O(1)$  independent modes having  $\omega \approx O(1/r_+)$  for each  $\ell$  and  $m$ , which, given Eq. (3.21), are all supported within  $\sim r_+$  from the stretched horizon. Therefore, they are all horizon modes. As stated earlier, the entropy of a black hole (or de Sitter spacetime) is dominated by the number of independent states of these high  $\ell$  modes

$$S \sim N_{\text{dof}} \sum_{\ell=0}^{O(r_+/l_{\text{s}})} \sum_{m=-\ell}^{\ell} O(1) \sim \frac{r_+^2}{l_{\text{p}}^2}, \quad (3.22)$$

so the dynamics of the black hole (de Sitter) microstates cannot be described by a semiclassical theory. Indeed, the dynamics of these modes is expected to be nonlocal in the spatial directions along the horizon [52,53].

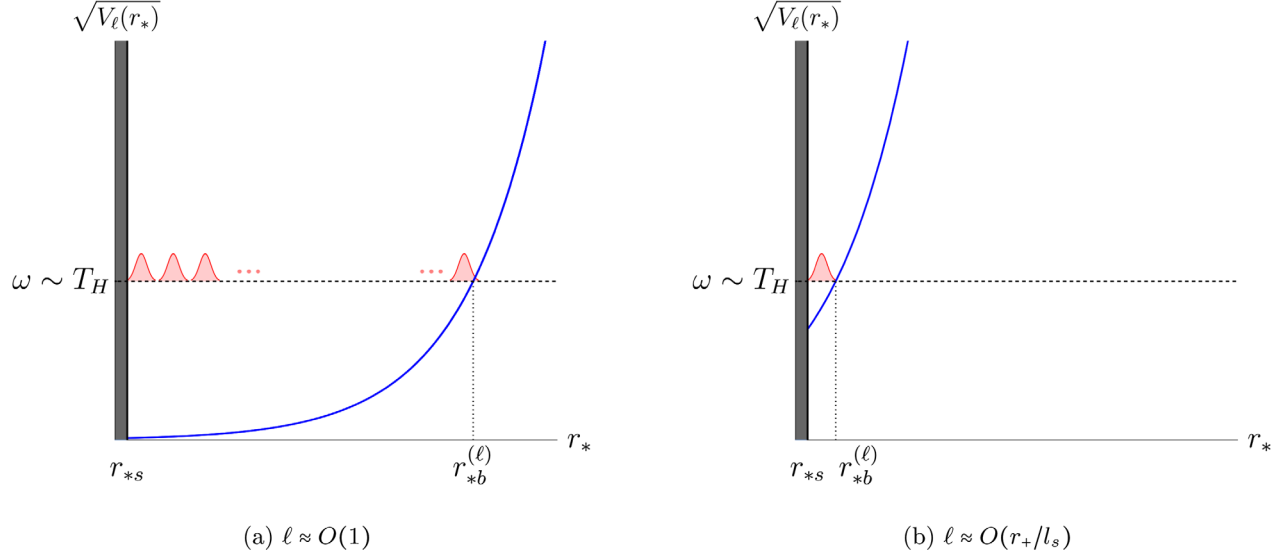


FIG. 3. A schematic depiction of approximately orthogonal, independent modes having  $\omega \approx O(T_H)$  and localized in the zone for a field of negligible mass. (a) For  $\ell \approx O(1)$ , there are  $\sim \ln(r_+/l_s)$  independent ingoing and outgoing modes. Their approximate basis can be taken to be wave packets of width  $\sim r_+$  in  $r_*$ , distributed uniformly over the classically allowed region between  $r_{*s}$  and  $r_{*b}^{(\ell)}$ , where  $r_{*b}^{(\ell)} - r_{*s} \sim r_+ \ln(r_+/l_s)$ . (b) For modes with  $\ell \approx O(r_+/l_s)$ , the classically allowed region is near the stretched horizon with  $r_{*b}^{(\ell)} - r_{*s} \sim r_+$ , so all these modes must be regarded as horizon modes.

In general, we call modes localized near the stretched horizon, i.e., within  $\sim r_+$  in  $r_*$ , horizon modes while those away from it we call zone modes. We use the term “zone modes” also in spacetimes other than the black hole spacetime, including de Sitter spacetime, even though there may not be real zones in these spacetimes. In the case where the region near the horizon is connected to another (ambient/bath) region across a barrier of the effective potential, we restrict the use of the term zone modes to those modes that are contained in the region near the horizon. For example, for a Schwarzschild black hole in asymptotically flat spacetime, zone modes only refer to modes within the zone  $r \lesssim r_z$  of the black hole. Modes located outside the barrier,  $r \gtrsim r_z$ , are not directly involved in the near horizon dynamics, and we call them *far modes*.

It is important to realize that the terminologies introduced above are associated with the spatial position of modes at a given time, or more precisely a time interval of width  $\Delta t \lesssim O(r_+)$  within which the system can be regarded as approximately static. This implies, for example, that a zone mode according to the classification at time  $t_1$  can be a horizon, zone, or far mode (or a superposition of them) according to the classification at another time  $t_2$ . In particular, if the mode is well localized in the zone at time  $t_1$  and is propagating toward the stretched horizon, then it will be a horizon mode according to the classification at  $t_2$ , a time after this mode has reached the stretched horizon. Generally, a horizon mode remains as a horizon mode for a long time, although it occasionally becomes an outgoing zone mode through interactions at the stretched horizon.

### C. Vacuum microstates of semiclassical theories

Consider a state specified by the occupation numbers of all the levels in Eq. (3.1) (with the precise values of  $\omega_n^{(\ell)}$  modified by interactions). This should be understood as a state of full quantum gravity. In particular, a typical pure state given in this way represents a microstate of a black hole or de Sitter vacuum. We now connect this description to the picture in semiclassical theory, wherein a vacuum state takes the form of a mixed, (approximately) thermal state.

For this purpose, we first consider a subset of zone modes which is relevant for describing a physical object (or objects) falling into the horizon. In the context of de Sitter spacetime, this means an object accelerating away from the observer toward the cosmological horizon. We can choose a subset of modes to accommodate *any* semiclassical object; the object is then described as an excitation of this set of modes over the semiclassical vacuum state.<sup>11</sup> We call these modes *hard modes*. They comprise only a tiny subset of the zone modes, since even the most entropic configuration of semiclassical matter has the entropy suppressed by powers of  $l_P/r_+$  compared with the entropy of the Hawking cloud [7,86], of which the entropy of the zone modes comprises an  $O(1)$  fraction. The rest of the zone modes and the horizon modes are together called *soft modes*.

<sup>11</sup>This subset can be chosen to represent bound or metastable states (or an object made out of them). We may even choose it to represent a microscopic black hole, although such a construction would have to be made more precise.

Similar to the zone and horizon modes, the hard and soft modes are defined at a specific time  $t$ . In particular, a mode defined as a hard (or soft) mode at time  $t$  may not be a hard (soft) mode at another time. This issue, however, is less relevant for these modes than for the zone and horizon modes, since the concept is used almost exclusively to construct an effective theory describing the region behind the horizon, which is erected at a given instantaneous time  $t$  as we will see below.

We label hard modes collectively by  $\alpha$ , which includes all possible quantum numbers like species, level, and orbital and spin angular momenta. The state of these modes is specified by giving the occupation number  $n_\alpha$  ( $\geq 0$ ) for each  $\alpha$ , which we denote by  $|\{n_\alpha\}\rangle$  and normalize such that

$$\langle \{m_\alpha\} | \{n_\alpha\} \rangle = \delta_{\{m_\alpha\}, \{n_\alpha\}} = \delta_{mn}, \quad (3.23)$$

where  $m$  and  $n$  are shorthand notations of  $\{m_\alpha\}$  and  $\{n_\alpha\}$ . We assume that different hard-mode states are observationally distinguishable in that two different states do not have identical quantum numbers within the uncertainties  $\Delta E$ ,  $\Delta p$ , and so on.

Suppose that the system consists only of zone and horizon modes; examples include de Sitter spacetime.<sup>12</sup> In this class of systems, the state of the entire system is generally given as an entangled state of hard and soft modes. Because of the energy constraint, the state of soft modes that comes with the hard-mode state  $|\{n_\alpha\}\rangle$  must have energy  $E - E_n$ , up to the uncertainty  $\Delta E$  required by quantum mechanics, which is typically of the order of the Hawking temperature  $T_H$ . Here,  $E$  is the total energy of the system, e.g.,  $E_{\text{dS}}$  in Eq. (3.16) for de Sitter spacetime, and

$$E_n = E_{\{n_\alpha\}} = \sum_\alpha n_\alpha \omega_\alpha \quad (3.24)$$

is the energy of the hard-mode state  $|\{n_\alpha\}\rangle$ .

The relevant Hilbert space of the system is then given by

$$\mathcal{H}(E) = \bigoplus_n (|\{n_\alpha\}\rangle \otimes \mathcal{H}_{\text{soft}}^{(n)}), \quad (3.25)$$

where  $\mathcal{H}_{\text{soft}}^{(n)}$  is the Hilbert space spanned by the soft-mode states that carry energy  $E - E_n$  within the uncertainty  $\Delta E$ . Given that the number of hard modes is much smaller than the number of relevant soft modes, the effective dimension of the Hilbert space  $\mathcal{H}_{\text{soft}}^{(n)}$  is given by

$$\ln \dim \mathcal{H}_{\text{soft}}^{(n)} = S(E - E_n), \quad (3.26)$$

<sup>12</sup>This is always the case if the effective potential increases monotonically as  $r_*$  moves away from the stretched horizon. Another system exhibiting a similar behavior is Rindler spacetime, although in this case the system has a planar rather than spherical symmetry.

where  $S(E)$  is the entropy density of the system at energy  $E$ . For de Sitter spacetime, it is given by the Gibbons-Hawking entropy

$$S(E) = S_{\text{dS}}(E) = \pi E^2 l_{\text{p}}^2, \quad (3.27)$$

which gives the standard expression of  $S_{\text{GH}} = \pi \alpha^2 / l_{\text{p}}^2$  for  $E = E_{\text{dS}} = \alpha / l_{\text{p}}^2$ .

A typical state described in Sec. III A corresponds to a typical state in the Hilbert space  $\mathcal{H}(E)$  of Eq. (3.25). For our purposes, we need not be very precise about what we mean by typical, but for concreteness one might imagine a state that is typical in  $\mathcal{H}(E)$  under the Haar measure. Denoting a set of generic orthonormal basis states of  $\mathcal{H}_{\text{soft}}^{(n)}$  by  $|\psi_{i_n}^{(n)}\rangle$  ( $i_n = 1, \dots, e^{S(E-E_n)}$ ), the state we are interested in can be written as

$$|\Psi(E)\rangle = \sum_n \sum_{i_n=1}^{e^{S(E-E_n)}} c_{ni_n} |\{n_\alpha\}\rangle |\psi_{i_n}^{(n)}\rangle, \quad (3.28)$$

where the real and imaginary parts of complex coefficients  $c_{ni_n}$ 's can be viewed as taking random values following independently the Gaussian distributions with

$$\begin{aligned} \langle \text{Re } c_{ni_n} \rangle &= \langle \text{Im } c_{ni_n} \rangle = 0, \\ \sqrt{\langle (\text{Re } c_{ni_n})^2 \rangle} &= \sqrt{\langle (\text{Im } c_{ni_n})^2 \rangle} = \frac{1}{\sqrt{2} e^{S_{\text{sys}}}}. \end{aligned} \quad (3.29)$$

Here, the brackets represent the ensemble average over (a sufficiently large portion of) the  $(n, i_n)$  space, and

$$e^{S_{\text{sys}}} = \sum_n e^{S(E-E_n)} \equiv z e^{S(E)} \quad (3.30)$$

is the number of independent states of the form of Eq. (3.28) with

$$z = \sum_n e^{-\frac{E_n}{T_H}}, \quad (3.31)$$

where we have used  $\partial S(E)/\partial E = 1/T_H$  and  $E_n \ll E$ .<sup>13</sup> In other words, we can say that  $c_{ni_n}$  is a complex Gaussian random variable, which implies that the phases of  $c_{ni_n}$ 's are distributed uniformly.

Note that by taking the basis of  $\mathcal{H}_{\text{soft}}^{(n)}$  for each  $n$ , we find that soft-mode states are orthogonal

$$\langle \psi_{i_m}^{(m)} | \psi_{j_n}^{(n)} \rangle = \delta_{mn} \delta_{i_m j_n}. \quad (3.32)$$

This is because states  $|\psi_{i_n}^{(n)}\rangle$  with different  $n$  can be observationally discriminated, which follows from the

<sup>13</sup>We have used the equal sign for a relation that becomes exact in the thermodynamic limit.

distinguishability of different hard-mode states  $|\{n_\alpha\}\rangle$  as well as the fact that the state of combined hard and soft mode system is specified with the minimal uncertainty. In particular, even if some observables (e.g.,  $E$  and  $p$ ) may have to be intrinsically coarse grained (by  $\Delta E$  and  $\Delta p$ ),  $|\psi_{i_n}^{(n)}\rangle$ 's with different  $n$  can still be discriminated at the semiclassical level, and hence are orthogonal.<sup>14</sup>

We also note that  $S_{\text{sys}} = S(E) + \ln z$  is equal to  $S(E)$  at the leading order in  $l_p/r_+$  and  $1/E_n r_+$ . This implies that the standard interpretation of the Gibbons-Hawking entropy as the entropy of de Sitter spacetime persists. A similar comment applies to a black hole, for which the density of soft-mode states is given by the Bekenstein-Hawking entropy

$$S(E) = S_{\text{bh}}(E) = 4\pi E^2 l_p^2, \quad (3.33)$$

although in this case the entire system also has degrees of freedom outside the zone, and  $S_{\text{sys}} = S(E) + \ln z$  represents only the entropy of the black hole system (i.e., the zone and horizon modes) without including the contribution from the far modes.

We now take a complete set of orthonormal states of the form of Eq. (3.28), i.e., states having the energy  $E$  within  $\Delta E$ , in a generic basis:

$$|\Psi_A(E)\rangle = \sum_n e^{S(E-E_n)} \sum_{i_n=1} c_{ni_n}^A |\{n_\alpha\}\rangle |\psi_{i_n}^{(n)}\rangle \quad (A = 1, \dots, e^{S_{\text{sys}}}), \quad (3.34)$$

where

$$\langle \Psi_A(E) | \Psi_B(E) \rangle = \delta_{AB} \iff \sum_n e^{S(E-E_n)} \sum_{i_n=1} c_{ni_n}^{A*} c_{ni_n}^B = \delta_{AB}. \quad (3.35)$$

In general, these states provide a basis for microstates of a semiclassical vacuum. For example, if the spectrum  $\omega_\alpha$  is given such that it represents hard modes inside the stretched de Sitter horizon, then the set  $\{|\Psi_A(E)\rangle\}$  forms a basis for the microstates of the de Sitter vacuum.

We stress that in order for the states of Eq. (3.34) to be microstates of the spacetime, they need to be taken generically in the space of  $\mathcal{H}(E)$ . For example, if we took  $e^{S_{\text{sys}}}$  orthonormal states that are, or approximately are, product states

<sup>14</sup>We assume that such coarse grainings would be performed using smoothing functions which damp very rapidly outside the windows of order  $\Delta E$  and  $\Delta p$  so that Eq. (3.23) is valid with sufficient accuracy. With this assumption, Eq. (3.32) is also valid at the same level of accuracy.

$$|\Psi(E)\rangle \approx |\{n_\alpha\}\rangle |\psi_{i_n}^{(n)}\rangle, \quad (3.36)$$

then these states would still form a ‘‘basis’’ of the microstates in Eq. (3.34) in the sense that all the states of the form in Eq. (3.34) can be obtained by superposing them, although none of them is by itself a microstates of the spacetime under consideration—these special, exponentially rare, states (states whose entanglement structure is significantly different from generic states) are ‘‘firewall’’ [3] states which do not represent the spacetime under consideration. This is because spacetime is a manifestation of the entanglement structure of a holographic boundary state [87–92], and entanglement cannot be represented as a linear operator—the concept of a linear vector space comprising the microstates of a spacetime is only an approximate one [40,93]. This issue will be discussed further in the next section.

Since a semiclassical theory can describe only the dynamics of hard modes, it concerns only about the state of these modes. Thus, the vacuum state  $\rho_{\text{vac}}(E)$  in a semiclassical theory is obtained by taking a typical vacuum microstate and then tracing out the soft modes. Specifically, it is given by

$$\begin{aligned} \rho_{\text{vac}}(E) &= \text{Tr}_{\text{soft}} |\Psi_A(E)\rangle \langle \Psi_A(E)| \\ &= \sum_n \left( \sum_{i_n=1} e^{S(E-E_n)} |c_{ni_n}^A|^2 \right) |\{n_\alpha\}\rangle \langle \{n_\alpha\}| \\ &= \sum_n \frac{e^{-\frac{E_n}{T_H}}}{z} |\{n_\alpha\}\rangle \langle \{n_\alpha\}| + O(e^{-\frac{1}{2}S(E)}). \end{aligned} \quad (3.37)$$

Here, in the last equation we have used

$$\epsilon_n^{AB} \equiv z e^{\frac{E_n}{T_H}} \sum_{i_n=1} e^{S(E-E_n)} c_{ni_n}^{A*} c_{ni_n}^B - \delta_{AB} \approx O(e^{-\frac{1}{2}S(E-E_n)}), \quad (3.38)$$

which can be derived from Eq. (3.29). More precisely, when we vary  $A$  and  $B$ ,  $\epsilon_n^{AB}$  behave as complex Gaussian random variables with mean 0 and variance  $e^{-S(E-E_n)}$ , but obeying

$$(\epsilon_n^{AB})^* = \epsilon_n^{BA}, \quad \sum_n \frac{e^{-\frac{E_n}{T_H}}}{z} \epsilon_n^{AB} = 0, \quad (3.39)$$

where the second relation follows from Eq. (3.35). In the case of de Sitter spacetime, the thermal state in Eq. (3.37) gives the vacuum state describing a static patch of the de Sitter spacetime with Hubble radius  $\alpha = E l_p^2$ .

The hard and soft modes described here provide a refinement of the hard and soft modes defined in Refs. [21–25]. In Refs. [21–25] a simple frequency space criterion was used to define the hard and soft modes, while



in our new definition here, the hard modes are chosen to be a subset of the zone modes whose dynamics we intend to describe at the semiclassical level. This makes it possible, for example, to describe the dynamics of zone modes with  $\omega \sim O(T_H)$  using a semiclassical theory, which was not possible with the previous definition. For many practical purposes, however, the two definitions are interchangeable. For a small object falling into the horizon, for example, the difference between the two definitions is not significant if we choose the frequency cutoff to be sufficiently larger than  $T_H$ . We can then employ the same construction for spacetime beyond the horizon, which we will discuss in Sec. V.

### 1. Evaporating black hole

The situation is more complicated if the region near the horizon is coupled to an ambient/bath system. An important example of this is a black hole in asymptotically flat spacetime; other examples include a small black hole in asymptotically AdS spacetime and a large AdS black hole coupled to a separate bath system. In this case, the near horizon system, consisting of the zone and horizon, evolves in time, and this evolution modifies the vacuum states.

For concreteness, let us focus on a black hole in asymptotically flat spacetime. The state of the entire system then involves the hard, soft, and far (located outside the zone,  $r > r_z$ ) modes. Thus, denoting orthonormal basis states of the far modes by  $|\phi_a\rangle$ , one would consider the state of the system to be given by<sup>15</sup>

$$|\Psi_A(M)\rangle \stackrel{?}{=} \sum_n e^{S_{\text{bh}}(M-E_n)} \sum_{i_n=1}^{S_{\text{rad}}} c_{ni_n a}^A |\{n_\alpha\}\rangle |\psi_{i_n}^{(n)}\rangle |\phi_a\rangle, \quad (3.40)$$

where we have assumed that the black hole system, comprising the hard and soft modes, is at rest and has energy  $M$  (up to the uncertainty of order  $T_H$ ). Here,  $e^{S_{\text{rad}}}$  is the number of independent far-mode states relevant here, i.e., those significantly entangled with the black hole system (typically Hawking radiation emitted earlier from the black hole), and  $A$  is the index for microstates running over

$$A = 1, \dots, e^{S_{s+r}}, \quad S_{s+r} = S_{\text{sys}} + S_{\text{rad}}, \quad (3.41)$$

with  $S_{\text{sys}}$  given by Eq. (3.30) with  $S(E) = S_{\text{bh}}(E)$ . The coefficients  $c_{ni_n a}^A$  satisfy the properties analogous to those in Eq. (3.29)

<sup>15</sup>We assume that the standard issues for a factorization of Hilbert space in quantum field theory, such as those associated with short distance divergences and constraints from gauge invariance, are dealt with appropriately.

$$\begin{aligned} \langle \text{Re } c_{ni_n a}^A \rangle &= \langle \text{Im } c_{ni_n a}^A \rangle = 0, \\ \sqrt{\langle (\text{Re } c_{ni_n a}^A)^2 \rangle} &= \sqrt{\langle (\text{Im } c_{ni_n a}^A)^2 \rangle} = \frac{1}{\sqrt{2}e^{S_{s+r}}}, \end{aligned} \quad (3.42)$$

with brackets representing the ensemble average over (a sufficiently large portion of) the  $(n, i_n, a)$  space. Note that  $|\Psi_A(M)\rangle$  represent microstates of the system with the black hole put in the semiclassical vacuum, and a generic state in the Hilbert space of dimension  $e^{S_{s+r}}$  has the black hole of mass  $M$ . Since black hole evaporation is a thermodynamically irreversible process [94,95], most of these microstates do not become a state with a larger black hole in empty space when evolved backward in time—there is some junk radiation around it. This, however, does not change the fact that there are  $e^{S_{s+r}}$  independent microstates relevant for the discussion here.

The fact that the height of the potential barrier is finite, however, implies that only modes with  $\omega < \omega_{\text{barrier}}^{(\ell)}$  are thermalized in the zone. Given that the dynamics at the stretched horizon is strongly coupled, outgoing modes (i.e., modes moving toward larger  $r$ ) with  $\omega > \omega_{\text{barrier}}^{(\ell)}$  can still be viewed as obeying the thermal distribution, but this is not the case for ingoing modes with  $\omega > \omega_{\text{barrier}}^{(\ell)}$ . At the microscopic level, this implies that the microcanonical ensemble in Sec. III A is taken with the extra constraint that the occupation numbers of ingoing modes with  $\omega > \omega_{\text{barrier}}^{(\ell)}$  are zero, leading to<sup>16</sup>

$$\begin{aligned} |\Psi_A(M)\rangle \stackrel{?}{\propto} & \sum_n \prod_{\substack{\alpha \in \text{ingoing} \\ \omega > \omega_{\text{barrier}}^{(\ell)}}} \delta_{n_\alpha, 0} \\ & \times \sum_{i_n=1}^{e^{S_{\text{bh}}(M-E_n)}} \sum_{a=1}^{S_{\text{rad}}} c_{ni_n a}^A |\{n_\alpha\}\rangle |\psi_{i_n}^{(n)}\rangle |\phi_a\rangle, \end{aligned} \quad (3.43)$$

where the  $\propto$  symbol represents “up to a normalization constant,” and the question mark above it indicates that this relation is tentative and will be updated momentarily. (The effect of the lack of ingoing soft modes with  $\omega > \omega_{\text{barrier}}^{(\ell)}$  is negligible for our purpose.) This corresponds to taking the Hartle-Hawking [96] and Unruh [54] vacua for hard modes with  $\omega < \omega_{\text{barrier}}^{(\ell)}$  and  $> \omega_{\text{barrier}}^{(\ell)}$ , respectively. The effect of the constraint on soft modes is negligible, since the vast

<sup>16</sup>Strictly speaking, there are small but nonzero amplitudes for outgoing hard modes with  $\omega > \omega_{\text{barrier}}^{(\ell)}$  to be reflected back from the potential barrier. We mostly ignore this effect because it is not essential for our discussion. Including it, however, is straightforward; instead of taking the terms with  $\exists \alpha', n_{\alpha'} \neq 0$  to be exactly absent, we keep these terms with small coefficients (compared to those of the terms with  $\forall \alpha', n_{\alpha'} = 0$ ). Note that the size of these coefficients in general depends strongly on  $\{n_{\alpha'}\}$ , reflecting the  $\omega_{\alpha'}$  dependence of the reflection amplitudes.

majority of the relevant modes have  $\omega \sim T_H$  and  $\ell \gg 1$ , and hence  $\omega \ll \omega_{\text{barrier}}^{(\ell)}$ .

The story, however, does not end here. Because of the coupling between the black hole system (zone + horizon) and the asymptotically flat spacetime around the region  $r \sim r_z$ , thermal quanta in the zone “leak” into the latter. This occurs mostly via  $s$ -wave modes that tunnel through the potential barrier. We emphasize that this process, occurring near the edge of the zone [97,98] (for related discussions, see Refs. [99–101]), is governed by semiclassical physics—it does not involve strongly coupled, intrinsically quantum gravitational physics in any significant way. Given a black hole microstate in Eq. (3.43) (slightly modified due to backreaction; see below), the emission of quanta into the asymptotic region occurs unitarily following the dynamics of standard quantum field theory. The apparent violation of unitarity in Hawking’s analysis [1] occurs because we cannot calculate the configuration of zone mode quanta using semiclassical theory due to the strong dynamics *near the stretched horizon*. It is this incalculability that makes the semiclassical description of Hawking radiation, obtained after tracing out the soft modes, intrinsically thermal and hence leading to a mixed final state. Physics away from the stretched horizon can indeed be fully semiclassical, even within and at the edge of the zone.<sup>17</sup>

The emission of Hawking particles to the ambient space at  $r \sim r_z$  gives a backreaction to the state of the black hole. In this region, quanta of the zone region is leaked into the ambient space through tunneling the potential barrier (and also via thermal hopping to some extent). This removes some of the quanta that would be reflected back to the zone by the potential, producing a deficit in ingoing zone mode quanta relative to those in the states of Eq. (3.43), i.e., with

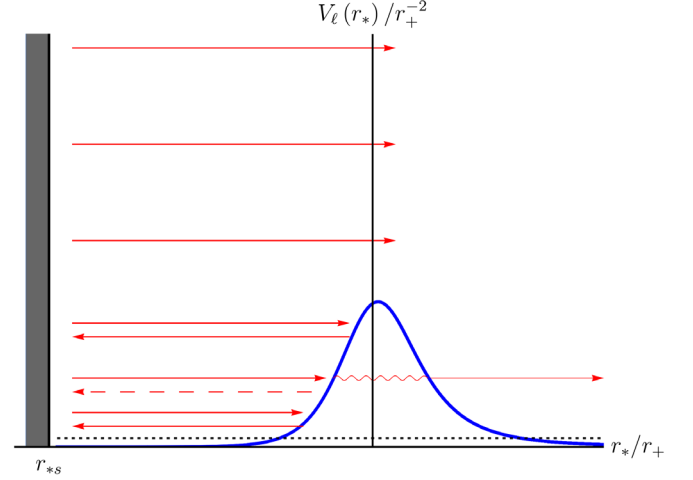


FIG. 4. A schematic depiction of quanta occupying various modes. Solid arrows indicate that the occupation numbers of the modes are determined by the thermodynamic consideration as described in Sec. III A. For zone modes, this applies to the outgoing mode as well as the ingoing modes with  $\omega \lesssim \omega_{\text{barrier}}^{(\ell)}$ , except that backreaction of Hawking emission at  $r \sim r_z$  leads to a deficit for some of these ingoing modes as indicated by the dashed arrow.

the Hartle-Hawking vacuum. Note that the process occurs only for low energy fields of mass  $m_\Phi \lesssim T_H$ . Given that an  $O(1)$  number of quanta are emitted within each time interval of order  $1/T_H \sim r_+$ , there are  $\sim \ln(r_+/l_s)$  quanta missing throughout the zone for each low energy field  $\Phi$  of  $m_\Phi \lesssim T_H$ . Denoting the annihilation operators for hard mode quanta by  $b_{\bar{\alpha}}$  (see below for more detail), we finally find that the microstates of an evaporating black hole are given by

$$|\Psi_A(M)\rangle \propto \prod_{\bar{\alpha} \in \text{ingoing}} b_{\bar{\alpha}} \sum_n \prod_{\alpha' \in \text{ingoing}} \delta_{n_{\alpha'}, 0} \sum_{i_n=1}^{e^{S_{\text{bh}}(M-E_n)}} \sum_{a=1}^{S_{\text{rad}}} c_{ni,a}^A |\{n_\alpha\}\rangle |\psi_{i_n}^{(n)}\rangle |\phi_a\rangle, \quad (3.44)$$

where the number of annihilation operators  $b_{\bar{\alpha}}$  in the product is at most of order  $\ln(r_+/l_s)$  for each field; the precise number depends on the choice of the hard modes. A schematic depiction of the occupation of various modes is given in Fig. 4.

The deficit of ingoing modes described above implies that there is a negative energy flux carrying negative

entropy for each field of  $m_\Phi \lesssim T_H$ . Note that these energy and entropy are measured with respect to the thermal, Hartle-Hawking vacuum state in Eq. (3.43) for modes with  $\omega \sim T_H$ . The flux has negative entropy because with the lack of some of the zone mode quanta, the number of independent states realizing the most probable configuration is smaller by  $e^{O(1)\ln(r_+/l_s)}$  for each field. Thus, the microstate index  $A$  in Eq. (3.44) runs effectively only for

$$A = 1, \dots, A_{\text{max}}, \quad \ln A_{\text{max}} - S_{\text{tot}} \sim -\ln \frac{r_+}{l_s}, \quad (3.45)$$

where  $S_{\text{tot}}$  is the coarse-grained entropy of the total system ignoring the backreaction

<sup>17</sup>This implies that the process of black hole mining [102,103] also occurs unitarily, which is governed by semiclassical physics if it is performed away from the stretched horizon. It also implies that the semiclassical calculation of the gray body factor, such as that in Ref. [104], is valid as long as  $r_+/l_s$  is sufficiently large; see below.

$$e^{S_{\text{tot}}} = \sum_n \prod_{\substack{\alpha \in \text{ingoing} \\ \omega > \omega_{\text{barrier}}^{(\ell)}}} \delta_{n_{\alpha}, 0} e^{S_{\text{bh}}(M-E_n)} e^{S_{\text{rad}}} \equiv z' e^{S_{\text{bh}}(M)+S_{\text{rad}}}, \quad (3.46)$$

and we have taken the number of low energy degrees of freedom relevant for the emission process to be of  $O(1)$ .<sup>18</sup> As discussed in Refs. [97,98], this negative entropy is essential for the unitarity of the black hole evolution, in particular for the black hole to keep relaxing into a lower mass black hole while absorbing the negative energy flux. Again, we emphasize that semiclassical physics is sufficient to understand the unitary emission process at  $r \sim r_z$  and the resulting emergence of a negative energy and entropy flux. The unknown UV physics enters only in the process occurring at the stretched horizon, in which the ingoing negative energy-entropy flux is absorbed into horizon modes and the black hole relaxes into its semiclassical vacuum state.

So far, we have assumed that our black hole is large, in particular  $\ln(r_+/l_s) \gg 1$ . In this limit, the difference of energies between adjacent discrete levels for modes relevant for Hawking emission,  $\sim r_+ / \ln(r_+/l_s)$ , is much smaller than the uncertainty of energy of Hawking quanta,  $\sim r_+$ , so that the effect of the discreteness of energy levels is negligible in calculating the spectrum of Hawking radiation. In particular, the semiclassical calculation of the spectrum, including the gray body factor, persists with high precision. If the value of  $\ln(r_+/l_s)$  is reduced, however, the effect of the discreteness of levels may become important. In particular, if the size of the classically allowed region  $r_{*b}^{(\ell)} - r_{*s} \approx 2r_+ \ln(r_+/l_s)$ , becomes smaller than a half wavelength of a Hawking quantum  $\lambda_H/2 \approx \pi/T_H \approx 4\pi^2 r_+$ , i.e.

$$r_+ \lesssim e^{2\pi^2} l_s, \quad (3.47)$$

then the effect may become non-negligible. (This condition can also be obtained by requiring  $\Delta\omega$  in Eq. (2.24) to be larger than  $T_H$ .) A naive guess is that in this regime, the energy of each Hawking quantum is larger than that obtained by the semiclassical calculation, since the frequency of the lowest energy level is expected to become larger than  $T_H$ . An interesting point is that the black hole enters the regime of Eq. (3.47) before its mass is reduced to the Planck mass, since this condition can be written as

$$M \lesssim \frac{e^{2\pi^2} l_s}{2l_p^2}. \quad (3.48)$$

For  $l_s/l_p \approx 20$ , as we might expect in our universe, the right-hand side is  $4 \times 10^9$  times larger than the Planck mass

<sup>18</sup>This implies that the range of  $i_n$  in Eq. (3.44) is, strictly speaking, smaller than that in Eqs. (3.40) or (3.43). For the lack of a better notation, we interpret the sum in Eq. (3.44), and analogous expressions later, to include this minor effect.

(corresponding to the black hole of  $T_H \approx 1 \times 10^8$  GeV). We leave further discussion of this issue, including its possible phenomenological implications, for the future.

#### D. Excited states

Before concluding this section, let us discuss semiclassical excitations in the zone. The states we have considered so far are typical states in a suitably defined microcanonical ensemble. For a black hole, for example, the relevant ensemble consists of the states that contain a fixed energy  $E$  in a spatial region  $r \lesssim r_z$  within an uncertainty  $\Delta E$ . These states are all vacuum states from the point of view of semiclassical theory as indicated by the fact that the Bekenstein-Hawking entropy is associated with the background (vacuum) spacetime in semiclassical theory.

We can, however, consider atypical states obtained by acting annihilation and/or creation operators of hard modes

$$b_\gamma = \sum_n \sqrt{n_\gamma} |\{n_\alpha - \delta_{\alpha\gamma}\}\{n_\alpha\}|, \quad (3.49)$$

$$b_\gamma^\dagger = \sum_n \sqrt{n_\gamma + 1} |\{n_\alpha + \delta_{\alpha\gamma}\}\{n_\alpha\}| \quad (3.50)$$

A state obtained in this way is not typical as it does not have the most probable configuration among the states in the ensemble. On these vacuum states, where  $\gamma$  specifies the mode which is annihilated/created by the operator. (States obtained by acting annihilation operators can become relevant when one considers backreaction of the Hawking emission or black hole mining process. In particular, some of the operators  $b_\alpha^{(z)}$  in Eq. (3.44) can be superpositions of the  $b_\gamma$  operators, if we choose the hard modes to include the relevant zone modes.)

We are interested in semiclassical excitations whose backreaction to the geometry is negligible, or regarded as being small, for example a baseball falling into an astronomical black hole. This implies that the number of creation/annihilation operators that can be acted on a vacuum state is limited, so that their algebra defined on a vacuum state does not close in a strict mathematical sense. This type of structure is in fact common in erecting a semiclassical theory in quantum gravity, and here we simply treat the space of states obtained in this way as a Hilbert space; for a mathematically more rigorous treatment, see, e.g., Ref. [105]. In holography, one can think of this space as a code subspace embedded in a physical Hilbert space [106–108], although we do not discuss the error correcting nature of operators in this paper.

Strictly speaking, the space of semiclassical excitations generated by  $b_\gamma$  and  $b_\gamma^\dagger$  acted on each vacuum state is not orthogonal to the space of vacuum microstates; namely, the Hilbert space cannot be strictly written as  $\mathcal{H}_{\text{exc}} \otimes \mathcal{H}_{\text{vac}}$  [25]. One can see this by calculating inner products between states obtained by exciting black hole microstates in Eq. (3.43):

$$\langle \Psi_A(M) | b_\beta^\dagger b_\gamma | \Psi_B(M) \rangle = \delta_{\beta\gamma} \sum_n \prod_{\substack{\alpha' \in \text{ingoing} \\ \omega > \omega_{\text{barrier}}^{(\ell)}}} \delta_{n_{\alpha'}, 0} \sum_{i_n=1}^{e^{S_{\text{bh}}(M-E_n)}} \sum_{a=1}^{e^{S_{\text{rad}}}} n_\gamma c_{ni_n a}^{A*} c_{ni_n a}^B, \quad (3.51)$$

$$\langle \Psi_A(M) | b_\beta b_\gamma^\dagger | \Psi_B(M) \rangle = \delta_{\beta\gamma} \sum_n \prod_{\substack{\alpha' \in \text{ingoing} \\ \omega > \omega_{\text{barrier}}^{(\ell)}}} \delta_{n_{\alpha'}, 0} \sum_{i_n=1}^{e^{S_{\text{bh}}(M-E_n)}} \sum_{a=1}^{e^{S_{\text{rad}}}} (n_\gamma + 1) c_{ni_n a}^{A*} c_{ni_n a}^B. \quad (3.52)$$

These are not proportional to  $\delta_{AB}$  in general, so that excited states built on different vacuum microstates are not necessarily orthogonal. However, for  $A \neq B$ , the right-hand sides of the above equations are exponentially suppressed by a factor of  $e^{-S_{\text{tot}}/2}$ , where we have taken

$$\langle \Psi_A(M) | \Psi_B(M) \rangle = \delta_{AB} \Leftrightarrow \sum_n \prod_{\substack{\alpha' \in \text{ingoing} \\ \omega > \omega_{\text{barrier}}^{(\ell)}}} \delta_{n_{\alpha'}, 0} \sum_{i_n=1}^{e^{S_{\text{bh}}(M-E_n)}} \sum_{a=1}^{e^{S_{\text{rad}}}} c_{ni_n a}^{A*} c_{ni_n a}^B = \delta_{AB}. \quad (3.53)$$

Therefore, the deviation from the product space structure is exponentially small.

Incidentally, the annihilation and creation operators in Eqs. (3.49) and (3.50) satisfy the standard commutation relations

$$[b_\beta, b_\gamma^\dagger] = \delta_{\beta\gamma} \sum_n |\{n_\alpha\}\rangle \langle \{n_\alpha\}|, \quad [b_\beta, b_\gamma] = [b_\beta^\dagger, b_\gamma^\dagger] = 0 \quad (3.54)$$

as operators, without having an exponentially small correction. In erecting a semiclassical theory, one regards all the microstates  $|\Psi_A(M)\rangle$  as representing the same geometry  $\mathcal{M}$ . This allows us to define field operators

$$\Phi_\Gamma(x) = \sum_{\gamma'} (b_\gamma u_{\gamma'}(x) + b_\gamma^\dagger v_{\gamma'}(x)), \quad (3.55)$$

where we have split the index  $\gamma$  into the index for species  $\Gamma$  and that for other quantum numbers  $\gamma'$ , e.g., a component of spin, whose structure may depend on  $\Gamma$ :  $\gamma = (\Gamma, \gamma')$ . Here,  $u_{\gamma'}(x)$  and  $v_{\gamma'}(x)$  are mode functions defined in the allowed region of  $\mathcal{M}$ . Because of Eq. (3.54), these field operators and their conjugate momenta obey the standard equal-time commutation relations, so that the resulting semiclassical theory respects causality exactly. For black hole and de Sitter spacetimes, this theory can be used to describe physics in the regions  $r > r_s$  and  $< r_s$ , respectively.

#### IV. HOLOGRAPHIC DESCRIPTION

So far, we have assumed that the fundamental description of a system represents spacetime in the allowed region (with the horizon degrees of freedom included). For example, this corresponds to taking a distant view for a black hole and a static patch view for de Sitter spacetime. Why is this the case?

In this section, we discuss this issue from the perspective of holography. We assume that these spacetimes arise in setups in which the holographic description has only a

single boundary; in particular, the black hole is formed by a gravitational collapse, and de Sitter spacetime arises in a cosmological context, for example as a universe created by bubble nucleation [109] which is filled with a positive cosmological constant. These setups are, arguably, more “realistic” than those with multiple boundaries, which will be discussed in Sec. VI. We conclude the section with a discussion on how the (effective) boundary Hilbert spaces describing the spacetimes considered here are related to the infinite-dimensional “fundamental” Hilbert space.

#### A. Black hole spacetime

We begin by considering a collapse formed black hole in an asymptotically AdS spacetime. We would like to know what spacetime picture can be obtained from the boundary CFT by reconstructing the bulk in a simple manner. By simple reconstruction, we mean reconstruction of the bulk using only low-complexity, causally propagating operators and sources in the CFT [16]. In particular, we are interested in obtaining a “gauge-fixed” bulk description where the spacetime is foliated by equal-time hypersurfaces, using which the canonical formulation of quantum mechanics can be employed.

One way to obtain such a description is to “pull” the boundary into the bulk by coarse graining boundary degrees of freedom [13,14,110]. Based on intuition from tensor networks [107,108,111], one can consider a series of states [112] defined on successfully “renormalized” boundaries obtained by moving the original boundary inward to the bulk. The coarse-grained degrees of freedom are



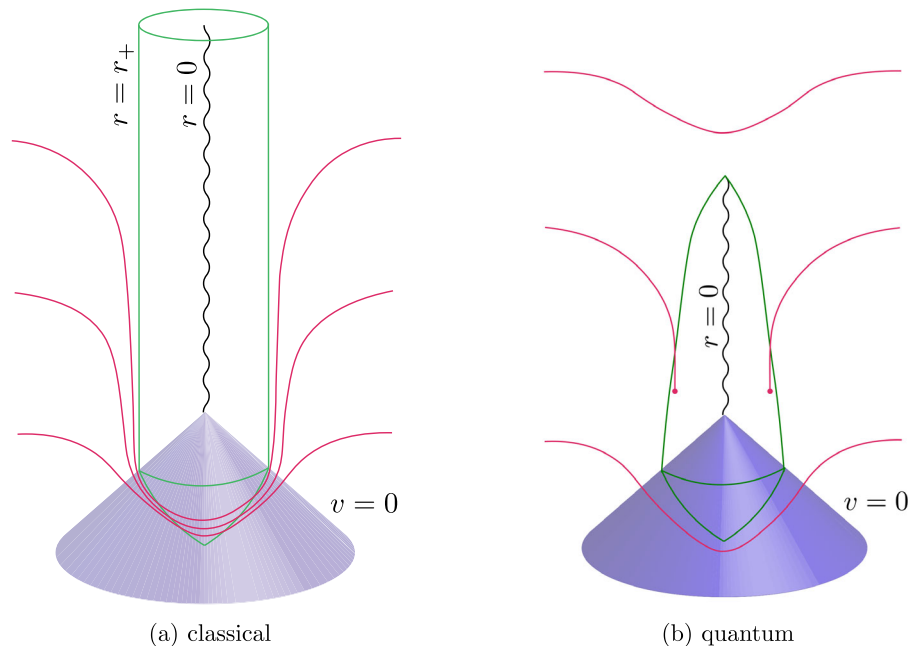


FIG. 5. Holographic slices in a black hole spacetime in ingoing Eddington-Finkelstein coordinates  $(v, r)$  obtained (a) without and (b) with quantum effects in the bulk. These slices correspond to equal-time hypersurfaces in the bulk, which approach Schwarzschild time slices near the black hole. The figures are taken from Refs. [13,14].

distributed locally on a renormalized boundary, although the dynamics they obey are not necessarily local.

This procedure is expected to work beyond the AdS/CFT context, which we assume to be the case, and henceforth we will not necessarily assume that the spacetime is asymptotically AdS. A particular way of performing this renormalization is to “continuously” coarse grain boundary degrees of freedom uniformly throughout the renormalized boundary space. This leads to a specific spacelike or null hypersurface swept by a series of renormalized boundaries called a *holographic slice* [13], which we will discuss in more detail below. This surface plays the role of an equal-time hypersurface in the bulk, providing a gauge fixing necessary for the canonical formulation.

In Fig. 5(a), we depict holographic slices in a black hole spacetime obtained without including quantum effects in the bulk. We find that the slices do not enter the black hole interior, hence providing a distant view of the black hole. This is because the black hole horizon, sufficiently after the black hole has stabilized, plays a role of a barrier [113] for extremal surfaces, which are used to move renormalized boundaries. In fact, the slices approach Schwarzschild time slices near the black hole; they then stay near the horizon for long time and are eventually capped off at  $r = 0$ .

In order to describe an evaporating black hole, quantum effects in the bulk must be included. The procedure of forming a holographic slice can be extended to incorporate these effects [14]. In this case, the holographic slice is smoothly capped off at  $r = 0$  initially, but at some

(boundary) time it becomes asymptoting to a quantum extremal surface located (approximately) at the horizon; and as time progresses further, it again becomes a surface without a hole. This is depicted in Fig. 5(b). This behavior of holographic slices, in fact, can be used to define what we mean by the “formation” and “evaporation” of a black hole. The holographic slices approximate Schwarzschild time slices in this case as well, giving a distant description of the black hole. Note that the region swept by the holographic slices is essentially the simple wedge [16], which can be reconstructed in a simple manner using the Hamilton-Kabat-Lifschytz-Lowe (HKLL) procedure [11,12,114] together with boundary time evolutions with sources.

In any event, the description of a black hole that naturally results from a holographic theory with a single boundary is an exterior/distant one. In this description, the unitarity of black hole evolution is not an issue, since the stretched horizon behaves as a regular material surface from the point of view of quantum information flow. The question, rather, is in what sense the near empty interior region exists for an infalling observer as predicted by general relativity. We will come back to this issue in Sec. V.

### B. de Sitter (or cosmological) spacetime

To discuss holographic descriptions beyond AdS/CFT, we need to introduce a “boundary” in spacetime on which a holographic state can be defined. One way to do this is to consider a renormalized boundary deep in the bulk—called a renormalized leaf [13,14] in this context—and

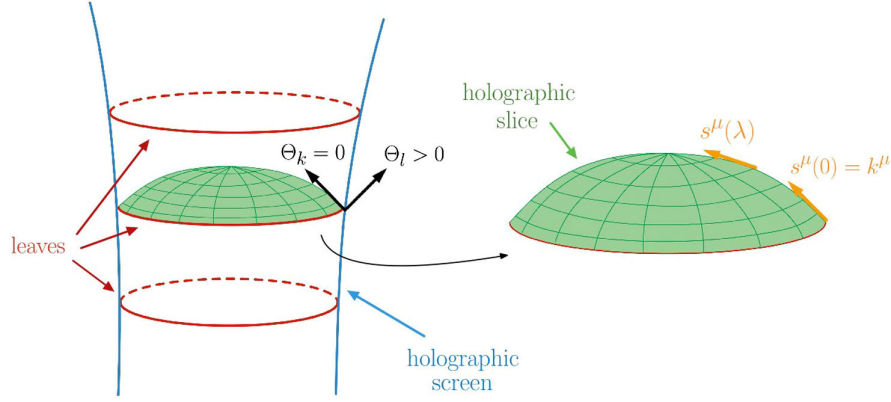


FIG. 6. A holographic theory resides on a holographic screen, which is a hypersurface foliated by quantum marginally antitrapped (or trapped) surfaces called leaves. At a given boundary time, a boundary state lives on a leaf, from which one can reconstruct a bulk equal-time hypersurface using the flow equation in Eq. (4.1).

“unrenormalize” it by successively integrating in relevant degrees of freedom, making the holographic slice grow toward the “nonrenormalized” boundary. This procedure depends on the background spacetime, but we can perform it in each branch of a state representing a well-defined spacetime in the semiclassical limit.

### 1. General picture

To be specific, let us adopt the scheme in Refs. [13,14], in which successively renormalized boundaries  $\sigma(\lambda)$ , parameterized by “renormalization scale”  $\lambda$ , are obtained by the flow equation

$$\frac{dx^\mu}{d\lambda} = s^\mu, \quad (4.1)$$

where  $x^\mu$  are the embedding coordinates of the codimension-2 surface  $\sigma(\lambda)$ , on which the holographic states are defined, and  $s^\mu$  is the evolution vector given by

$$s^\mu = \frac{1}{2}(\Theta_k l^\mu + \Theta_l k^\mu). \quad (4.2)$$

Here,  $\{k^\mu, l^\mu\}$  are the future-directed null vectors orthogonal to  $\sigma(\lambda)$ , normalized such that  $k_\mu l^\mu = -2$ ,<sup>19</sup> and  $\Theta_{k,l}$  represent quantum expansions in the corresponding directions [115], which reduce to the classical expansions  $\theta_{k,l}$  in the limit that bulk quantum effects are ignored.<sup>20</sup> We assume that  $\sigma(\lambda)$  is neither quantum trapped nor

antitrapped, so that the quantum expansion  $\Theta_s$  associated with the evolution vector  $s^\mu$  satisfies

$$\Theta_s = \Theta_k \Theta_l \leq 0. \quad (4.3)$$

This condition is needed for the consistency of the interpretation that the flow (toward larger  $\lambda$ ) corresponds to coarse graining of the boundary degrees of freedom.

Since a holographic slice is nothing but a codimension-1 surface swept by  $\sigma(\lambda)$ , we can extend it “outward” using the flow equation. For a generic spacetime, we can perform this extension up to the point where one of the  $\Theta_{k,l}$  becomes zero, i.e., until  $\sigma(\lambda)$  becomes a quantum marginally trapped or antitrapped surface. Suppose that  $\Theta_k = 0$  throughout this surface, which we call  $\sigma(0)$ . Then, the dimension of the Hilbert space associated with states on the holographic slice is bounded [115,116] by the generalized entropy of  $\sigma(0)$ , so that the  $\sigma(0)$  can be identified as an equal-time surface of a nonrenormalized boundary—or a leaf—on which a nonrenormalized boundary state is given.<sup>21</sup> A codimension-1 surface foliated by such leaves is called a holographic screen [117], or  $Q$ -screen in the quantum context [118], on which a holographic theory for general spacetimes is supposed to live [40–42]. This framework is indeed consistent with the hypothetical relationship [112] between bulk spacetime and the entanglement structure of a boundary state. A sketch of a holographic screen, leaves, and a holographic slice is given in Fig. 6.

In some cases, nonrenormalized leaves cannot be reached with a finite evolution in  $\lambda$ . In particular, this occurs in asymptotically AdS and flat spacetimes. In an asymptotically AdS spacetime, the metric in the asymptotic region can be expanded in a Fefferman-Graham series [119]

<sup>19</sup>This condition does not fix the normalizations of  $k^\mu$  and  $l^\mu$  separately, but it is sufficient to ensure the validity of the following treatment. In particular,  $s^\mu$  is invariant under rescalings of  $k^\mu$  and  $l^\mu$  satisfying  $k_\mu l^\mu = -2$ . Below, we fix this freedom conveniently in each setup when we give explicit expressions for  $k^\mu$ ,  $l^\mu$ ,  $\theta_{k,l}$ , and  $\Theta_{k,l}$ .

<sup>20</sup>More precisely, we must use a modified version of the quantum expansion which includes a bulk entropy contribution from an “exterior” region, as described in Ref. [14].

<sup>21</sup>If we follow the flow further beyond  $\lambda = 0$ , then we lose this property; in this sense,  $\sigma(0)$  is the “maximally unrenormalized” leaf.

$$ds^2 = \frac{L^2}{z^2} \{g_{ab}(x^a, z) dx^a dx^b + dz^2\}, \quad (4.4)$$

where  $L$  is the AdS length scale,  $a, b = 0, \dots, d-1$ , and  $g_{ab}(x^a, z) = g_{ab}^{(0)}(x^a) + z^2 g_{ab}^{(2)}(x^a) + \dots$  with  $g_{ab}^{(0)}$  being the conformal boundary metric. Suppose that  $\sigma(\lambda)$  is taken to be a constant  $t = x^0$  surface at a constant  $z = \epsilon$ . The null normals are then given by  $k_\mu = (L/\epsilon)(-dt + dz)$  and  $l_\mu = (L/\epsilon)(-dt - dz)$ , yielding  $\Theta_k = -(d-1)/L$  and  $\Theta_l = (d-1)/L$  up to higher order corrections in  $\epsilon$ . Here, we have used the fact that the quantum expansions  $\Theta_{k,l}$  approach classical expansions  $\theta_{k,l}$  in the asymptotic region because of the lack of matter there. This implies that a leaf of any regularized boundary located at  $z > 0$  is, in fact, a renormalized one, and that the nonrenormalized holographic screen can lie formally only at spacelike infinity,<sup>22</sup> where a nonregularized holographic CFT lives in AdS/CFT. It also gives the evolution vector

$$s^\mu = \frac{(d-1)\epsilon}{L^2} \frac{\partial}{\partial z} + O(\epsilon^2), \quad (4.5)$$

showing that the holographic slice extending inward from a regularized boundary evolves initially in the  $z$  direction, up to corrections suppressed by  $\epsilon$ ; see Fig. 7(a). This behavior is the same as that of the conventional holographic renormalization group flow [120–124] in the AdS/CFT correspondence. We also find that the flow freezes,  $|dz/d\lambda| \rightarrow 0$ , as  $z \rightarrow 0$ .

For an asymptotically flat spacetime, the metric in the asymptotic region can be expanded in the Bondi-Sachs form [125,126] as

$$ds^2 = -\frac{V}{r} e^{2\beta} du^2 - 2e^{2\beta} du dr + r^2 h_{AB}(dx^A - U^A du)(dx^B - U^B du), \quad (4.6)$$

where  $A, B = 1, \dots, d-1$ , and each function admits a large  $r$  expansion of the form  $V = r + O(1)$ ,  $\beta = O(r^{-2})$ ,  $U^A = O(r^{-2})$ , and  $h_{AB} = O(1)$ . As in the case of an asymptotically AdS spacetime, let us consider  $\sigma(\lambda)$  which is a constant time slice of the surface  $r = R$ . The null normals are then given by  $k_\mu = du$  and  $l_\mu = (V/r)du + 2dr$ , leading to  $\Theta_k = -2/R$  and  $\Theta_l = 2/R$  at the leading order in  $1/R$ . We thus find that a leaf of a regularized holographic screen is a renormalized one, and the flow vector is

$$s^\mu = \frac{2}{R} \left( -\frac{\partial}{\partial u} + \frac{\partial}{\partial r} \right) + O\left(\frac{1}{R^2}\right), \quad (4.7)$$

<sup>22</sup>We adopt a definition of the nonrenormalized holographic screen such that it is a hypersurface foliated by leaves, on which at least one of the  $\Theta_{k,l}$  vanishes.

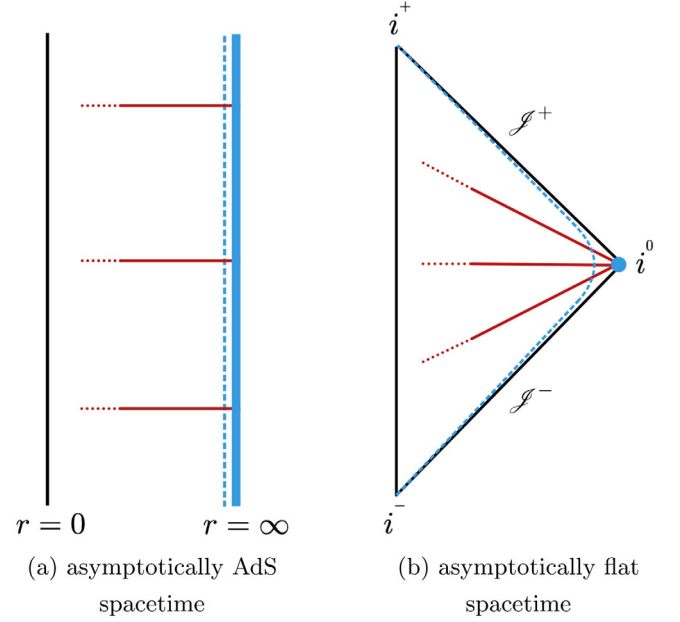


FIG. 7. Renormalized holographic screens, on which regularized holographic theories reside, are timelike hypersurfaces in an asymptotically AdS and flat spacetimes, which are depicted by the dotted blue lines in the Penrose diagrams in (a) and (b), respectively. In both cases, holographic slices agree with the time slices given by timelike Killing vectors in the asymptotic regions. In the nonrenormalized limit, physics occurring in a finite time is described by holographic theories located at the conformal boundary ( $r = \infty$ ) and spatial infinity ( $i^0$ ) in the cases of asymptotically AdS and flat spacetimes, respectively.

showing that the generated flow follows a hypersurface of equal Minkowski time in the inward radial direction, up to corrections suppressed by  $1/R$ , and that it freezes as  $R \rightarrow \infty$ . This is illustrated in Fig. 7(b), from which we see that any process occurring in the bulk within a finite time interval can be described by a boundary theory located at spatial infinity  $i^0$ . By taking the time interval to infinity, one recovers an  $S$ -matrix description in the bulk. This view is consistent with discussions of flat space holography advanced, e.g., in Refs. [127–131].

## 2. de Sitter spacetime

Let us now discuss de Sitter spacetime. As we will see below, care is needed to consider holography of de Sitter spacetime in a static patch. Our view is that in this context, the concept of exact de Sitter spacetime arises only as a result of idealization, very much analogous to an eternal single-sided black hole in an asymptotically flat spacetime (which does not exist because of Hawking radiation).<sup>23</sup>

For now, we bypass this issue by considering de Sitter spacetime in a cosmological setup. In particular, we consider

<sup>23</sup>This is consistent with the expectation from string theory that there is no absolutely stable de Sitter vacuum [132,133].

an empty bubble universe in which there is a positive vacuum energy density  $\rho_\Lambda$ . The interior of the bubble is an open Friedmann-Lemaître-Robertson-Walker (FLRW) universe described by the metric

$$ds^2 = -d\tau^2 + a(\tau)^2 \left[ d\chi^2 + \left\{ \frac{1}{\sqrt{-\kappa}} \sinh(\sqrt{-\kappa}\chi) \right\}^2 d\Omega^2 \right]. \quad (4.8)$$

Here,

$$a(\tau) = \sqrt{\frac{-\kappa}{\tilde{\Lambda}}} \sinh(\sqrt{\tilde{\Lambda}}\tau) \quad (4.9)$$

is the scale factor, where  $\tilde{\Lambda}$  is related to the vacuum energy density by  $\tilde{\Lambda} = 8\pi\rho_\Lambda l_p^2/3$ , and  $\kappa < 0$  is a curvature parameter related to the physical curvature radius by  $r_{\text{curv}}(\tau) = a(\tau)/\sqrt{-\kappa}$ . The holographic screen is located at the apparent horizon, where  $\theta_k = 0$  and  $\theta_l > 0$ :

$$\chi = \chi_{\text{sc}}(\tau) = \frac{1}{\sqrt{-\kappa}} \ln \left[ \coth \frac{\sqrt{\tilde{\Lambda}}\tau}{2} \right], \quad (4.10)$$

where we have taken  $k^\mu$  and  $l^\mu$  to be the future-directed ingoing and outgoing null vector orthogonal to a leaf, respectively, and we have indexed leaves by FLRW times at their locations. Here, we have ignored the difference between classical and quantum expansions, which is valid because we consider the semiclassical vacuum state, implying that Gibbons-Hawking radiation is not extracted as semiclassical radiation.

At late times, the universe behaves like a de Sitter spacetime:

$$a(\tau) \approx \frac{\sqrt{-\kappa}}{2\sqrt{\tilde{\Lambda}}} e^{\sqrt{\tilde{\Lambda}}\tau} \quad \text{for } \tau \gg \frac{1}{\sqrt{\tilde{\Lambda}}}. \quad (4.11)$$

The metric in Eq. (4.8) then becomes flat slicing of the de Sitter spacetime; see Fig. 8. On the other hand, the physical distance to the holographic screen approaches the Hubble radius  $1/\sqrt{\tilde{\Lambda}}$ :

$$\begin{aligned} a(\tau)\chi_{\text{sc}}(\tau) &= \frac{1}{\sqrt{\tilde{\Lambda}}} \sinh(\sqrt{\tilde{\Lambda}}\tau) \ln \left[ \coth \frac{\sqrt{\tilde{\Lambda}}\tau}{2} \right] \\ &\xrightarrow{\tau \gg \frac{1}{\sqrt{\tilde{\Lambda}}}} \frac{1}{\sqrt{\tilde{\Lambda}}} \left( 1 - \frac{2}{3} e^{-2\sqrt{\tilde{\Lambda}}\tau} \right). \end{aligned} \quad (4.12)$$

The holographic theory thus describes a static patch of the late-time de Sitter spacetime, i.e., the patch inside the holographic screen.

The holographic slices approach static de Sitter time slices at late times. To see this, we can calculate  $k^\mu$  and  $l^\mu$  for arbitrary  $(\tau, \chi)$

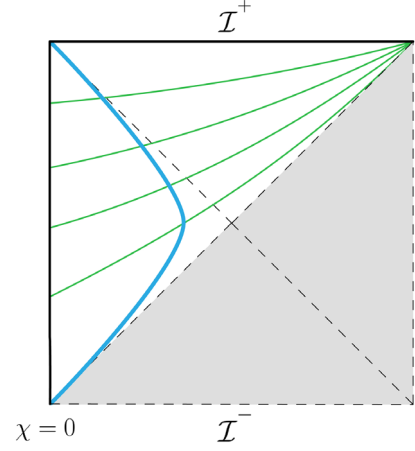


FIG. 8. Penrose diagram of de Sitter spacetime. A de Sitter bubble corresponds to the upper-left half of this diagram, where FLRW equal-time hypersurfaces are drawn by green lines. The holographic screen is located along the trajectory of the apparent horizon, which is depicted by the solid blue curve.

$$k^\mu = \begin{pmatrix} 1 \\ -\sqrt{\frac{\tilde{\Lambda}}{-\kappa}} \frac{1}{\sinh(\sqrt{\tilde{\Lambda}}\tau)} \end{pmatrix}, \quad l^\mu = \begin{pmatrix} 1 \\ \sqrt{\frac{\tilde{\Lambda}}{-\kappa}} \frac{1}{\sinh(\sqrt{\tilde{\Lambda}}\tau)} \end{pmatrix} \quad (4.13)$$

and their associated quantum expansions

$$\begin{cases} \Theta_k \approx \theta_k = 2\sqrt{\tilde{\Lambda}} \left[ \coth(\sqrt{\tilde{\Lambda}}\tau) - \frac{\coth(\sqrt{-\kappa}\chi)}{\sinh(\sqrt{\tilde{\Lambda}}\tau)} \right], \\ \Theta_l \approx \theta_l = 2\sqrt{\tilde{\Lambda}} \left[ \coth(\sqrt{\tilde{\Lambda}}\tau) + \frac{\coth(\sqrt{-\kappa}\chi)}{\sinh(\sqrt{\tilde{\Lambda}}\tau)} \right]. \end{cases} \quad (4.14)$$

The evolution vector is then

$$\begin{pmatrix} s^\tau \\ s^\chi \end{pmatrix} = 2\sqrt{\tilde{\Lambda}} \begin{pmatrix} \coth(\sqrt{\tilde{\Lambda}}\tau) \\ -\sqrt{\frac{\tilde{\Lambda}}{-\kappa}} \frac{\coth(\sqrt{-\kappa}\chi)}{\sinh^2(\sqrt{\tilde{\Lambda}}\tau)} \end{pmatrix}. \quad (4.15)$$

We are interested in the spacetime region well inside the holographic screen

$$\chi \ll \chi_{\text{sc}}(\tau) \xrightarrow{e^{\sqrt{\tilde{\Lambda}}\tau} \gg 1} \sqrt{-\kappa}\chi \ll 2e^{-\sqrt{\tilde{\Lambda}}\tau} \quad (4.16)$$

at late times

$$\alpha = \frac{1}{\sqrt{\tilde{\Lambda}}} \ll r_{\text{curv}}(\tau) \rightarrow e^{\sqrt{\tilde{\Lambda}}\tau} \gg 1, \quad (4.17)$$

where  $\alpha$  is the Hubble radius of the late time universe. In this region, the relation between the FLRW coordinates  $(\tau, \chi)$  and the de Sitter static coordinates  $(t, r)$  is given by



$$\begin{cases} \alpha e^{\frac{t}{\alpha}} = \sqrt{\alpha^2 - r^2} e^{\frac{t}{\alpha}}, \\ \frac{\alpha}{2} \sqrt{-\kappa\chi} e^{\frac{t}{\alpha}} = r, \end{cases} \quad (4.18)$$

leading to the  $s^\mu$  vector in the static coordinates

$$\begin{pmatrix} s^t \\ s^r \end{pmatrix} \approx \begin{pmatrix} -\frac{8}{3\alpha^2} \frac{r^2}{(1-r^2/\alpha^2)^2} e^{-\frac{2t}{\alpha}} \\ -\frac{2}{r} (1 - \frac{r^2}{\alpha^2}) \end{pmatrix} \approx \begin{pmatrix} -\frac{8r^2}{3\alpha^2} e^{-\frac{2t}{\alpha}} \\ -\frac{2}{r} \end{pmatrix}. \quad (4.19)$$

We thus find that both  $s^t$  and  $s^r$  are negative, and  $|s^t| \ll |s^r|$  in the relevant region.

From the above analysis, we conclude that in the holographic description of de Sitter spacetime, the region swept by holographic slices is the interior of the static patch. Thus, what is analogous to the exterior of a single-sided black hole is the interior of a static patch in de Sitter spacetime. This indicates that the interior of the static patch is the region in which semiclassical field operators can be reconstructed in a simple manner. We note that the same conclusion can also be obtained by regularizing de Sitter spacetime in different ways, for example by considering a big-bang universe filled with two fluid components with the equation-of-state parameters  $w = -1$  and  $w > -1$ , or with a single fluid component of  $w = -1 + \epsilon$ , where  $\epsilon (> 0)$  is taken sufficiently small that the system can be viewed as in a de Sitter vacuum [40,41]. This provides a justification for the description of de Sitter spacetime adopted in Secs. II and III, focusing on a single static patch.<sup>24</sup>

It is important that the holographic description based on a single static patch, which we may call “single-sided” de Sitter spacetime, assumes an appropriate physical regularization. Had we started with exact de Sitter spacetime, then the location of a leaf would be on the bifurcation surface or the future horizon. In this case, a subregion of the leaf would have degenerate extremal surfaces, all of which are located on the future horizon and have areas equal to the volume of the subregion. This would imply that holographic slices sweep only a codimension-1 surface in the bulk, i.e., the future horizon, failing to reconstruct the codimension-0 spacetime [40].<sup>25</sup> For  $\Theta_k = 0$  and  $\Theta_l \neq 0$ , this can also be seen from the fact that  $s^\mu \propto k^\mu$ .

The picture of a single-sided de Sitter spacetime is shown in Fig. 9, in which the region near the bifurcation surface and the past horizon should be viewed as regularization dependent. This picture can also be obtained if we begin with a renormalized holographic screen foliated by

<sup>24</sup>It is also comforting that there is a perturbative positive energy theorem in a static patch of de Sitter spacetime [134].

<sup>25</sup>More precisely, if we use global information of the boundary state, the interior of the static patch may be reconstructed. However, the entanglement wedge of a boundary subregion in this case contains either the entirety or none of the interior, depending on the size of the subregion. This indicates a “singular” nature of the limit of de Sitter spacetime.

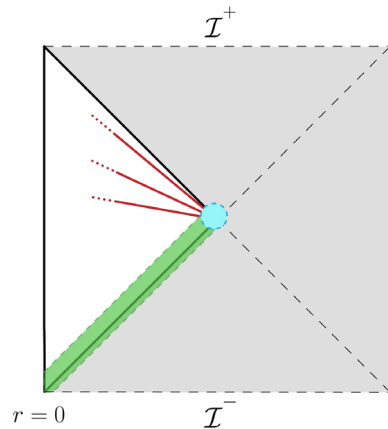


FIG. 9. A holographic description of “single-sided” de Sitter spacetime. Holographic slices agree (approximately) with equal-time hypersurfaces in static coordinates. The smaller the amount of UV renormalization becomes, the closer renormalized leaves are to the bifurcation surface. The geometry in the vicinity of the past horizon (indicated by light green) and the bifurcation surface (light blue) is regularization dependent.

renormalized leaves located deep inside the static patch and then push the screen outward by unrenormalizing it using the flow equation. Renormalized leaves then approach the bifurcation surface (or the future horizon if the assumed deviation from exact de Sitter spacetime is significant), but they never get there. Given the nondecoupling of bulk gravity on these leaves, we expect that the holographic theory on the screen is gravitational, but we do not make any further speculation about this theory here.

### 3. de Sitter entropy

The fact that holographic states in a single-sided de Sitter spacetime are analogous to those in a single-sided black hole suggests that we can interpret the de Sitter entropy in an analogous manner to the black hole case [135]. At the classical level, de Sitter spacetime is parametrized by one continuous number: the Hubble radius  $\alpha$ . At the quantum level, this freedom leads to the corresponding independent quantum states, which are discretized. The number of independent states corresponding to the value of the Hubble radius between  $\alpha$  and  $\alpha + \delta\alpha$  is

$$\mathcal{N} \sim e^{S_{\text{GH}}} \frac{\delta\alpha}{\alpha}, \quad (4.20)$$

where

$$S_{\text{GH}} = \frac{\pi\alpha^2}{l_{\text{P}}^2} \quad (4.21)$$

is the Gibbons-Hawking entropy. As in standard statistical mechanics, this result does not depend on the detailed

choice of  $\delta\alpha$  (unless  $\delta\alpha$  is taken to be exponentially small in  $S_{\text{GH}}$ ).

Now, suppose that the timescale for the evolution of a de Sitter microstate is given by  $\Delta t$ . Using the energy of Eq. (3.16), we find that the required uncertainty of  $\alpha$ , determined by  $\Delta E\Delta t \sim 1$ , is

$$\Delta\alpha \sim \frac{l_{\text{P}}^2}{\Delta t}. \quad (4.22)$$

We expect that  $\Delta t \sim 1/T_{\text{H}} \sim \alpha$  because of large redshift between the location of the stretched horizon, where everything is controlled by the string scale, and the location at which the time  $t$  is measured,  $r = 0$ . This gives

$$\Delta\alpha \sim \frac{l_{\text{P}}^2}{\alpha}. \quad (4.23)$$

In the semiclassical regime  $\alpha \gg l_{\text{P}}$ , this uncertainty is smaller than the Planck length,  $\Delta\alpha \ll l_{\text{P}}$ , which is consistent with the fact that the Hubble radius can be precisely specified (treated classically) in a semiclassical theory.

Note that the situation described above is analogous to the black hole case. For a black hole,  $E = M$  and  $r_+ \sim Ml_{\text{P}}^2$  give  $\Delta r_+ \sim l_{\text{P}}^2/\Delta t$ . Assuming  $\Delta t \sim 1/T_{\text{H}} \sim Ml_{\text{P}}^2$ , this leads to

$$\Delta r_+ \sim \frac{1}{M}. \quad (4.24)$$

In the semiclassical regime  $M \gg 1/l_{\text{P}}$ , we find  $\Delta r_+ \ll l_{\text{P}}$ .

There is one apparent difference between the de Sitter and black hole cases. Since the de Sitter spacetime is an ‘‘inside-out’’ version of the black hole spacetime, simple operators can cause excitations inside the stretched horizon,  $r < r_{\text{s}}$ , and with many such excitations the geometry will be backreacted. In particular, we can form a black hole inside the de Sitter horizon, leading to a different semiclassical geometry. We will now try to understand microscopic entropies of such geometries and their relations to the de Sitter entropy.<sup>26</sup>

We naturally expect that the original theory on the holographic screen can accommodate all such solutions. This interpretation is consistent if we regard a holographic theory of de Sitter spacetime to be associated with a fixed vacuum energy  $\rho_{\Lambda}$  (within uncertainty), and not a fixed horizon radius. In this case, the solution with a black hole has an entropy smaller than the solution with no black hole [39], so the dimension of the Hilbert space of the holographic theory is

$$\dim \mathcal{H} \sim \sum_{\Delta S \leq \Delta S_{\text{max}}} e^{S_{\text{GH}} - \Delta S} \Rightarrow \ln \dim \mathcal{H} \sim S_{\text{GH}} \quad (4.25)$$

<sup>26</sup>A similar issue was also discussed in Ref. [45].

even including all the spacetimes with varying sizes of black holes. Here,  $\Delta S$  is the entropy deficit of a spacetime with a black hole(s),  $S_{\text{GH}}$  is the Gibbons-Hawking entropy expressed in terms of the vacuum energy  $\rho_{\Lambda}$  or the Hubble radius  $\alpha_0$  of the de Sitter spacetime without a black hole

$$S_{\text{GH}} = \frac{\pi\alpha_0^2}{l_{\text{P}}^2} = \frac{3}{8\rho_{\Lambda}l_{\text{P}}^4}, \quad (4.26)$$

and  $\Delta S_{\text{max}}$  is the entropy deficit when the cosmic and black hole horizons have the same radius, which occurs with  $r_{+, \text{bh}} = r_{+, \text{dS}} = \alpha_0/\sqrt{3}$ :

$$\Delta S_{\text{max}} = S_{\text{GH}} - 2 \times \frac{\pi\alpha_0^2}{3l_{\text{P}}^2} = \frac{\pi\alpha_0^2}{3l_{\text{P}}^2} = \frac{1}{8\rho_{\Lambda}l_{\text{P}}^4}, \quad (4.27)$$

where the factor of 2 comes from the fact that we have both cosmological and black hole horizons. The fact that  $\Delta S_{\text{max}}$  is positive serves as a consistency check for our interpretation.<sup>27</sup>

In fact, because of the inside-out nature, forming black holes inside the de Sitter horizon is analogous to forming (small) black holes outside the black hole horizon, specifically in the zone, while keeping the total energy fixed (though the latter necessarily breaks the spherical symmetry). Such excited states can be included in the Hilbert space associated with the central black hole without changing its dimensions,  $\ln \dim \mathcal{H} = S_{\text{bh}}$ , at the leading order.

### C. Holographic Hilbert spaces

Before concluding this section, we discuss the structure of holographic boundary Hilbert spaces in a more general setting. The picture presented here builds on the analyses performed in Refs. [13,40] and is suggested by relations between bulk geometries and boundary entanglement entropies.

<sup>27</sup>In  $d + 1$  dimensions, this occurs when the black hole mass becomes

$$M = \frac{d-1}{d-2} \left( \frac{d-2}{d} \right)^{d/2} \frac{\text{vol}(\Omega_{d-1})}{8\pi l_{\text{P}}^{d-1}} \alpha_0^{d-2} \quad (4.28)$$

with

$$r_{+, \text{bh}} = r_{+, \text{dS}} = \sqrt{\frac{d-2}{d}} \alpha_0. \quad (4.29)$$

Here,  $\text{vol}(\Omega_{d-1}) = 2\pi^{d/2}/\Gamma(d/2)$  is the volume (area) of the  $(d-1)$ -dimensional unit sphere. The maximal entropy deficit is thus

$$\Delta S_{\text{max}} = \frac{\text{vol}(\Omega_{d-1})\alpha_0^{d-1}}{4l_{\text{P}}^{d-1}} \left[ 1 - 2 \times \left( \frac{d-2}{d} \right)^{\frac{d-1}{2}} \right], \quad (4.30)$$

which is indeed positive for all  $d \geq 2$ .

### 1. Effective boundary Hilbert spaces $\mathcal{H}_{\text{eff}}(\mathcal{A})$

Consider a set of states associated with a leaf  $\sigma$  characterized by its volume  $\mathcal{A}$  on the boundary. ( $\mathcal{A}$  is an area from the point of view of bulk spacetime.) Holography implies that the number of degrees of freedom on the leaf is given by

$$\mathcal{N} = \frac{\mathcal{A}}{4G_{\text{N}}}, \quad (4.31)$$

which we will assume to be uniformly distributed over the leaf. Here, we have used  $G_{\text{N}}$  instead of  $l_{\text{p}}$ , since discussion in this subsection does not depend on the number of spacetime dimensions. We denote the Hilbert space comprising states of these  $\mathcal{N}$  degrees of freedom by  $\mathcal{H}_{\text{eff}}(\mathcal{A})$ , so that

$$\ln \dim \mathcal{H}_{\text{eff}}(\mathcal{A}) = \frac{\mathcal{A}}{4G_{\text{N}}}. \quad (4.32)$$

The reason for the subscript ‘‘eff’’ will become clear later.

We expect that states of these  $\mathcal{N}$  degrees of freedom can represent various different spacetimes, or more precisely the domain of dependence  $D_{\sigma}$  of a spacelike hypersurface bounded by  $\sigma$  in these spacetimes. In addition, for a given bulk spacetime, there may be many independent states in  $\mathcal{H}_{\text{eff}}(\mathcal{A})$  that span the space of microstates for the spacetime, as in the case of black hole and de Sitter spacetimes. How can these happen?

Let us assume that the entanglement entropy of subregions of a boundary state  $|\psi\rangle$  dual to a semiclassical geometry can be calculated via the Hubeny-Rangamani-Ryu-Takayanagi (HRRT) prescription [87–89] (or its quantum extension [90–92]). In particular, we assume that the boundary and bulk Hilbert spaces can be appropriately factorized, which may involve a gauge choice or the introduction of edge modes [136–141]. Here we consider the ‘‘classical limit,’’ meaning that all the subregions we consider contain  $O(\mathcal{N})$  degrees of freedom. Given a bulk spacetime, one can then find the corresponding entanglement entropies for all subregions of the boundary. The collection of all boundary subregions and their corresponding entanglement entropies will be referred to as the entanglement structure of the state, which we denote by  $\mathcal{S}(|\psi\rangle)$ .

First, we note that for a given entanglement structure  $\mathcal{S}_0$ , we can always find a basis of the Hilbert space  $\mathcal{H}_{\text{eff}}(\mathcal{A})$  in which all basis states have the specified entanglement structure. This is because by applying local unitaries to a state, one can generate  $e^{\mathcal{N}}$  orthogonal states while preserving the entanglement structure of the original state. This fact, however, does not mean that these  $e^{\mathcal{N}}$  states span an  $e^{\mathcal{N}}$ -dimensional space of microstates for bulk spacetime corresponding to the entanglement structure  $\mathcal{S}_0$ . Indeed, by generically superposing  $e^{O(\mathcal{N})}$  of these states, one would

obtain a state whose entanglement structure is drastically different from  $\mathcal{S}_0$ , so that the resulting state is dual to a completely different spacetime, if it represents bulk spacetime at all.

Of course, given an entanglement structure, there exists a subspace of dimension  $e^{O(\mathcal{N}^p)}$  with  $p < 1$  in which generic states have this same entanglement structure up to  $O(\mathcal{N}^p)$  corrections. This is because we generally have

$$\mathcal{S}\left(\sum_{i=1}^{e^M} c_i |\psi_i\rangle\right) = \mathcal{S}_0 + O(M), \quad (4.33)$$

where  $\mathcal{S}(|\psi_i\rangle) = \mathcal{S}_0$  for all  $i$ , so that for  $M = O(\mathcal{N}^p)$  the corrections are suppressed by powers of  $\mathcal{N}$  compared to  $\mathcal{S}_0$ , which is of order  $\mathcal{N}$ . The subspace obtained in this way, however, comprises only an exponentially small subset of  $\mathcal{H}_{\text{eff}}(\mathcal{A})$ ; in particular, it is a measure zero subset of  $\mathcal{H}_{\text{eff}}(\mathcal{A})$  in the classical limit.

A nontrivial thing is that for  $\mathcal{N} \gg 1$ , there exist subspaces of dimension  $e^{O(\mathcal{N})}$  in which generic states have the same entanglement structure  $\mathcal{S}_0$  up to small corrections. Specifically, for such a subspace spanned by basis states  $|\psi_i\rangle$  ( $i = 1, \dots, e^{Q\mathcal{N}}$ ), we have

$$\mathcal{S}\left(\sum_{i=1}^{e^{Q\mathcal{N}}} c_i |\psi_i\rangle\right) = \mathcal{S}_0 + O(\mathcal{N}^p; p < 1), \quad (4.34)$$

where  $Q \leq 1$  does not scale with  $\mathcal{N}$ . The existence of these subspaces with entanglement structures invariant under superpositions is expected from canonical typicality (also referred to as the general canonical principle) [142,143], which states that generic states in such a subspace have the same reduced density matrix for small subsystems (up to small corrections). This is the case despite the fact that the size of the subspace is large enough that one would naively think that superpositions would ruin the entanglement structure at  $O(\mathcal{N})$ . The proof of this statement is purely kinematical and hence applies generally. In fact, according to canonical typicality the correction term in Eq. (4.34) is exponentially small,  $O(e^{-Q\mathcal{N}/2})$ .

In Fig. 10, we show a sketch of the collection  $\mathcal{C}$  of states in  $\mathcal{H}_{\text{eff}}(\mathcal{A})$  which have the same entanglement structure  $\mathcal{S}_0$  up to corrections higher order in  $1/\mathcal{N}$

$$\mathcal{C} = \{|\psi\rangle | \mathcal{S}(|\psi\rangle) = \mathcal{S}_0 + O(\mathcal{N}^p; p < 1)\} \quad (4.35)$$

in the case that they form a subspace of dimension  $e^{Q\mathcal{N}}$  with  $Q < 1$ . In addition to this subspace represented by the pink plane,  $\mathcal{C}$  contains  $e^{\mathcal{N}} - e^{Q\mathcal{N}}$  states (slightly ‘‘thickened’’ in  $\mathcal{H}_{\text{eff}}(\mathcal{A})$ ) orthogonal to it, which are schematically represented by the red arrow. Furthermore, in the  $e^{Q\mathcal{N}}$ -dimensional subspace, there are exponentially rare states that do not have the entanglement structure  $\mathcal{S}_0$ . These states can be obtained by fine-tuning coefficients when we

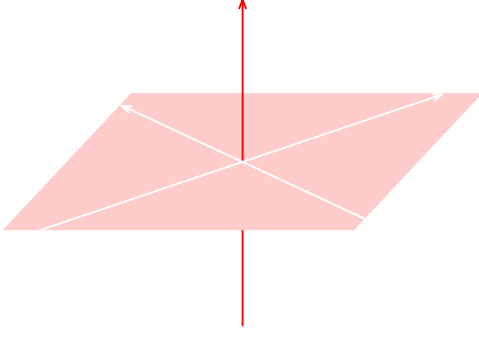


FIG. 10. A sketch of the collection of states in holographic Hilbert space  $\mathcal{H}_{\text{eff}}(\mathcal{A})$  which have the same entanglement structure at leading order in  $1/\mathcal{N}$ . It forms an  $e^{Q\mathcal{N}}$ -dimensional subspace of  $\mathcal{H}_{\text{eff}}(\mathcal{A})$  (represented by the pink plane) except that exponentially rare nongeneric states are excluded (white arrows), and that  $e^{\mathcal{N}} - e^{Q\mathcal{N}}$  isolated states orthogonal to it are added (red arrow).

expand  $|\psi\rangle$  in terms of the basis states  $|\psi_i\rangle$  of the subspace and are represented by white arrows.

In Ref. [40], it was argued that it is this  $e^{Q\mathcal{N}}$ -dimensional subspace that comprises the space of microstates for a spacetime. In particular, for  $Q < 1$  the corresponding entanglement structure  $S_0$  can be nonmaximal,<sup>28</sup> and generic states in this subspace can have dual bulk spacetimes which are simply reconstructable. In this case, even if one considers an exponentially large superposition of microstates, geometric operators are effectively linear so long as the state is generic within the subspace [93]. On the other hand, if  $Q = 1$ , the “subspace” is the whole Hilbert space  $\mathcal{H}_{\text{eff}}(\mathcal{A})$ , so applying Page’s analysis [144], we see that the only entanglement structure consistent with Eq. (4.34) is that of maximal entropy. In this case, the resulting spacetime is not in a simple wedge, and reconstruction of the bulk requires some level of nonlinearity, or state dependence.<sup>29</sup>

The structure discussed above allows for a single holographic Hilbert space  $\mathcal{H}_{\text{eff}}(\mathcal{A})$  to harbor effective subspaces dual to different geometries. In fact, similarly to Eq. (4.25), one can show that  $\mathcal{H}_{\text{eff}}(\mathcal{A})$  satisfying Eq. (4.32) can support a number of  $e^{Q\mathcal{N}}$ -dimensional subspaces with  $Q < 1$ . Geometric operators are approximately linear in each of these subspaces, which gives the effective linear space of microstates for a fixed semiclassical geometry.

## 2. Fundamental boundary Hilbert space $\mathcal{H}_{\text{UV}}$

So far, we have considered boundary Hilbert space with fixed volume  $\mathcal{A}$ :  $\mathcal{H}_{\text{eff}}(\mathcal{A})$ . However, general spacetime

<sup>28</sup>By the entanglement structure being maximal, we mean that for any subregion  $A$  its entanglement entropy  $S_A$  is maximal, i.e.,  $S_A = \|A\|/4G_{\text{N}}$ , at the leading order in  $1/\|A\|$ . Here  $\|A\|$  is the volume of  $A$  on the boundary.

<sup>29</sup>This is indeed the case for the interior of a black hole; see Sec. VA.

involves boundary evolution in which the volume of a leaf (a boundary equal-time surface) changes [42]. Such an evolution can occur in the “fundamental” Hilbert space  $\mathcal{H}_{\text{UV}}$  which contains, at least effectively,  $\mathcal{H}_{\text{eff}}(\mathcal{A})$ ’s with different  $\mathcal{A}$ ’s<sup>30</sup>:

$$\mathcal{H}_{\text{UV}} \supset \{\mathcal{H}_{\text{eff}}(\mathcal{A}) \mid \dim \mathcal{H}_{\text{eff}}(\mathcal{A}) \in \mathbb{N}\}. \quad (4.36)$$

A naive possibility is to literally have  $\mathcal{H}_{\text{UV}} \supset \bigoplus_{\mathcal{A}} \mathcal{H}_{\text{eff}}(\mathcal{A})$ , but this need not be the case. In fact, motivated by the relation between geometric objects in the bulk and quantum information theoretic quantities on the boundary, which has been learned in AdS/CFT and is expected to apply beyond, one can imagine that  $\mathcal{H}_{\text{eff}}(\mathcal{A})$ ’s are contained in  $\mathcal{H}_{\text{UV}}$  in a more intricate manner.

In Refs. [13,112], it was envisioned that  $\mathcal{H}_{\text{eff}}(\mathcal{A})$  is embedded in  $\mathcal{H}_{\text{UV}}$  as an effective subspace (in the sense discussed before; see, e.g., Fig. 10) in which a generic state has the property

$$\sum_i S_i = \frac{\mathcal{A}}{4G_{\text{N}}}, \quad (4.37)$$

where  $S_i$  represents the entanglement entropy of the state in a sufficiently small subregion,  $A_i$ , of the holographic space  $\Omega$ , on which  $\mathcal{H}_{\text{UV}}$  is defined; the sum runs over all of these small subregions such that  $\Omega = \cup_i A_i$  and  $A_i \cap A_j = \emptyset$  ( $i \neq j$ ). This allows us to consider holographic states of all bulk spacetimes, and also their dynamics, in a single Hilbert space  $\mathcal{H}_{\text{UV}}$ .

The Hilbert space  $\mathcal{H}_{\text{UV}}$  will be defined by introducing a short distance cutoff  $\delta$  in  $\Omega$  and then sending  $\delta$  to zero, so that

$$\dim \mathcal{H}_{\text{UV}} \rightarrow \infty. \quad (4.38)$$

For  $(d+1)$ -dimensional asymptotically AdS and flat bulk spacetimes,  $S_i$  will behave as

$$S_i \sim \frac{\|\partial A_i\|}{\delta^{d-2}} \quad \text{and} \quad S_i = f(A_i) \frac{\|A_i\|}{\delta^{d-1}}, \quad (4.39)$$

respectively. Here,  $\|x\|$  represents the volume of the object  $x$ ,  $\partial A_i$  is the boundary of  $A_i$ , and  $f(A_i)$  is a function of  $O(1)$ .<sup>31</sup>

According to this picture, states representing a cosmological spacetime with finite leaf area comprise a tiny effective subspace of  $\mathcal{H}_{\text{UV}}$ , obtained from a generic state in  $\mathcal{H}_{\text{UV}}$  by an infinite number (in the limit  $\delta \rightarrow 0$ )

<sup>30</sup>A semiclassical description is valid only when  $\dim \mathcal{H}_{\text{eff}}(\mathcal{A}) \gg 1$ , but we will be sloppy about it in writing Eq. (4.36).

<sup>31</sup>In asymptotically flat spacetime, for example, the ratio of  $f(A_i)$  for  $A_i$  being a half of  $\Omega$ ,  $A_{1/2}$ , to that for  $A_i$  being a cutoff size region,  $A_\delta$ , is  $f(A_{1/2})/f(A_\delta) = (1/\sqrt{\pi})\{\Gamma[d/2]/\Gamma[(d+1)/2]\}$ .



of fine-tunings.<sup>32</sup> It will be interesting to study the dynamics in  $\mathcal{H}_{UV}$  in the thermodynamic limit, using information from semiclassical theory.

## V. ANALYTIC EXTENSION OF SPACETIME IN QUANTUM GRAVITY

In a single-sided system, the spacetime region obtained from analytic extension in general relativity—e.g., the interior of a black hole and the region outside a static patch in de Sitter spacetime—emerges as a collective phenomenon involving horizon (and possibly other) modes. This can occur because a huge gravitational red/blueshift at the stretched horizon makes the string dynamics relevant in a static description, which makes the state take a generic, universal form across all low energy species. In this section, we review this construction [21–25] and refine it to include the effect of black hole evolution analyzed in Sec. III C.

### A. The interior of a black hole

Consider a state of a black hole of mass  $M$  at some time  $t = t_*$ . We assume that it is in the semiclassical vacuum state, which is achieved typically more than one scrambling time  $t_{\text{scr}}$  after the last perturbation, where<sup>33</sup>

$$t_{\text{scr}} = 2r_+ \left[ \ln \frac{r_+}{l_p} + O(1) \right]. \quad (5.3)$$

As discussed in Sec. III A, we assume that the mass of the black hole (as well as other quantities such as the momentum) is specified with the maximal precision allowed by the uncertainty principle,  $\Delta M \sim T_H$ . If the state of the system involves a superposition of a black hole of a wider range of masses, then our consideration below

<sup>32</sup>A similar picture was discussed for de Sitter spacetime in Ref. [145].

<sup>33</sup>This expression is obtained from the fact that the scrambling time represented in ingoing Eddington-Finkelstein time  $v = t + r_*$  is [33]

$$v_{\text{scr}} = \frac{1}{2\pi T_H} \ln S_{\text{bh}} \approx 4r_+ \ln \frac{r_+}{l_p}. \quad (5.1)$$

The scrambling time in Schwarzschild time is then related to this time by [24]

$$t_{\text{scr}} = v_{\text{scr}} + r_{*s} \approx 2r_+ \ln \frac{r_+}{l_p}, \quad (5.2)$$

where the  $r_{*s}$  term in the middle expression comes from the fact that the scrambling time in Eq. (5.1) is defined as the minimal time needed to recover information about an object falling into the stretched horizon (at  $r_* = r_{*s}$ ) at a location sufficiently far from the horizon,  $r_* \approx O(r_+)$ . In the last equation, we have used  $r_{*s} \approx -2r_+ \ln(r_+/l_p)$ , obtained from Eq. (2.17) by identifying  $l_s$  with  $l_p$ .

applies to each of the branches containing a black hole with minimal uncertainties.<sup>34</sup>

Ignoring evolution effects for now, a black hole vacuum state is given as in Eq. (3.40). This state picks out a set of special states of the combined system of soft and far modes: those multiplied by the hard-mode states  $|\{n_\alpha\}\rangle$  in Eq. (3.40). We denote these states using the double-ket symbol

$$|\{n_\alpha\}_A\rangle\rangle = \zeta_n^A \sum_{i_n=1}^{e^{S_{\text{bh}}(M-E_n)}} \sum_{a=1}^{e^{S_{\text{rad}}}} c_{ni_n a}^A |\psi_{i_n}^{(n)}\rangle |\phi_a\rangle. \quad (5.4)$$

Here,  $A$  is the index for the microstate, specified by the coefficients  $c_{ni_n a}^A$  in Eq. (3.40), and  $\zeta_n^A$  is the normalization constant

$$\zeta_n^A = \frac{1}{\sqrt{\sum_{i_n=1}^{e^{S_{\text{bh}}(M-E_n)}} \sum_{a=1}^{e^{S_{\text{rad}}}} c_{ni_n a}^{A*} c_{ni_n a}^A}} = \sqrt{z} e^{\frac{E_n}{2T_H}} \left( 1 - \frac{1}{2} \tilde{\epsilon}_n^{AA} \right), \quad (5.5)$$

where  $z$  is given by Eq. (3.31). The last expression is obtained by using statistical properties of  $c_{ni_n a}^A$  in Eq. (3.42), and  $\tilde{\epsilon}_n^{AB}$  is the quantity analogous to  $\epsilon_n^{AB}$  in Eq. (3.38):

$$\begin{aligned} \tilde{\epsilon}_n^{AB} &\equiv z e^{\frac{E_n}{T_H}} \sum_{i_n=1}^{e^{S_{\text{bh}}(M-E_n)}} \sum_{a=1}^{e^{S_{\text{rad}}}} c_{ni_n a}^{A*} c_{ni_n a}^B - \delta_{AB} \\ &\approx O\left(e^{-\frac{1}{2}\{S_{\text{bh}}(M-E_n)+S_{\text{rad}}\}}\right), \end{aligned} \quad (5.6)$$

which satisfies

$$(\tilde{\epsilon}_n^{AB})^* = \tilde{\epsilon}_n^{BA}, \quad \sum_n \frac{e^{-\frac{E_n}{T_H}}}{z} \tilde{\epsilon}_n^{AB} = 0. \quad (5.7)$$

Substituting Eq. (5.4) into Eq. (3.40), we see that the state can be written in the thermofield double form

$$|\Psi_A(M)\rangle\rangle = \frac{1}{\sqrt{z}} \sum_n e^{-\frac{E_n}{2T_H}} |\{n_\alpha\}\rangle |\{n_\alpha\}_A\rangle\rangle, \quad (5.8)$$

up to exponentially suppressed corrections of order  $e^{-\{S_{\text{bh}}(M)+S_{\text{rad}}\}/2}$ . We can also check that the states  $|\{n_\alpha\}_A\rangle\rangle$  are orthonormal in  $\{n_\alpha\}$  as well as the microstate index  $A$ , up to exponentially small corrections [25]:

$$\langle\langle \{m_\alpha\}_A | \{n_\alpha\}_B \rangle\rangle = \delta_{mn} \eta_n^{AB}, \quad (5.9)$$

<sup>34</sup>In practice, different branches decohere with environment, so focusing on a single branch is phenomenologically forced on us when we discuss the physics of the black hole itself, such as its interior, using the pure state language. Of course, when discussing more global aspects, such as the full unitarity of a black hole formation and evaporation process, we need to take into account all these branches [83,84]. However, the effect from such a superposition, i.e., a superposition of “macroscopically distinguishable” black holes, is subdominant in entropic consideration, compared with the Bekenstein-Hawking entropy associated with a black hole with minimal uncertainties.

where

$$n_n^{AB} \equiv \begin{cases} 1 & \text{for } A = B \\ \tilde{\epsilon}_n^{AB} \approx O(e^{-\frac{1}{2}\{S_{\text{bh}}(M-E_n)+S_{\text{rad}}\}}) & \text{for } A \neq B. \end{cases} \quad (5.10)$$

Since the spectrum of the states  $|\{n_\alpha\}\rangle$  represents semiclassical physics in the zone of the single-sided black hole, the corresponding states  $|\{n_\alpha\}_A\rangle\rangle$  of the soft and far modes can be identified as the states in the second exterior of an effective two-sided black hole, given by Eq. (5.8).

With this identification, one can construct annihilation and creation operators for modes in the second exterior

$$\tilde{b}_\gamma^A = \sum_n \sqrt{n_\gamma} \|\{n_\alpha - \delta_{\alpha\gamma}\}_A\rangle\rangle \langle\langle \{n_\alpha\}_A \| \quad (5.11)$$

$$\begin{aligned} &= \sum_n \sqrt{n_\gamma} \zeta_{n_\gamma}^A \zeta_n^{A*} \sum_{i_{n-\gamma}=1}^{e^{S_{\text{bh}}(M-E_{n-\gamma})}} \sum_{j_n=1}^{e^{S_{\text{bh}}(M-E_n)}} \sum_{a,b=1}^{e^{S_{\text{rad}}}} c_{n_\gamma i_{n-\gamma} a}^A \\ &\quad \times c_{n_{j_n b}^{A*}} |\psi_{i_{n-\gamma}}^{(n-\gamma)}\rangle |\phi_a\rangle \langle \psi_{j_n}^{(n)} | \langle \phi_b|, \end{aligned} \quad (5.12)$$

$$\tilde{b}_\gamma^{A\dagger} = \sum_n \sqrt{n_\gamma + 1} \|\{n_\alpha + \delta_{\alpha\gamma}\}_A\rangle\rangle \langle\langle \{n_\alpha\}_A \| \quad (5.13)$$

$$\begin{aligned} &= \sum_n \sqrt{n_\gamma + 1} \zeta_{n_\gamma}^A \zeta_n^{A*} \sum_{i_{n+\gamma}=1}^{e^{S_{\text{bh}}(M-E_{n+\gamma})}} \sum_{j_n=1}^{e^{S_{\text{bh}}(M-E_n)}} \sum_{a,b=1}^{e^{S_{\text{rad}}}} c_{n_\gamma i_{n+\gamma} a}^A \\ &\quad \times c_{n_{j_n b}^{A*}} |\psi_{i_{n+\gamma}}^{(n+\gamma)}\rangle |\phi_a\rangle \langle \psi_{j_n}^{(n)} | \langle \phi_b|, \end{aligned} \quad (5.14)$$

in addition to those in the first exterior, Eqs. (3.49) and (3.50). Here,  $n_{\pm\gamma} \equiv \{n_\alpha \pm \delta_{\alpha\gamma}\}$ , and we have used the same symbol  $\gamma$  to specify both the first and second exterior modes. From these operators, one can then construct annihilation and creation operators for interior modes, as well as the infalling time evolution operator, through an appropriate Bogoliubov transformation [21–25].<sup>35</sup> We will see this construction more explicitly below for an evaporating black hole.

<sup>35</sup>This construction differs from a similar construction in Refs. [26,31,32], in which the degrees of freedom that are identified as those in the first exterior of the effective two-sided black hole increase as the black hole evaporates; in particular, Hawking radiation emitted earlier composes degrees of freedom in the first exterior. In the construction here, the number of degrees of freedom composing the first exterior (zone modes) decreases as the evaporation progresses; in particular, early Hawking radiation is identified as a part of the degrees of freedom in the second exterior. The idea that the construction of interior operators involves early Hawking radiation was promoted in Ref. [29], but its specific realization is different here. In contrast to the picture laid out in Ref. [29], the second exterior of the effective two-sided geometry arises primarily from degrees of freedom directly associated with the black hole (soft modes), and the involvement of Hawking radiation is indirect (although it is significant for an old black hole, i.e., a black hole that is nearly maximally entangled with the rest of the system). In particular, the structure of entanglement is not bipartite between the first exterior and early Hawking radiation degrees of freedom as envisioned in Ref. [29].

In this picture, the second exterior of the black hole emerges effectively as a collective phenomenon associated with the soft and far modes. Note that while the energy of each soft or early Hawking mode may be tiny,  $\lesssim T_H$ , their collective excitation can have a much larger energy,  $\gg T_H$ . In particular, a “quasiparticle” created by  $\tilde{b}_\gamma^{A\dagger}$  in the second exterior has energy  $E_\gamma$ , which is negative and  $|E_\gamma| > T_H$ . The fact that this energy is negative implies that the Hamiltonian  $H$  of the original single-sided black hole is mapped to the generator of the timelike isometry in the effective two-sided picture

$$H \mapsto H_R - H_L, \quad (5.15)$$

where  $H_R$  and  $H_L$  are the Hamiltonian operators acting on the first and second exteriors, respectively. As we will see below, excitation in the black hole interior is a superposition of quasiparticle excitations built by  $\tilde{b}_\gamma^A$ 's and  $\tilde{b}_\gamma^{A\dagger}$ 's and original excitations in the zone.

It is worth mentioning that we have no freedom in choosing the basis in the space of  $|\{n_\alpha\}_A\rangle\rangle$ 's. In other words, the interpretation that  $|\{n_\alpha\}_A\rangle\rangle$  is a state in which  $n_\alpha$  of the mode corresponding to  $\alpha$  in the first exterior is excited in the second exterior, is not invariant under unitary transformation

$$|\{n_\alpha\}_A\rangle\rangle \rightarrow |\{n'_\alpha\}_A\rangle\rangle = \sum_{n'} U_{nn'} |\{n'_\alpha\}_A\rangle\rangle, \quad (5.16)$$

so that a state obtained by acting  $\tilde{b}_\gamma^A$  or  $\tilde{b}_\gamma^{A\dagger}$  on  $|\Psi_A(M)\rangle\rangle$  in Eq. (3.40) cannot be regarded as the vacuum state at the semiclassical level. This is because the minimal uncertainty condition imposed on  $|\Psi_A(M)\rangle\rangle$  allows us to physically distinguish between different  $|\{n_\alpha\}\rangle\rangle$ 's, and hence there is no ambiguity in defining the corresponding  $|\{n_\alpha\}_A\rangle\rangle$ 's. Our framework, therefore, does not suffer from the “frozen vacuum” problem posed in Ref. [146].

We also comment that  $|\phi_a\rangle$  in Eq. (5.4) may not consist of only the early Hawking radiation; if a major portion of the radiation interacts with other degrees of freedom, such degrees of freedom must also be included in  $|\phi_a\rangle$ . This is because  $|\phi_a\rangle$ 's represent states of all the degrees of freedom that are entangled significantly with the hard and soft modes. It follows that the degrees of freedom that were once the second-exterior degrees of freedom keep playing that role even if they no longer take the form of Hawking radiation. This situation, however, does not last forever. The role as a second exterior emerges only relationally with respect to the hard modes. Thus, after the black hole is evaporated completely, the Hawking radiation as well as any other degrees of freedom that have interacted with it become regular matter that does not have any spacetime interpretation.

Finally, we emphasize that the construction described here is performed in a theory with gravity in the bulk,

implying that the precise spectrum of the horizon and zone modes (represented by the boundary conditions on the stretched horizon in the low energy field theory) are not arbitrary; rather, they are determined by a consistent UV theory. In other words, the construction here should be viewed as being performed in a holographic boundary theory (see, e.g., footnote below). The physics in a non-gravitational bulk can be reproduced if we send  $l_p \rightarrow 0$  (with fixed  $r_+ = 2Ml_p^2$ ), which implies  $l_s \rightarrow 0$  because of Eq. (3.12). The change from the distant description to the infalling description considered here is then reduced to the change of reference frames in quantum field theory (from the Rindler to Minkowski frame for  $r_+ \rightarrow \infty$ ).

### 1. Effective theory of the interior of an evaporating black hole

As we have seen in Sec. III C, for an evaporating black hole some of the ingoing zone modes are not excited in the black hole vacuum; see, e.g., Eq. (3.44) and Fig. 4. Strictly speaking, however, the coefficients of the terms involving zone-mode states in which these modes are excited are not exactly zero (though they are exponentially suppressed); see footnote 16. We can therefore define the double-ket states as in Eq. (5.4) for all  $\{n_\alpha\}$  and construct annihilation and creation operators, Eqs. (5.11)–(5.14), for the second exterior mode corresponding to any hard mode  $\gamma$  in the first exterior.

Nevertheless, it is true that in the black hole vacuum state, the hard-mode states in which these ingoing modes are excited do not have sizable coefficients. We thus have to use Eq. (5.4) in Eq. (3.44) instead of Eq. (3.40), and we obtain

$$|\Psi_A(M)\rangle \propto \prod_{\tilde{\alpha} \in \substack{\text{ingoing} \\ m_\Phi \lesssim T_H, \omega \sim T_H}} b_{\tilde{\alpha}} \sum_n \prod_{\substack{\alpha' \in \substack{\text{ingoing} \\ \omega > \omega_{\text{barrier}}^{(\ell)}}}} \delta_{n_{\alpha'}, 0} e^{-\frac{E_n}{2T_H}} \times |\{n_\alpha\}\rangle |\{n_\alpha\}_A\rangle, \quad (5.17)$$

up to exponentially small corrections. We thus find that the ingoing modes  $\alpha'$  with the frequency larger than the barrier height are virtually not entangled with the corresponding second exterior modes in the vacuum microstate.

This implies that the spacetime is not smooth across the future horizon of the second exterior, or the past horizon of the first exterior, since the missing entanglement is essential for the connectedness of the spacetime there. The spacetime region which the effective theory built on the state in Eq. (5.17) describes is thus the shaded region in Fig. 11, which we refer to as region  $K$ . Here,  $U$  and  $V$  are the Kruskal-Szekeres coordinates erected at  $t = t_*$ :

$$ds^2 = -dUdV + r^2 d\Omega_{d-1}^2, \quad (5.18)$$

which are given in the near horizon region by

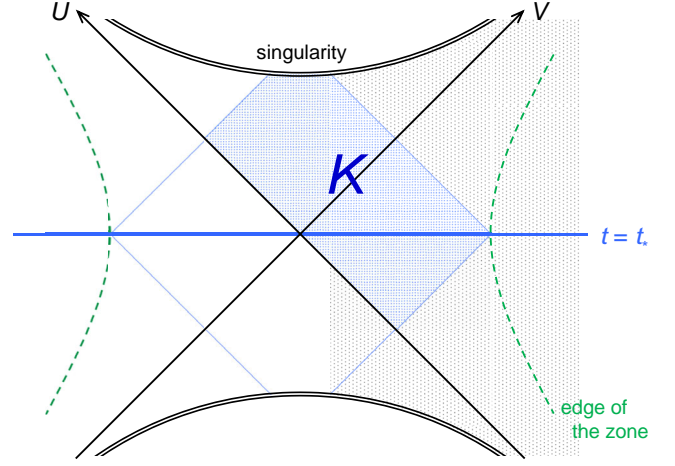


FIG. 11. The spacetime region  $K$  described by the effective theory of the interior erected at a boundary (Schwarzschild) time  $t = t_*$ .

$$\begin{cases} U = -Re^{-\omega} \\ V = Re^{\omega}, \end{cases} \quad \begin{cases} U = Re^{-\omega} \\ V = Re^{\omega}, \end{cases} \\ \begin{cases} U = Re^{-\omega} \\ V = -Re^{\omega}, \end{cases} \quad \begin{cases} U = -Re^{-\omega} \\ V = -Re^{\omega} \end{cases} \quad (5.19)$$

for Region I ( $U < 0, V > 0$ ), Region II ( $U, V > 0$ ), Region III ( $U > 0, V < 0$ ), and Region IV ( $U, V < 0$ ), respectively, with

$$R = 2\sqrt{r_+|r - r_+|}, \quad \omega = \frac{1}{2r_+}(t - t_*). \quad (5.20)$$

With the state given by Eq. (5.17), the interior of the black hole is in the semiclassical vacuum in the infalling frame.

We now study in more detail how the interior region is described in this effective theory. Let us first take the Schrödinger picture (as we have implicitly been doing). We want to understand what an object located in the zone and falling toward the black hole will experience as it crosses the horizon. An excited state with  $N$  particles in the zone can be obtained by applying appropriate superpositions of creation operators  $b_\gamma^\dagger$  on a black hole vacuum state:

$$|\Psi(t = t_*)\rangle \equiv \prod_{i=1}^N \left( \sum_{\gamma} f_\gamma^{(i)} b_\gamma^\dagger \right) |\Psi_A(M)\rangle, \quad (5.21)$$

where  $|\Psi_A(M)\rangle$  is the black hole vacuum microstate in Eq. (3.44), and  $f^{(i)}$  is the wave function of the  $i$ -th particle represented in the  $\gamma$  space. This state can be straightforwardly mapped to that in the effective theory—we simply have to take  $|\Psi_A(M)\rangle$  to be that in Eq. (5.17) instead of Eq. (3.44). We denote this state in the effective theory by  $|\Psi(\tau = 0)\rangle$ :

$$|\Psi(t = t_*)\rangle \mapsto |\Psi(\tau = 0)\rangle, \quad (5.22)$$

by choosing the origin of the infalling time  $\tau$  to match the boundary time  $t = t_*$ .

The operators used to interpret a state in the effective theory is given by  $b_\gamma$  and  $b_\gamma^\dagger$  in Eqs. (3.49) and (3.50) and  $\tilde{b}_\gamma^A$  and  $\tilde{b}_\gamma^{A\dagger}$  in Eqs. (5.11) and (5.13). In particular, we can form infalling mode operators out of superpositions of these operators

$$a_\xi^A = \sum_\gamma (\alpha_{\xi\gamma} b_\gamma + \beta_{\xi\gamma} b_\gamma^\dagger + \zeta_{\xi\gamma} \tilde{b}_\gamma^A + \eta_{\xi\gamma} \tilde{b}_\gamma^{A\dagger}), \quad (5.23)$$

$$a_\xi^{A\dagger} = \sum_\gamma (\beta_{\xi\gamma}^* b_\gamma + \alpha_{\xi\gamma}^* b_\gamma^\dagger + \eta_{\xi\gamma}^* \tilde{b}_\gamma^A + \zeta_{\xi\gamma}^* \tilde{b}_\gamma^{A\dagger}), \quad (5.24)$$

where  $\xi$  is the label in which the frequency  $\omega$  with respect to boundary time  $t$  is traded with the frequency  $\Omega$  associated with infalling time  $\tau$ , and  $\alpha_{\xi\gamma}$ ,  $\beta_{\xi\gamma}$ ,  $\zeta_{\xi\gamma}$ , and  $\eta_{\xi\gamma}$  are the Bogoliubov coefficients calculable using the standard field theory method [54,55].<sup>36</sup>

To obtain the generator of infalling time evolution  $\tilde{H}$ , we start from the generator of boundary time evolution

$$H = H_{\text{soft}} + \sum_\gamma \omega_\gamma b_\gamma^\dagger b_\gamma + H_{\text{int}}(\{b_\gamma\}, \{b_\gamma^\dagger\}), \quad (5.26)$$

where  $H_{\text{soft}}$  gives time evolution of the soft modes as well as interactions between the hard and soft modes. The infalling time evolution generator is then given by

$$\tilde{H} = \sum_\xi \Omega_\xi a_\xi^{A\dagger} a_\xi^A + \tilde{H}_{\text{int}}(\{a_\xi^A\}, \{a_\xi^{A\dagger}\}), \quad (5.27)$$

where  $\tilde{H}_{\text{int}}(\{a_\xi^A\}, \{a_\xi^{A\dagger}\})$  is determined by matching it with  $H_{\text{int}}(\{b_\gamma\}, \{b_\gamma^\dagger\})$ . The state in the effective theory evolves as

$$|\Psi(\tau)\rangle = e^{-i\tilde{H}\tau} |\Psi(\tau = 0)\rangle. \quad (5.28)$$

<sup>36</sup>For a massless scalar field, for example, Eq. (5.23) takes the form

$$a_\xi^A = \pm \frac{i}{2\pi\sqrt{\Omega_\xi T_H}} \int_0^\infty d\omega_\gamma \left[ \frac{\Xi}{\sqrt{1 - e^{-\frac{\omega_\gamma}{T_H}}}} b_\gamma + \frac{\Xi^*}{\sqrt{e^{\frac{\omega_\gamma}{T_H}} - 1}} b_\gamma^\dagger - \frac{\Xi^*}{\sqrt{1 - e^{-\frac{\omega_\gamma}{T_H}}}} \tilde{b}_\gamma^A - \frac{\Xi}{\sqrt{e^{\frac{\omega_\gamma}{T_H}} - 1}} \tilde{b}_\gamma^{A\dagger} \right] \quad (5.25)$$

in the near horizon limit. Here, we have adopted the continuum notation for the sum over the frequency, and  $\Xi = (\Omega_\xi/2\pi T_H)^{\pm \frac{i\omega_\gamma}{2\pi T_H}} \Gamma(1 \pm \frac{\omega_\gamma}{2\pi i T_H}) / |\Gamma(1 \pm \frac{\omega_\gamma}{2\pi i T_H})|$  is a pure phase. The  $\pm$  symbol in these equations takes  $+$  and  $-$  for ingoing and outgoing modes, respectively.

Note that the generator in Eq. (5.27) contains a term that involves both  $b_\gamma$  and  $\tilde{b}_\gamma^A$  operators, e.g.,  $b_\gamma^\dagger \tilde{b}_\gamma^A$ , even in the free part. Thus, it acts on the hard and soft (and far) modes simultaneously, as can be seen in Eqs. (5.12) and (5.14).

The effective theory is defined with the initial state on  $\Sigma$ , the zone at  $t = t_*$ , and its mirror in the second exterior  $\tilde{\Sigma}$ , and the spacetime is not smoothly connected across the  $U$  axis (or formally has an infinite energy shock wave along the  $U$  axis). We can, therefore, trust results obtained using the theory only in the region

$$K = D(\Sigma \cup \tilde{\Sigma}) \cap \{(U, V) | V > 0\}, \quad (5.29)$$

shown in Fig. 11, where  $D(X)$  represents the domain of dependence of  $X$ . This is, however, all we need to describe the fate of the fallen object. Note that here we consider the fate of an object falling into a black hole in its semiclassical vacuum. Also, if an object leaves the region  $D(\Sigma \cup \tilde{\Sigma})$  before hitting the singularity, we can choose to erect an effective theory at a different time such that the entire trajectory of the object (until it hits the singularity) is contained in the region described by the effective theory; such a choice is always possible [21].

The description of the interior which is more along the lines of a conventional quantum field theory treatment can be obtained by adopting the Heisenberg picture [25]. With the operators in Eqs. (5.23) and (5.24), quantum field operators at  $\tau = 0$  are given by

$$\tilde{\Phi}_\Gamma(\mathbf{x}, 0) = \sum_{s, \Omega, \mathbf{L}} (f_s(\Omega, \mathbf{L}) \varphi_{\Omega, \mathbf{L}}(\mathbf{x}) a_\xi^A + g_s(\Omega, \mathbf{L}) \varphi_{\Omega, \mathbf{L}}^*(\mathbf{x}) a_{\xi^c}^{A\dagger}), \quad (5.30)$$

where we have decomposed index  $\xi$  into  $\Gamma$ ,  $s$ ,  $\Omega$ , and  $\mathbf{L}$  which represent species, spin, frequency, and orbital angular momentum quantum numbers, respectively, with  $\xi^c$  representing the  $CPT$  conjugate of  $\xi$ . Here,  $f_s(\Omega, \mathbf{L})$  and  $g_s(\Omega, \mathbf{L})$  are the standard factors providing Lorentz representation of the field (Dirac spinors, polarization vectors, etc), and  $\varphi_{\Omega, \mathbf{L}}(\mathbf{x})$  are the spatial wave functions, determined by matching  $\tilde{\Phi}_\Gamma(\mathbf{x}, 0)$  with quantum field operators of the original theory at  $t = t_*$ .<sup>37</sup>

The Heisenberg picture field operators can then be defined as

$$\tilde{\Phi}_\Gamma(\mathbf{x}, \tau) = e^{i\tilde{H}\tau} \tilde{\Phi}_\Gamma(\mathbf{x}, 0) e^{-i\tilde{H}\tau}, \quad (5.31)$$

where  $\tilde{H}$  is given in Eq. (5.27). The quantities we are interested in are the correlators

<sup>37</sup>This match requires information of the coefficients in Eqs. (5.23) and (5.24).



$$\begin{aligned} & \langle \tilde{\Phi}_{\Gamma_1}(x_1) \tilde{\Phi}_{\Gamma_2}(x_2) \cdots \tilde{\Phi}_{\Gamma_n}(x_n) \rangle \\ & = \text{Tr}[\tilde{\rho}(0) \tilde{\Phi}_{\Gamma_1}(x_1) \tilde{\Phi}_{\Gamma_2}(x_2) \cdots \tilde{\Phi}_{\Gamma_n}(x_n)], \end{aligned} \quad (5.32)$$

where  $x_i = \{\mathbf{x}_i, \tau_i\}$ , and  $\tilde{\rho}(0) = |\Psi(\tau=0)\rangle\langle\Psi(\tau=0)|$ . Since these are expectation values in the state at a fixed time  $\tau=0$ , we must adopt the in-in formalism rather than the more conventional in-out formalism to calculate them. This ultimately comes from the fact that the  $S$ -matrix cannot be defined at the semiclassical level for an object falling into a black hole.

Using the Schwinger-Keldysh method, Eq. (5.32) can be written as a path integral over an appropriate closed time contour with the boundary condition given by  $\tilde{\rho}(0)$ . We can also calculate it using perturbation theory in the canonical in-in formalism. Note that fields in Eq. (5.32) need not be in the interior of the black hole; they only need to be in the region  $K$ . We can thus compute correlators between fields inside and outside the horizon using Eq. (5.32).

In the construction of the effective theory described so far, we have used an input from semiclassical theory to determine the coefficients in Eqs. (5.23) and (5.24). However, we expect that this is ultimately not needed. In particular, we expect that these coefficients are determined (though not uniquely) by the requirement that the generator  $\tilde{H}$  can be written in the local form in terms of the original quantum fields [22]; namely, when  $a_\xi^A$ 's in  $\tilde{H}$  are represented by  $b_\gamma$ 's and  $\tilde{b}_\gamma^{A'}$ 's which in turn are represented by quantum fields and their canonical conjugates,  $\tilde{H}$  takes a local form in the first and emergent second exteriors. To find the coefficients directly from the boundary theory, one possibility is to use a physical probe to construct the infalling Hamiltonian, along the lines of Refs. [147,148]. We leave a detailed study of these issues for the future.

## 2. State dependence and intrinsic ambiguity

The operators in the effective theory constructed so far,  $\tilde{b}_\gamma^A$ ,  $\tilde{b}_\gamma^{A'}$ ,  $a_\xi^A$ , and  $a_\xi^{A'}$ , depended on the vacuum microstate, indexed by  $A$ , on which the excited states are built. This dependence, however, can be relaxed in such a way that a single set of operators can describe an object in the zone even if it is entangled arbitrarily with the vacuum microstates [25,30] (see also Ref. [32]).

The Hilbert space spanned by all the independent vacuum microstates of a black hole of mass  $M$  is given by

$$\mathcal{M} = \left\{ \sum_{A=1}^{e^{S_{\text{tot}}}} a_A |\Psi_A(M)\rangle \middle| a_A \in \mathbb{C}, \sum_{A=1}^{e^{S_{\text{tot}}}} |a_A|^2 = 1 \right\}, \quad (5.33)$$

where  $S_{\text{tot}}$  is given by Eq. (3.46). We consider a subspace of  $\mathcal{M}$  spanned by  $e^{S_{\text{eff}}}$  independent microstates

$$\hat{\mathcal{M}} = \left\{ \sum_{A'=1}^{e^{S_{\text{eff}}}} a_{A'} |\Psi_{A'}(M)\rangle \middle| a_{A'} \in \mathbb{C}, \sum_{A'=1}^{e^{S_{\text{eff}}}} |a_{A'}|^2 = 1 \right\}, \quad (5.34)$$

where  $S_{\text{eff}} < S_{\text{tot}}$ . By choosing the bases of  $\mathcal{M}$  and  $\hat{\mathcal{M}}$  appropriately, we can take  $\{|\Psi_{A'}(M)\rangle\}$  to be a subset of  $\{|\Psi_A(M)\rangle\}$ , so  $\langle\Psi_{A'}(M)|\Psi_{B'}(M)\rangle = \delta_{A'B'}$ .

We now define the following operators associated with the Hilbert subspace  $\hat{\mathcal{M}}$

$$\tilde{\mathcal{B}}_\gamma = \sum_{A'=1}^{e^{S_{\text{eff}}}} \tilde{b}_\gamma^{A'}, \quad \tilde{\mathcal{B}}_\gamma^\dagger = \sum_{A'=1}^{e^{S_{\text{eff}}}} \tilde{b}_\gamma^{A'\dagger}, \quad (5.35)$$

where  $\tilde{b}_\gamma^{A'}$  and  $\tilde{b}_\gamma^{A'\dagger}$  are given by Eqs. (5.12) and (5.14). These operators will work as desired if the excited states obtained by acting  $\tilde{b}_\gamma^{A'}$ 's and  $\tilde{b}_\gamma^{A'\dagger}$ 's on  $|\Psi_{A'}(M)\rangle$  are effectively orthogonal for different  $A'$ 's. In fact, we can show that the algebra of these operators in the black hole Hilbert space built on  $\hat{\mathcal{M}}$ , i.e., the space spanned by the vacuum microstates of  $\hat{\mathcal{M}}$  and the states in which these microstates have been excited, is the same as that of the mode operators in the second exterior of the corresponding two-sided black hole, up to correction of order [25]

$$\epsilon = \max \left\{ \frac{e^{\frac{E_{\text{max}}}{2T_{\text{H}}}}}{e^{\frac{1}{2}(S_{\text{bh}}(M)+S_{\text{rad}})}}, \frac{e^{\frac{E_{\text{max}}}{2T_{\text{H}}+S_{\text{eff}}}}}{e^{S_{\text{bh}}(M)+S_{\text{rad}}}} \right\}, \quad (5.36)$$

where  $E_{\text{max}}$  is the maximum energy which an excitation can carry in the semiclassical theory. Given that  $E_{\text{max}} \ll M$ , we find that the error is exponentially small for

$$S_{\text{eff}} < S_{\text{bh}}(M) + S_{\text{rad}} \approx S_{\text{tot}}, \quad (5.37)$$

where the symbol  $<$  here means that  $S_{\text{eff}}$  is smaller than  $S_{\text{bh}}(M) + S_{\text{rad}}$  and that the fractional difference between  $S_{\text{eff}}$  and  $S_{\text{bh}}(M) + S_{\text{rad}}$  is not exponentially small, specifically

$$S_{\text{bh}}(M) + S_{\text{rad}} - S_{\text{eff}} \gg \frac{E_{\text{max}}}{2T_{\text{H}}}. \quad (5.38)$$

Below, we use the symbols  $<$  and  $>$  to mean similar relations.

Similarly, we can define infalling mode operators

$$\mathcal{A}_\xi = \sum_\gamma (\alpha_{\xi\gamma} b_\gamma + \beta_{\xi\gamma} b_\gamma^\dagger + \zeta_{\xi\gamma} \tilde{\mathcal{B}}_\gamma + \eta_{\xi\gamma} \tilde{\mathcal{B}}_\gamma^\dagger), \quad (5.39)$$

$$\mathcal{A}_\xi^\dagger = \sum_\gamma (\beta_{\xi\gamma}^* b_\gamma + \alpha_{\xi\gamma}^* b_\gamma^\dagger + \eta_{\xi\gamma}^* \tilde{\mathcal{B}}_\gamma + \zeta_{\xi\gamma}^* \tilde{\mathcal{B}}_\gamma^\dagger), \quad (5.40)$$

which act linearly in the black hole Hilbert space built on  $\hat{\mathcal{M}}$ . Here, the coefficients  $\alpha_{\xi\gamma}$ ,  $\beta_{\xi\gamma}$ ,  $\zeta_{\xi\gamma}$ , and  $\eta_{\xi\gamma}$  are the same as those in Eqs. (5.23) and (5.24). The matrix elements of

products of  $\mathcal{A}_\xi$  and  $\mathcal{A}_\xi^\dagger$  in the black hole Hilbert space built on  $\hat{\mathcal{M}}$ , then, are the same as the corresponding field theory values on the two-sided black hole background, up to corrections suppressed by  $\epsilon$  in Eq. (5.36).<sup>38</sup>

The existence of operators  $\mathcal{A}_\xi$  and  $\mathcal{A}_\xi^\dagger$  allows us to erect the effective theory such that the dependence of operators on states is invisible in the effective theory. Suppose that the state at  $t = t_*$  is an entangled state between semiclassical excitations and black hole vacuum microstates

$$|\Psi(t_*)\rangle = \sum_{A=1}^{S_{\text{tot}}} \sum_I d_{AI}(t_*) |\Psi_{A,I}(M)\rangle, \quad (5.41)$$

where  $|\Psi_{A,I}(M)\rangle$  represents the state in which the semiclassical excitation  $I$  exists on the black hole vacuum state  $|\Psi_A(M)\rangle$ . Even in this case, given that the logarithm of the dimension of the excitation Hilbert space,  $S_{\text{exc}}$ , is much smaller than  $S_{\text{tot}}$ , we can write the state using the Schmidt decomposition as

$$|\Psi(t_*)\rangle = \sum_{I=1}^{\mathcal{K}} g_I |\Psi_{A(I),I}(M)\rangle, \quad (5.42)$$

where  $\sum_{I=1}^{\mathcal{K}} |g_I|^2 = 1$ ,  $g_I > 0$ , and  $\mathcal{K}$  is the Schmidt number. The point is that since  $\mathcal{K}$  satisfies

$$\mathcal{K} \leq S_{\text{exc}} \ll S_{\text{bh}}(M) + S_{\text{rad}}, \quad (5.43)$$

$\hat{\mathcal{M}}$  can always be taken to contain the space of vacuum microstates spanned by  $\{|\Psi_{A(I)}(M)\rangle | I = 1, \dots, \mathcal{K}\}$ . This guarantees that the effective theory respects the standard tenet of quantum mechanics that physical observables are given by linear operators acting on the Hilbert space of the theory.

Since  $S_{\text{eff}}$  only needs to satisfy Eq. (5.37), one might think that a single, fixed set of  $\mathcal{A}_\xi$ ,  $\mathcal{A}_\xi^\dagger$  operators can cover the states built on most of the black hole vacuum microstates in  $\mathcal{M}$  of Eq. (5.33) by taking

$$S_{\text{eff}} = c \{S_{\text{bh}}(M) + S_{\text{rad}}\} \quad (5.44)$$

with  $c$  close to (but not exponentially close to) 1. This is, however, not the case. The dimension of the space  $\hat{\mathcal{M}}_\perp$  of vacuum microstates that are orthogonal to the states in  $\hat{\mathcal{M}}$  is

<sup>38</sup>Precisely speaking, the operators  $\mathcal{A}_\xi$  and  $\mathcal{A}_\xi^\dagger$  (and  $\tilde{\mathcal{B}}_\gamma$  and  $\tilde{\mathcal{B}}_\gamma^\dagger$ ) can be used for states built on a larger vacuum microstate space. Specifically, the algebra of these operators is the same as the corresponding semiclassical algebra in the space of black hole states built on a typical state in subspace  $\hat{\mathcal{M}}'$  of  $\mathcal{M}$  as long as  $\dim(\hat{\mathcal{M}}' \cap \hat{\mathcal{M}}) \gg \dim \hat{\mathcal{M}}'/2$  [25].

$$\dim \hat{\mathcal{M}}_\perp = e^{S_{\text{tot}}} - e^{S_{\text{eff}}}, \quad (5.45)$$

which is much larger than  $\dim \hat{\mathcal{M}} = e^{S_{\text{eff}}}$  even for  $c$  close to 1. In fact, for  $c > 1/2$ , there is a simple relation between the fraction of  $\mathcal{M}$  which a fixed set of operators can cover and the size of error for using these operators:

$$\frac{\dim \hat{\mathcal{M}}}{\dim \mathcal{M}} \approx \epsilon. \quad (5.46)$$

This makes it clear that we cannot use a fixed set of operators to cover a significant fraction of states in  $\mathcal{M}$  while keeping the error  $\epsilon$  small. In fact, to cover all states in  $\mathcal{M}$  by fixed sets of operators, we need double exponentially large number,  $O(e^{e^{S_{\text{tot}} - S_{\text{eff}}}})$ , of sets.<sup>39</sup>

Finally, we note that the construction of the effective theory described here has an intrinsic ambiguity coming from the fact that the actions of infalling mode operators are not strictly orthogonal to  $\hat{\mathcal{M}}$  in the space of black hole microstates. Specifically, we find that the inner product between states obtained by operating mode operators on two vacuum microstates  $|\Psi_A(M)\rangle$  and  $|\Psi_B(M)\rangle$  reproduces the field theory value multiplied by  $\delta_{AB}$ , but with corrections of order  $\epsilon$  in Eq. (5.36) which are not proportional to  $\delta_{AB}$ . The existence of these exponentially suppressed corrections means that an excited state cannot have an exact and unambiguous association with a unique vacuum microstate.

The fact that the corrections are only of order  $\epsilon$ , however, implies that up to these exponentially suppressed corrections, the mode operators  $b_\gamma$ ,  $b_\gamma^\dagger$ ,  $\tilde{\mathcal{B}}_\gamma$ ,  $\tilde{\mathcal{B}}_\gamma^\dagger$ ,  $\mathcal{A}_\xi$ , and  $\mathcal{A}_\xi^\dagger$  act only on the excitation index  $I$ , and not on the vacuum index  $A'$ . In other words, ignoring these corrections, the Hilbert space can be viewed as

$$\mathcal{H} \approx \mathcal{H}_{\text{exc}} \otimes (\mathcal{H}_{\text{vac}} \cong \hat{\mathcal{M}}), \quad (5.47)$$

where these mode operators act only on  $\mathcal{H}_{\text{exc}}$ , and an excited state can be associated “uniquely” with a vacuum microstate. The exponentially suppressed corrections discussed here constitute an intrinsic ambiguity of semiclassical physics, resulting from the fact that the black hole system (consisting of zone and horizon modes as well as the relevant degrees of freedom of far modes) is finite dimensional.

### 3. Consistency with the semiclassical expectation

Suppose that a falling object hits the stretched horizon at  $t = t_*$  when viewed from the exterior. The object then disappears from the zone, but this does not mean that the

<sup>39</sup>This is related to the well-known fact that in a Hilbert space of dimension  $e^S \gg 1$ , there are  $O(e^S)$  approximately orthogonal states with exponentially small overlaps of  $O(e^{-S/2})$ .

state immediately becomes the quasi-equilibrium form of Eq. (3.44) with Eq. (3.42) at  $t = t_*$ ; rather, it stays in a state with excited horizon modes for a while. Now, consider that we erect an effective theory of the interior shortly after  $t_*$ :  $t = t_* + \delta t$ . In this case, the effective theory can have semiclassical excitations in the interior reflecting the fact that there is an object that has fallen into the horizon at  $t_*$ . These include everything that the object does to the spacetime region described by the effective theory; in particular, the excitations existing in the effective theory need not be the object itself.<sup>40</sup>

On the other hand, if we erect an effective theory more than one scrambling time after the last disturbance to the stretched horizon,  $t = t_* + \Delta t$  with  $\Delta t > t_{\text{scr}}$ , then we expect that the effective theory finds the semiclassical vacuum in the interior, since the black hole state has already equilibrated by then. (There can be zone mode excitations corresponding to semiclassical objects in the zone.) Is this consistent with the semiclassical expectation?

In Fig. 12, we show by the central blue diamond the spacetime region described by the effective theory erected at  $t = t_*$ , at which the object reaches the stretched horizon. The coordinates  $U$  and  $V$  are given by Eqs. (5.19) and (5.20). Now, consider the Kruskal-Szekeres coordinates  $\tilde{U}$  and  $\tilde{V}$  adapted to the effective theory erected at  $t = t_* + \Delta t$ , which are related to  $U$  and  $V$  by

$$\tilde{U} = U e^{\frac{1}{2r_+}\Delta t}, \quad \tilde{V} = V e^{-\frac{1}{2r_+}\Delta t} \quad (5.48)$$

(so that  $\tilde{U} = U$  and  $\tilde{V} = V$  for  $\Delta t = 0$ ). The spacetime region that can be described by this theory, depicted by the red shaded region in Fig. 12 for  $\Delta t > 0$ , must thus satisfy

$$\tilde{U} \lesssim r_z = O(r_+) \quad \Rightarrow \quad U \lesssim r_+ e^{-\frac{1}{2r_+}\Delta t}. \quad (5.49)$$

It is then clear that the object that fell into the horizon at  $t = t_*$  spends only tiny proper time in this region, given by

$$\Delta\tau \lesssim r_+ e^{-\frac{1}{2r_+}\Delta t}. \quad (5.50)$$

Since holography limits the maximum amount of information an object can handle to be of  $O(1)$  per Planck time, this

<sup>40</sup>Suppose, for example, that the object hits the stretched horizon at  $t = t_*$  with almost the speed of light. Then, the object will not appear in the effective theory erected at  $t = t_* + \delta t$ , since it is squeezed into the  $U$  axis so strongly that it cannot exist in the effective theory, a theory that cannot describe physics below the Planck length. Even in this case, however, the object may emit a high energy quantum in the outward direction shortly after it crosses the horizon. The quantum would then have to appear in the effective theory as a signal emerging from somewhere on the  $U$  axis to the positive  $V$  direction, if  $\delta t$  is sufficiently small. (This modification of the boundary condition along the  $U$  axis must occur through a map of horizon excitations in the microscopic theory to the effective theory, which requires a UV physics.)

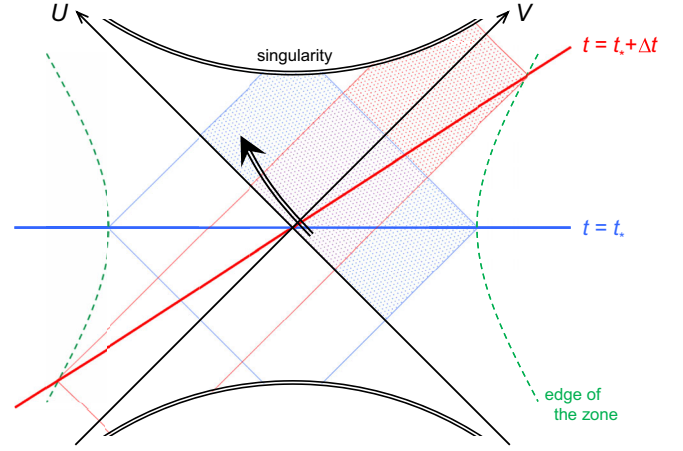


FIG. 12. The spacetime regions described by effective theories erected at  $t = t_*$  (blue) and  $t = t_* + \Delta t$  with  $\Delta t > 0$  (red). The black arrow indicates a falling object that reaches the stretched horizon at  $t = t_*$ .

implies that the object cannot cause any physical effect in an effective theory erected after  $t = t_* + \Delta t_{\text{max}}$ , with

$$\Delta t_{\text{max}} = 2r_+ \ln \frac{r_+}{l_P}, \quad (5.51)$$

where the expression is reliable up to fractional corrections of order  $1/\ln(r_+/l_P)$ . Comparing this with the scrambling time in Eq. (5.3), we find that the picture is indeed consistent with what we expect from semiclassical theory.

#### 4. Young black hole and the role of the Page time

The construction of the effective theory described so far applies to a black hole of any age. However, if the black hole is young, i.e., if it is not maximally entangled with other systems, then we can have an alternative construction in which the  $\tilde{b}_\gamma^A$  and  $\tilde{b}_\gamma^{A\dagger}$  operators act only on soft-mode states, without involving far mode degrees of freedom [23,25].

This is done by projecting operators in Eqs. (5.11) and (5.13) on the space of soft-mode states using the Petz map [149,150]. Specifically, we can take

$$\begin{aligned} \tilde{b}_\gamma^A &= \sigma^{-1/2} \left[ \sum_a \langle \phi_a | \left( \sum_n \sqrt{n_\gamma} \| \{n_\alpha - \delta_{\alpha\gamma}\}_A \rangle \right) \right. \\ &\quad \left. \times \langle \langle \{n_\alpha\}_A \rangle \rangle | \phi_a \rangle \right] \sigma^{-1/2}, \end{aligned} \quad (5.52)$$

$$\begin{aligned} \tilde{b}_\gamma^{A\dagger} &= \sigma^{-1/2} \left[ \sum_a \langle \phi_a | \left( \sum_n \sqrt{n_\gamma + 1} \| \{n_\alpha + \delta_{\alpha\gamma}\}_A \rangle \right) \right. \\ &\quad \left. \times \langle \langle \{n_\alpha\}_A \rangle \rangle | \phi_a \rangle \right] \sigma^{-1/2}. \end{aligned} \quad (5.53)$$

Here,

$$\sigma = \sum_a \langle \phi_a | \left( \sum_n \|\{n_a\}_A\| \langle\langle \{n_a\}_A \rangle\rangle \right) | \phi_a \rangle, \quad (5.54)$$

and  $\|\{n_a\}_A\|$ 's are given by Eq. (5.4). Note that  $\tilde{b}_\gamma^A$ ,  $\tilde{b}_\gamma^{A\dagger}$ , and  $\sigma$  are operators acting only on the space of soft-mode states. Infalling mode operators acting only on the hard and soft modes can then be constructed by substituting these  $\tilde{b}_\gamma^A$  and  $\tilde{b}_\gamma^{A\dagger}$  in Eqs. (5.23) and (5.24).

The algebra of the above mode operators in the black hole Hilbert space, i.e., the space obtained by acting mode operators on the vacuum state in Eq. (3.44), follows that of semiclassical theory up to errors of order

$$\epsilon_{\text{young}} = \max \left\{ \frac{1}{e^{\frac{1}{2}S_{\text{bh}}(M)}}, \frac{e^{S_{\text{rad}}}}{e^{S_{\text{bh}}(M)}} \right\}. \quad (5.55)$$

Therefore, if the black hole is young, i.e.,  $S_{\text{rad}} < S_{\text{bh}}(M)$ , then the operators  $\tilde{b}_\gamma^A$  and  $\tilde{b}_\gamma^{A\dagger}$  in Eqs. (5.52) and (5.53) as well as  $a_\xi^A$  and  $a_\xi^{A\dagger}$  constructed using them can be used to describe the interior of the black hole (up to corrections suppressed exponentially in a macroscopic entropy). This elucidates the role of the Page time in constructing the effective theory: if a black hole is young, then interior operators *can be* represented purely using the black hole degrees of freedom (i.e., zone and horizon modes), while if it is old, then the operators *must* involve the early Hawking radiation (i.e., far modes).

The promotion of  $\tilde{b}_\gamma^A$ ,  $\tilde{b}_\gamma^{A\dagger}$ ,  $a_\xi^A$ , and  $a_\xi^{A\dagger}$  operators to act linearly in  $\hat{\mathcal{M}}$  can be made similarly as before. In particular, errors of the promoted operators are of order

$$\hat{\epsilon}_{\text{young}} = \max \left\{ \frac{1}{e^{\frac{1}{2}S_{\text{bh}}(M)}}, \frac{e^{S_{\text{rad}}+S_{\text{eff}}}}{e^{S_{\text{bh}}(M)}} \right\}, \quad (5.56)$$

so that these operators work correctly as long as

$$S_{\text{eff}} < S_{\text{bh}}(M) - S_{\text{rad}}. \quad (5.57)$$

This agrees with the result of the general analysis in Ref. [30].

Incidentally, in the way of constructing an effective theory of the interior described here, operators  $\tilde{b}_\gamma^A$  and  $\tilde{b}_\gamma^{A\dagger}$  cannot be represented using only far modes, even if the black hole is old. Namely, unlike Eqs. (5.52) and (5.53), projecting operators in Eqs. (5.11) and (5.13) on the space of far-mode states does not work. Technically, this is because of the energy constraint imposed on the black hole, i.e., the combined system of zone and horizon modes. The relation of this statement to entanglement wedge reconstruction, in which the interior of an old black hole is reconstructed only using the early Hawking radiation, will be discussed below.

### 5. Relation to entanglement wedge reconstruction

As we have seen, operators in our effective theory describing the interior must involve horizon degrees of freedom, regardless of the age of the black hole. On the other hand, the analysis [33–37] based on holographic entanglement wedge reconstruction [151–156] says that after the Page time operators acting on early radiation are sufficient to reconstruct a portion of the black hole interior. What is the relation between these two statements?

A key ingredient to understanding this is the boundary time evolution [24]. In general, entanglement wedge reconstruction assumes that we know the time evolution operator of the boundary theory; in models discussed in Refs. [33–37], for example, the Hamiltonian of a system consists of boundary conformal field theory as well as an auxiliary theory coupling to it. With this knowledge, one can reconstruct the state of some of the horizon modes at  $t = t_h$  from the state of the radiation at  $t = t_R$  if

$$t_R > t_h + t_{\text{scr}}. \quad (5.58)$$

Note that as discussed in Sec. III B, the notion of zone, horizon, and far modes is associated with a specific time  $t$ , so that a component of a horizon mode at some time can become a far mode at a later time.

A detailed way in which the reconstruction described above works was discussed in Ref. [24]. It is essentially the Hayden-Preskill protocol [52] applied to the horizon and zone modes. Recall that zone modes at  $t = t_h - t_{\text{sig}}$  either become horizon modes (for ingoing modes) or far modes (for outgoing modes) by  $t = t_h$ . Here,  $t_{\text{sig}} \approx |r_{*s}|$  is the signal propagation time: the time it takes for a massless quantum to propagate from the stretched horizon to the edge of the zone. Hence, radiation at  $t_R$  satisfying Eq. (5.58) can reconstruct some of the zone modes. In fact, we can arbitrarily choose which zone modes to reconstruct, and we can take them to be hard modes. This implies that radiation at  $t = t_R$  can reconstruct an object in the interior of the effective theory erected at  $t = t_*$  if

$$t_* < t_R - t_{\text{scr}} - t_{\text{sig}} \quad (5.59)$$

and  $t_* + t_{\text{scr}} > t_{\text{Page}}$ . In other words, an object in the black hole interior can be reconstructed from radiation at  $t = t_R$  if it exists in the inner wedge of the stretched horizon at

$$t = t_R - t_{\text{scr}}, \quad (5.60)$$

assuming  $t_R > t_{\text{Page}}$ . This is sketched in Fig. 13(a). The entanglement island,  $\mathcal{I}_{\text{rad}}$ , of the radiation,  $R(t_R)$ , represents the spacetime region in which some information about the region is reconstructed. Thus, its edge, i.e., the minimal quantum extremal surface for  $R(t_R)$ , is located near the stretched horizon at  $t = t_R - t_{\text{scr}}$ . Given that the



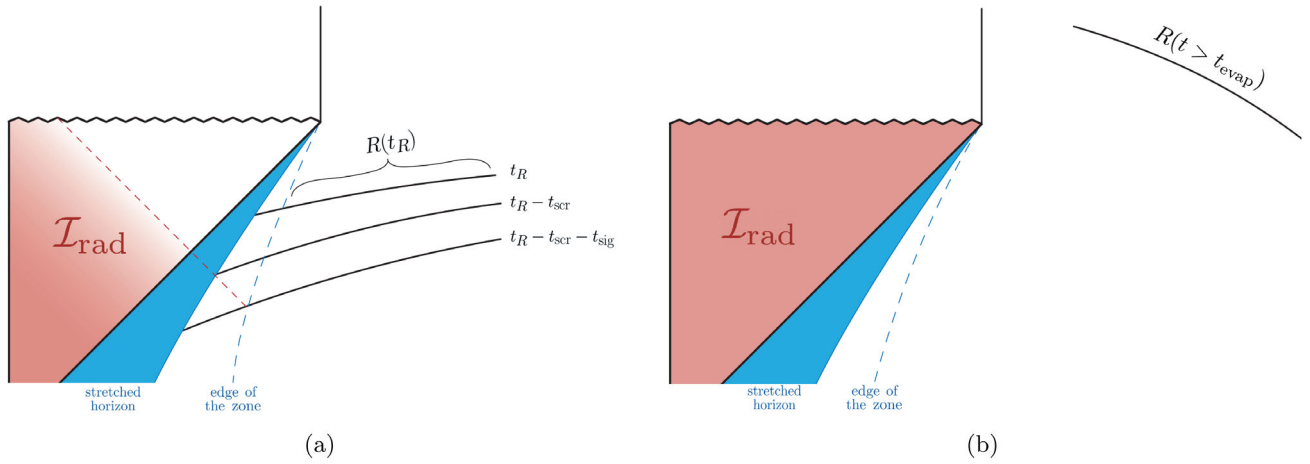


FIG. 13. (a) Radiation at  $t = t_R$ ,  $R(t_R)$ , contains information about some of the horizon and zone modes at  $t < t_R - t_{\text{scr}} - t_{\text{sig}}$ . This allows us to reconstruct a portion  $\mathcal{I}_{\text{rad}}$  of the interior spacetime corresponding to the inner wedge of the stretched horizon at  $t = t_R - t_{\text{scr}}$ . The amount of information that can be reconstructed, however, depends on the location of an object in ingoing time, which is indicated by the gradation of the red shade. (b) Radiation after the black hole is fully evaporated,  $R(t > t_{\text{evap}})$ , allows for reconstruction of the full interior region.

discussion for reconstruction here is intrinsically semi-classical, and hence does not have a precision of resolving the Planck scale, this is consistent with what was found in Refs. [33–35].

We stress that while the entanglement island  $\mathcal{I}_{\text{rad}}$  represents the interior region which one can reconstruct from the radiation  $R(t_R)$ , the amount of information that can be reconstructed, or the size of code subspace [106,107] one can construct, depends on the location in this region. Suppose one wants to reconstruct an interior object that has a significant amount of information, say having (coarse-grained) entropy  $S_{\text{obj}}$ . Then, the radiation must contain the corresponding amount of information. This modifies Eqs. (5.59) and (5.60) to

$$t_* < t_R - t_{\text{scr}} - t_{\text{sig}} - \frac{S_{\text{obj}}}{T_{\text{H}}} \quad (5.61)$$

and

$$t = t_R - t_{\text{scr}} - \frac{S_{\text{obj}}}{T_{\text{H}}}, \quad (5.62)$$

where  $t_R$  must satisfy  $t_R > t_{\text{Page}} + S_{\text{obj}}/T_{\text{H}}$ . This feature inherits from the use of the Hayden-Preskill protocol, and more fundamentally boundary time evolution, in the reconstruction. Because of the scrambling and quantum error correcting nature of black hole dynamics, one can arbitrarily choose *which* information to reconstruct [52], but the *amount* of information to be reconstructed is bounded from above by

$$S_{\text{obj}} \leq T_{\text{H}}(t_R - t_{\text{scr}} - v_{\text{obj}}), \quad (5.63)$$

where  $v_{\text{obj}}$  is the ingoing Eddington-Finkelstein time  $v$  at the location of the object, chosen so that  $v = t$  at the stretched horizon. This is represented by the gradation of the red shade in Fig. 13(a).

The discussion above makes it clear that entanglement wedge reconstruction is nothing but the statement of unitarity as viewed from the exterior. Indeed, the entanglement island of radiation after the black hole is fully evaporated can be viewed as the entire interior; see Fig. 13(b).<sup>41</sup>

The understanding of entanglement wedge reconstruction described above also elucidates why an operator acting only on radiation  $R(t_R)$ , in particular a unitary operator representing a physical manipulation of  $R(t_R)$ , can affect the interior of a black hole  $\mathcal{I}_{\text{rad}}$  in a seemingly acausal manner. In the manifestly unitary picture adopted so far in this paper, based on an external view of the black hole, an excitation  $X$  in the interior  $\mathcal{I}_{\text{rad}}$  is, in fact, “located” in the zone and/or stretched horizon at  $t < t_R$  (and possibly in the early radiation as well for an old black hole), which is timelike separated from  $R(t_R)$ . There is, therefore, no *a priori* reason why the operator generating  $X$  commutes with operators acting on the radiation at  $t = t_R$ . In particular, an operator  $\mathcal{O}_R$  acting on  $R(t_R)$  can be a “precursor” [157,158] of  $X$ ; i.e., when evolved backward in time,  $\mathcal{O}_R$  induces a change of the state of an object falling into the black hole at an earlier time, which changes the interior of the black hole.

<sup>41</sup>Strictly speaking, the existence of an island in this case is not rigorously established, since its edge cannot be obtained as a surface that is quantum extremized in the regime in which a semiclassical description is valid. However, from physical consideration as well as continuity of  $\mathcal{I}_{\text{rad}}$  in  $t$ , we expect that  $R(t)$  ( $t > t_{\text{evap}}$ ) has the entire interior region as the “island.”

On the other hand, in the global spacetime picture of general relativity, it is still true that radiation at  $t = t_R$  is spacelike separated from the interior region  $\mathcal{I}_{\text{rad}}$ . This raises the question of why there can be any unitary operator acting on  $R(t_R)$  which affects  $\mathcal{I}_{\text{rad}}$ , without completely jeopardizing the intuition coming from the semiclassical, global spacetime picture. This issue was studied in Ref. [159], in which it was argued that if the dynamics of a black hole is sufficiently complex (e.g., leading to a pseudorandom state), then the information about the interior  $\mathcal{I}_{\text{rad}}$  cannot be accessed by any simple operator acting on  $R(t_R)$  which does not have exponential computational complexity. This reproduces the causality of the semiclassical theory, assuming that it represents only a feature of simple operations performed on the system in each observer frame.

In the present context, the required complexity arises from the dynamics of the stretched horizon as viewed from the exterior of the black hole. Note that the time  $t_*$  at which the effective theory describing  $\mathcal{I}_{\text{rad}}$  is erected must satisfy Eq. (5.61). This implies that ingoing zone modes (as well as most of the horizon modes) at  $t = t_*$  must go through the complex horizon dynamics before they become radiation modes at  $t = t_R$ . In particular, this is true for the ingoing hard modes for any choice of hard modes describing the object carrying the entropy  $S_{\text{obj}}$ . The semiclassical picture emerging from our framework, therefore, respects causality, at least in the sense of Ref. [159].

In entanglement wedge reconstruction, reconstructing the interior on a Hawking radiation state involves backward time evolution, so that the resulting boundary operators acting on the radiation degrees of freedom are highly fragile; i.e., a small deformation of the operators destroys the success of the reconstruction. In the case of the effective theory, on the other hand, interior operators are given simply by Eqs. (5.12), (5.14), (5.23), and (5.24) in terms of hard, soft, and far mode operators at time  $t_*$ , which can be obtained easily in the boundary theory from the structure of entanglement of the whole state at  $t = t_*$ , if such a state is given.<sup>42</sup> In this “equal-time” conversion between the distant and infalling descriptions, which use modes at a given time  $t_*$  directly, there is no upper limit on the amount of information about the infalling object described, such as that in Eq. (5.63); the only restriction on the size of the object is the one coming from the validity of the semiclassical picture.

<sup>42</sup>A specific construction can go as follows. We can first construct zone mode operators  $b_\gamma$  and  $b_\gamma^\dagger$  using the HKLL procedure [11, 12, 114] or its extension [16]. We can then expand the full state  $|\Psi(t_*)\rangle$  at  $t = t_*$  in terms of the eigenstates of number operators  $b_\alpha^\dagger b_\alpha$ 's and identify the  $|\{n_\alpha\}\rangle$  states associated with  $|\Psi(t_*)\rangle$ . This allows us to define  $b_\gamma$  and  $b_\gamma^\dagger$  operators acting on  $|\{n_\alpha\}\rangle$ 's, and hence infalling operators  $a_\xi$  and  $a_\xi^\dagger$  through Eqs. (5.23) and (5.24). The effective theory of the interior (erected at  $t = t_*$ ) can then be obtained using the generator  $\tilde{H}$  of time evolution, given in Eq. (5.27).

## B. The outside of a de Sitter static patch

A construction analogous to the black hole interior can be applied to a state describing de Sitter spacetime [22]. As we have argued in Sec. IV B, a microstate representing empty de Sitter spacetime consists of the state of the “zone modes” inside a static patch as well as that of the horizon modes. Thus, by choosing a subset of the zone modes as hard modes, it can be written in the form of Eq. (3.34) with  $S(E - E_n) \rightarrow S_{\text{dS}}(E - E_n)$ . Here,  $S_{\text{dS}}(E)$  is given by Eq. (3.27), and  $E$  is the “energy” given in terms of the Hubble radius  $\alpha$  by  $E = \alpha/l_p^2$ ; see Eq. (3.16). As in the case of a black hole, we can define the “double-ket” states using soft modes

$$|\{n_\alpha\}_A\rangle\rangle = \zeta_n^A \sum_{i_n=1}^{e^{S_{\text{dS}}(E-E_n)}} c_{ni_n}^A |\psi_{i_n}^{(n)}\rangle, \quad (5.64)$$

where  $A = 1, \dots, e^{S_{\text{sys}}}$ , and  $\zeta_n^A$  is the normalization constant

$$\zeta_n^A = \frac{1}{\sqrt{\sum_{i_n=1}^{e^{S_{\text{dS}}(E-E_n)}} c_{ni_n}^{A*} c_{ni_n}^A}} = \sqrt{z} e^{\frac{E_n}{2T_H}} [1 + O(e^{-\frac{1}{2}S_{\text{dS}}(E-E_n)})], \quad (5.65)$$

where  $T_H = 1/2\pi\alpha$ , and  $z$  is given by Eq. (3.31). In terms of these states, the de Sitter vacuum microstate can be rewritten as

$$|\Psi_A(E)\rangle\rangle = \frac{1}{\sqrt{z}} \sum_n e^{-\frac{E_n}{2T_H}} |\{n_\alpha\}\rangle |\{n_\alpha\}_A\rangle\rangle. \quad (5.66)$$

Since the state  $|\{n_\alpha\}\rangle$  is specified by the occupation numbers  $n_\alpha$  of modes inside the stretched horizon of a static patch, which we refer to as the polar region here, the thermofield double state in Eq. (5.66) represents a vacuum state at  $\tau = 0$  of global de Sitter spacetime of which the original static patch is a portion. Here,  $\tau$  is global de Sitter time, or time associated with closed slicing. This implies that  $|\{n_\alpha\}_A\rangle\rangle$  should be identified as the state of the other hemisphere of the static patch at  $\tau = 0$ . States in which modes in the polar region are excited or deexcited are those obtained by acting corresponding creation operators Eq. (3.50) or annihilation operators Eq. (3.49), respectively, to a de Sitter vacuum microstate.<sup>43</sup> These states are naturally mapped to excited states in the effective theory built on Eq. (5.66).

As in the case of a black hole, we can introduce annihilation and creation operators for modes in the other hemisphere in the effective theory as

<sup>43</sup>Deexcited states become relevant if Gibbons-Hawking radiation is extracted by a material.

TABLE I. Correspondence between an evaporating black hole and cosmological de Sitter spacetime.

	Evaporating black hole	Cosmological de Sitter spacetime
Microscopic level	Zone region Stretched horizon Far region	Polar region Stretched horizon ...
Effective theory	Two-sided black hole The second exterior	Global de Sitter spacetime The other hemisphere

$$\tilde{b}_\gamma^A = \sum_n \sqrt{n_\gamma} \zeta_{n-\gamma}^A \zeta_n^{A*} e^{S_{\text{dS}}(E-E_{n-\gamma})} e^{S_{\text{dS}}(E-E_n)} \sum_{i_{n-\gamma}=1} \sum_{j_n=1} c_{n-\gamma i_{n-\gamma}}^A c_{n j_n}^{A*} \times |\psi_{i_{n-\gamma}}^{(n-\gamma)}\rangle \langle \psi_{j_n}^{(n)}|, \quad (5.67)$$

$$\tilde{b}_\gamma^{A\dagger} = \sum_n \sqrt{n_\gamma + 1} \zeta_{n+\gamma}^A \zeta_n^{A*} e^{S_{\text{dS}}(E-E_{n+\gamma})} e^{S_{\text{dS}}(E-E_n)} \sum_{i_{n+\gamma}=1} \sum_{j_n=1} c_{n+\gamma i_{n+\gamma}}^A c_{n j_n}^{A*} \times |\psi_{i_{n+\gamma}}^{(n+\gamma)}\rangle \langle \psi_{j_n}^{(n)}|, \quad (5.68)$$

which can be used to form annihilation and creation operators for global time slicing

$$a_\xi^A = \sum_\gamma (\alpha_{\xi\gamma} b_\gamma + \beta_{\xi\gamma} b_\gamma^\dagger + \zeta_{\xi\gamma} \tilde{b}_\gamma^A + \eta_{\xi\gamma} \tilde{b}_\gamma^{A\dagger}), \quad (5.69)$$

$$a_\xi^{A\dagger} = \sum_\gamma (\beta_{\xi\gamma}^* b_\gamma + \alpha_{\xi\gamma}^* b_\gamma^\dagger + \eta_{\xi\gamma}^* \tilde{b}_\gamma^A + \zeta_{\xi\gamma}^* \tilde{b}_\gamma^{A\dagger}), \quad (5.70)$$

where  $\xi$  is the label in which the frequency  $\omega$  with respect to static time  $t$  is traded with the frequency  $\Omega$  associated with global time  $\tau$ , and  $\alpha_{\xi\gamma}$ ,  $\beta_{\xi\gamma}$ ,  $\zeta_{\xi\gamma}$ , and  $\eta_{\xi\gamma}$  are the Bogoliubov coefficients (which, of course, differ from the black hole case). The correspondence between the black hole and de Sitter cases is summarized in Table I. Note that for de Sitter spacetime, there is no region corresponding to the region outside the zone of an evaporating black hole.

Physical quantities in global de Sitter spacetime can be calculated using the global time evolution operator

$$U(\tau) = e^{-i\tilde{H}\tau}, \quad (5.71)$$

where

$$\tilde{H} = \sum_\xi \Omega_\xi a_\xi^{A\dagger} a_\xi^A + \tilde{H}_{\text{int}}(\{a_\xi^A\}, \{a_\xi^{A\dagger}\}). \quad (5.72)$$

In the Heisenberg picture, this can be done by evolving quantum fields  $\tilde{\Phi}_a(\mathbf{x}, 0)$  formed from  $a_\xi^A$  and  $a_\xi^{A\dagger}$  at  $\tau = 0$  as

$$\tilde{\Phi}_a(\mathbf{x}, \tau) = U(\tau)^\dagger \tilde{\Phi}_a(\mathbf{x}, 0) U(\tau) \quad (5.73)$$

and sandwiching their products by the state  $|\Psi(\tau = 0)\rangle$  of the effective theory at  $\tau = 0$  obtained by matching with the state of the microscopic theory at  $t = 0$ . ( $|\Psi(\tau = 0)\rangle = |\Psi_A(E)\rangle$  if the system is in the semiclassical vacuum.) In the Schrödinger picture, the state must be evolved with  $U(\tau)$  while  $\tilde{\Phi}_a(\mathbf{x}, 0)$  are inserted at intermediate stages of the evolution.

The promotion of operators to a less state-dependent form can be made in a similar way to the black hole case, Eqs. (5.35), (5.39), and (5.40). The analysis of errors of operator algebras can be performed similarly. In particular, the global description in the effective theory has an intrinsic error of order

$$\epsilon = \max \left\{ e^{\frac{E_{\text{max}}}{27H}}, e^{\frac{E_{\text{max}} + S_{\text{eff}}}{27H}} \right\}, \quad (5.74)$$

where  $S_{\text{GH}}$  is the Gibbons-Hawking entropy,  $E_{\text{max}}$  is the maximum energy that an excitation can carry in the effective theory, and  $S_{\text{eff}}$  represents the size of the microscopic Hilbert space covered by the promoted operators, as defined in Eq. (5.34).

Like the case of a black hole, the effective theory of global de Sitter spacetime is intrinsically semiclassical in that the algebra of operators in the theory is defined only up to an uncertainty of  $e^{-S_{\text{GH}}/2}$ . This is consistent with the observation in Ref. [59] that symmetries of classical de Sitter spacetime cannot be implemented exactly in a finite-dimensional Hilbert space.

### 1. Consistency with the semiclassical expectation

Suppose an object hits the stretched horizon in the static patch description at some time  $t = t_{\text{obj}}$ . The state shortly after it will not take the form of Eq. (3.34) with random  $c_{ni_n}^A$  which leads to the vacuum state of Eq. (5.66). Instead, it stays in an excited state for a while, reflecting the existence of an object just outside the horizon. In the effective theory, this is represented as the existence of an excitation in the other hemisphere at  $\tau = 0$ . Such an excitation has a physical significance, since we can retrieve (a part of) information about it if the system leaves the de Sitter phase after  $t = t_{\text{obj}}$ , e.g., by tunneling into a Minkowski vacuum; see Fig. 14.

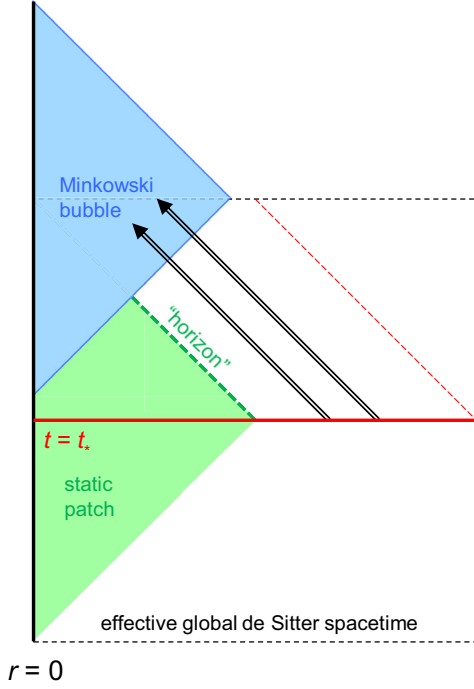


FIG. 14. Information outside the static patch in the effective theory can be retrieved if the system leaves the de Sitter phase at a later time, for example by tunneling into a Minkowski bubble universe.

According to the static patch description, the system is expected to relax into an equilibrium state of the form of Eq. (3.34) when one scrambling time<sup>44</sup>

$$t_{\text{scr}} = \alpha \left[ \ln \frac{\alpha}{l_{\text{P}}} + O(1) \right] \quad (5.77)$$

has passed after the object hits the stretched horizon. This implies that if we erect the effective theory at  $t_* > t_{\text{obj}} + t_{\text{scr}}$ , then the effective theory sees the vacuum in the other hemisphere. Is this consistent with the semiclassical expectation?

<sup>44</sup>To obtain this expression, we can consider an analogue of ingoing Eddington-Finkelstein time in de Sitter spacetime,  $v = t - r_*$ , and use the general expression for scrambling time to obtain

$$v_{\text{scr}} = \frac{1}{2\pi T_{\text{H}}} \ln S_{\text{dS}} \approx 2\alpha \ln \frac{\alpha}{l_{\text{P}}}. \quad (5.75)$$

The scrambling time in static time is then

$$t_{\text{scr}} = v_{\text{scr}} - r_{*s} \approx \alpha \ln \frac{r_+}{l_{\text{P}}}, \quad (5.76)$$

where we have used  $r_{*s} \approx \alpha \ln(r_+/l_{\text{P}})$ , obtained from Eq. (2.28) by identifying  $l_s$  with  $l_{\text{P}}$ .

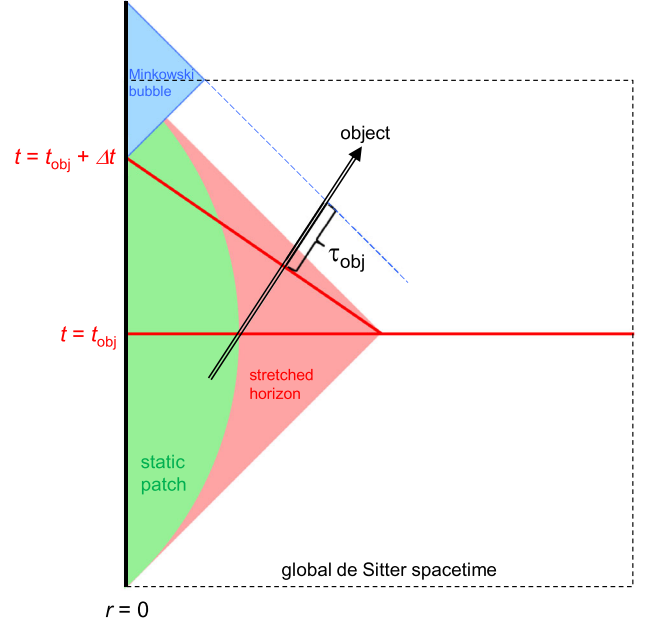


FIG. 15. An object falling into the stretched horizon at  $t = t_{\text{obj}}$  can try to send a physical signal to the observer at  $r = 0$  who enters a Minkowski bubble nucleated at  $r = 0$  at  $t = t_{\text{obj}} + \Delta t$ . The holographic static-patch description of de Sitter spacetime suggests that this is not possible if  $\Delta t$  is larger than the scrambling time  $t_{\text{scr}}$ . This expectation is consistent with the semiclassical picture, since for  $\Delta t > t_{\text{scr}}$  the object would have to send the signal within proper time  $\tau_{\text{obj}} \lesssim l_{\text{P}}$ , which is not possible.

In Fig. 15, we depict the Penrose diagram of the situation in which an object hits the stretched horizon at  $t_{\text{obj}}$ , and a Minkowski bubble is nucleated at the location of the observer,  $r = 0$ , at  $t = t_{\text{obj}} + \Delta t$ . We expect that if  $\Delta t > t_{\text{scr}}$ , the observer cannot receive any information from the object, since then the effective theory erected at  $t = t_{\text{obj}} + \Delta t$  will not have an excitation in the other hemisphere of the effective global de Sitter spacetime, and so there is nothing that can send a signal to the observer. From the viewpoint of the object, this implies that it cannot send any signal to the observer after it crosses the  $t = t_{\text{obj}} + \Delta t$  hypersurface.

This is indeed consistent with what semiclassical theory implies. We can show that in order for the signal to reach the Minkowski bubble, and hence the observer, the object must send it within the proper time

$$\tau_{\text{obj}} \lesssim l_{\text{P}} \quad (5.78)$$

after it crosses the  $t = t_{\text{obj}} + \Delta t$  hypersurface, but this is not possible because the holographic principle does not allow for the object to send physical information within  $l_{\text{P}}$ . Incidentally, this conclusion does not change even if the object is accelerated toward  $r = 0$  after it crosses the



stretched horizon, as long as the invariant acceleration is smaller than  $O(1/l_s)$ .<sup>45</sup>

## 2. Schwarzschild de Sitter spacetime

Our analysis applies straightforwardly to a Schwarzschild de Sitter spacetime which has a black hole of radius  $r_{\text{bh}}$  (centered at  $r = 0$ ) in de Sitter spacetime of radius  $r_{\text{dS}}$ , as long as  $r_{\text{bh}}$  is sufficiently smaller than  $r_{\text{dS}}$ . (The case in which  $r_{\text{bh}}$  and  $r_{\text{dS}}$  are comparable is analogous to the case of a near extremal black hole, which we do not consider in this paper.)

Specifically, if  $r_{\text{bh}} \ll r_{\text{dS}}$ , the black hole can be viewed as a small excitation for the purpose of erecting the effective theory outside the de Sitter horizon, so that the construction described so far goes through without significant modifications. The situation is similar for the effective theory of the black hole interior. A notable thing is that if the black hole is old enough that it is maximally entangled and that the majority of emitted Hawking radiation has reached the stretched de Sitter horizon, then operators describing the black hole interior must involve degrees of freedom associated with the de Sitter horizon.

## VI. INTRINSICALLY EXTENDED SPACETIMES

So far, we have been considering “realistic,” single-sided systems and seen how the effective two-sided pictures emerge as collective phenomena associated with the stretched horizon degrees of freedom. However, there is nothing theoretically wrong in considering “intrinsically two-sided” systems which involve two copies of holographic theories discussed so far. In fact, such systems have been considered in various contexts in black hole [160–171] and de Sitter [43–48, 172–177] physics.

In this section, we discuss how such intrinsically two-sided systems can be understood in the framework described so far. In particular, we will see how these and single-sided systems lead to similar semiclassical physics, despite the fact that states in the two cases have rather different structures at the microscopic level. While we focus here on the two-sided case for a black hole, we expect it is relatively straightforward to extend it to black holes with more than two exterior regions [178].

### A. Two-sided eternal black hole

Let us consider a static, two-sided eternal black hole. Strictly speaking, for a finite black hole mass, this system exists only as a large AdS black hole, but we can imagine that a two-sided black hole in asymptotically flat spacetime

also behaves in approximately the same manner if it is sufficiently large compared with the scale that we are interested in.

We specifically consider a thermofield double state which is prepared by the Euclidean path integral over a half of the time compactified on a circle of length  $\beta$  [160]. In the context of the AdS/CFT correspondence, this is a thermally entangled state of two CFTs. Similar to the case of a single-sided black hole, we divide modes in each side of the black hole into zone and horizon modes at the state preparation time, which we refer to as  $t = 0$ .<sup>46</sup> We then have four classes of modes, i.e., zone and horizon modes in both (right and left) sides of the black hole.

Assuming that the two sides of the black hole have the same microscopic structure (which corresponds to the case that the two CFTs are the same), the black hole vacuum state at  $t = 0$  as viewed from the exterior is given by

$$|\Psi(M)\rangle = \frac{1}{\sqrt{\sum_{\mu'} e^{-\beta E_{\mu'}}}} \sum_{\mu} e^{-\frac{\beta}{2} E_{\mu}} |\mu\rangle_{\text{R}} |\mu\rangle_{\text{L}}, \quad (6.1)$$

where  $|\mu\rangle_{\text{R}}$  represent energy eigenstates of the holographic theory describing the right side with energy  $E_{\mu}$  and similarly for  $|\mu\rangle_{\text{L}}$ . The mass  $M$  of the black hole is related to  $\beta$  by a smoothness condition for spacetime [179]:

$$M = \frac{r_+}{2l_p^2} \left( 1 + \frac{r_+^2}{L^2} \right), \quad \beta = \frac{4\pi r_+ L^2}{3r_+^2 + L^2}, \quad (6.2)$$

where  $L$  is the AdS length. In the context of holography, this relation can be viewed as arising from the requirement that the boundary spacetime can be smoothly extended to the bulk spacetime [180].

We note that the state  $|\Psi(M)\rangle$  has exactly zero energy under the two-sided (modular) Hamiltonian  $H = H_{\text{R}} - H_{\text{L}}$ , which is possible because the system is static under time evolution generated by  $H$ .<sup>47</sup> This time evolution corresponds to Schwarzschild time evolution in a single-sided black hole; see Eq. (5.15).

In general, the energy eigenstates  $|\mu\rangle_{\text{R}}$  do not agree with the number eigenstates of the zone and horizon modes for the right side (and similarly for the left side). However, to describe the dynamics of a zone mode that has energy  $E \gtrsim 1/\beta = T_{\text{H}}$ , this issue can be ignored. Consider, for example, a zone mode of energy  $\sim E$  localized at  $\bar{r}_*$  in the zone and having a Gaussian tail of the form  $\sim e^{-O(1)E^2(r_* - \bar{r}_*)^2}$ . The correction arising from the existence of the stretched horizon to the quantum theory of matter on

<sup>45</sup>Note that here we have considered the case in which the system was in a metastable de Sitter phase before entering a Minkowski phase. We expect that the situation is different if the earlier, de Sitter phase is only approximate in that there is noticeable time dependence of background spacetime, as in the case of slow-roll inflation. A different analysis will be needed for such cases.

<sup>46</sup>For a large AdS black hole, we define zone modes to be all the modes other than horizon modes defined as in Sec. III B.

<sup>47</sup>We can extend the state in Eq. (6.1) to the case in which two CFTs are different. In this case, the state must have some, though exponentially small, energy uncertainty, reflecting the fact that the spectra of the two CFTs are not identical.

classical black hole background is then suppressed by the exponential of  $E^2(\bar{r}_* - r_{*,s})^2$ . This implies that the effect of the deviation of zone-mode states from the thermofield double form coming from their interactions with the stretched horizon is negligible compared with the factor of  $e^{-\beta E/2}$  if the mode is sufficiently hard ( $E \gg 1/\beta$ ) or away from the stretched horizon ( $\bar{r}_* - r_{*,s} \gg \beta$ ).

We now take a subset of the zone modes on the right side as right-side hard modes, whose state we denote by  $|n\rangle_{\text{R}}$ . Here,  $n = \{n_\alpha\}$  represents the set of occupation numbers of these modes. The corresponding zone modes on the left side are then left-side hard modes, whose state is  $|n\rangle_{\text{L}}$ . The rest of the modes on the right and left sides are right- and left-side soft modes, respectively, whose states we denote by  $|\psi_i\rangle_{\text{R}}$  and  $|\psi_i\rangle_{\text{L}}$ . Here, the index  $i$  runs over all soft-mode states, not just those in a specific energy window, since we no longer impose the energy constraint for modes on one side.

The state in Eq. (6.1) can then be approximated, for the purpose of describing the dynamics of hard modes, as

$$|\Psi(M)\rangle \approx \frac{1}{\sqrt{Z}} \sum_n \sum_i e^{-\frac{\beta}{2}(E_n + E_i)} |n\rangle_{\text{R}} |\psi_i\rangle_{\text{R}} |n\rangle_{\text{L}} |\psi_i\rangle_{\text{L}}, \quad (6.3)$$

where  $E_n$  and  $E_i$  are the energies carried by  $|n\rangle_{\text{R}}$  and  $|\psi_i\rangle_{\text{R}}$  defined with respect to  $H$  (or  $H_{\text{R}}$ ), and

$$Z = \sum_n \sum_i e^{-\beta(E_n + E_i)}. \quad (6.4)$$

We can then build an effective theory of the interior on the vacuum microstate of Eq. (6.3).<sup>48</sup>

It is important to realize that the microscopic entanglement structure of  $|\Psi(M)\rangle$  in Eq. (6.3) is different from that of a microstate of a single-sided black hole, which takes the form of Eq. (3.28) (ignoring the entanglement with far modes). In particular, in Eq. (6.3), the hard modes on the right side, which we identify as “our” side, are entangled directly with those on the left side, while in Eq. (3.28) they are entangled with the soft modes on the same side.

The way in which the thermal nature is introduced is also different in two setups. In Eq. (6.3), the Boltzmann factors are introduced already at the microscopic level, while in Eq. (3.28) the coefficients take random values, and the Boltzmann factors for the hard modes arise only statistically after tracing out the soft modes; see Eq. (3.37). In the case of the two-sided black hole, tracing out the left side gives

$$\rho_{\text{R}} = \frac{1}{Z} \sum_n \sum_i e^{-\beta(E_n + E_i)} |n\rangle_{\text{R}} |\psi_i\rangle_{\text{R}} \langle n|_{\text{R}} \langle \psi_i|. \quad (6.5)$$

<sup>48</sup>In this particular case, we can take the hard modes to be the entire zone modes, though this is not the case in general.

Further tracing out the soft modes lead to

$$\rho_{\text{R,hard}} = \frac{1}{z} \sum_n e^{-\beta E_n} |n\rangle_{\text{RR}} \langle n|, \quad (6.6)$$

where  $z = \sum_n e^{-\beta E_n}$ . At the microscopic level, the two-sided black hole is a model of a single-sided black hole only in the sense that Eq. (6.6) takes the same form as Eq. (3.37).

Excited states in which there are objects in the right exterior (our side) of the black hole are obtained by acting with the annihilation/creation operators

$$b_{\text{R}\gamma} = \sum_n \sqrt{n_\gamma} | \{n_\alpha - \delta_{\alpha\gamma}\} \rangle_{\text{RR}} \langle \{n_\alpha\} |, \quad (6.7)$$

$$b_{\text{R}\gamma}^\dagger = \sum_n \sqrt{n_\gamma + 1} | \{n_\alpha + \delta_{\alpha\gamma}\} \rangle_{\text{RR}} \langle \{n_\alpha\} | \quad (6.8)$$

on the vacuum state in Eq. (6.3). As in the case of a single-sided black hole, describing the interior requires time evolution operator other than that generated by  $H = H_{\text{R}} - H_{\text{L}}$ . The appropriate generator  $\tilde{H}$  can be constructed using annihilation/creation operators acting on left-side hard modes

$$b_{\text{L}\gamma} = \sum_n \sqrt{n_\gamma} | \{n_\alpha - \delta_{\alpha\gamma}\} \rangle_{\text{LL}} \langle \{n_\alpha\} |, \quad (6.9)$$

$$b_{\text{L}\gamma}^\dagger = \sum_n \sqrt{n_\gamma + 1} | \{n_\alpha + \delta_{\alpha\gamma}\} \rangle_{\text{LL}} \langle \{n_\alpha\} | \quad (6.10)$$

as

$$\tilde{H} = \sum_\xi \Omega_\xi a_\xi^\dagger a_\xi + \tilde{H}_{\text{int}}(\{a_\xi\}, \{a_\xi^\dagger\}), \quad (6.11)$$

where

$$a_\xi = \sum_\gamma (\alpha_{\xi\gamma} b_{\text{R}\gamma} + \beta_{\xi\gamma} b_{\text{R}\gamma}^\dagger + \zeta_{\xi\gamma} b_{\text{L}\gamma} + \eta_{\xi\gamma} b_{\text{L}\gamma}^\dagger), \quad (6.12)$$

$$a_\xi^\dagger = \sum_\gamma (\beta_{\xi\gamma}^* b_{\text{R}\gamma} + \alpha_{\xi\gamma}^* b_{\text{R}\gamma}^\dagger + \eta_{\xi\gamma}^* b_{\text{L}\gamma} + \zeta_{\xi\gamma}^* b_{\text{L}\gamma}^\dagger) \quad (6.13)$$

are infalling mode operators with the coefficients  $\alpha_{\xi\gamma}$ ,  $\beta_{\xi\gamma}$ ,  $\zeta_{\xi\gamma}$ , and  $\eta_{\xi\gamma}$  taking the same values as those in Eqs. (5.23) and (5.24) for near horizon modes. The reason why the construction here need not involve soft modes is that the black hole vacuum state given in Eq. (6.3), in fact, factors into hard-mode and soft-mode parts:

$$|\Psi(M)\rangle \approx \left( \frac{1}{\sqrt{Z}} \sum_n e^{-\frac{\beta}{2}E_n} |n\rangle_R |n\rangle_L \right) \times \left( \frac{1}{\sqrt{\sum_j e^{-\beta E_j}}} \sum_i e^{-\frac{\beta}{2}E_i} |\psi_i\rangle_R |\psi_i\rangle_L \right), \quad (6.14)$$

so that the soft-mode piece can simply be ignored. Note that unlike the construction in Refs. [169,170], we do not claim that the generator  $\tilde{H}$  of infalling time evolution can be represented purely on the right operators.<sup>49</sup> The evolution with  $\tilde{H}$  in Eq. (6.11) allows us to describe the semiclassical physics in the domain of dependence of the union of zones on both sides at  $t = 0$ .

This erection of the effective theory without involving soft modes is essentially nothing other than the construction of the semiclassical theory from the beginning. The unknown UV physics, including the effect of ignoring horizon modes, is reflected in the choice of the vacuum in the effective theory. For example, for a large AdS black hole in thermal equilibrium with the ambient AdS space-time, the correct choice for typical soft-mode states is the Hartle-Hawking vacuum. This choice, however, cannot be derived from the low energy consideration alone.

We note that our starting point of Eq. (6.1) involves a choice of time. Suppose we evolve the state  $|\Psi(M)\rangle$  under the full, microscopic Hamiltonian  $H_R + H_L$ <sup>50</sup>:

$$|\Psi(M)\rangle \rightarrow e^{-i(H_R+H_L)t} |\Psi(M)\rangle = \frac{1}{\sqrt{\sum_{\mu'} e^{-\beta E_{\mu'}}}} \sum_{\mu} e^{-\frac{\beta}{2}(E_{\mu}+2it)E_{\mu}} |\mu\rangle_R |\mu\rangle_L. \quad (6.15)$$

In general, this state is not equivalent to  $|\Psi(M)\rangle$ ; the phases on the right-hand side have a physical meaning. For example, as  $t$  increases from 0, the maximal interior volume of a spatial surface anchored to the boundaries at time  $t$  grows, although for  $t \gtrsim e^{S_{\text{BH}}}$  the mixing between the hard and soft modes becomes so important that the naive semiclassical picture of interior volume growth ceases to apply [181].

In the two-sided state in Eq. (6.3), the von Neumann entropy of the right side is given by

<sup>49</sup>In the large  $N$  limit construction of Refs. [169,170], operators  $b_{R\gamma}$ ,  $b_{R\gamma}^\dagger$ ,  $b_{L\gamma}$ , and  $b_{L\gamma}^\dagger$  generate a type III<sub>1</sub> von Neumann algebra (which becomes type II<sub>∞</sub> at the next order in  $1/N$  [171]), so that Hilbert spaces for  $|n\rangle_R$ ,  $|n\rangle_L$ ,  $|\psi_i\rangle_R$ , and  $|\psi_i\rangle_L$  are not defined. This is not the case for finite  $N$  as envisioned here.

<sup>50</sup>It is important that this evolution is performed using the full, microscopic Hamiltonian (or equivalently performed in the boundary theory), since the analogous operator  $H_R + H_L$  in the bulk field theory is singular at the bifurcation surface.

$$S_R = -\text{Tr}[\rho_R \ln \rho_R] = \sum_n \sum_i \frac{e^{-\beta(E_n+E_i)}}{Z} [\beta(E_n + E_i) + \ln Z], \quad (6.16)$$

where  $\rho_R$  is given by Eq. (6.5). Approximating that the density of states of the soft modes does not depend on  $E_i$  and denoting it by  $S_{\text{soft}}(M)$ , we have up to  $O(1)$  coefficients

$$\sum_i \rightarrow \beta \int dE_i e^{S_{\text{soft}}(M)}, \quad (6.17)$$

where  $\beta$  is used to match the dimension. This gives

$$Z \sim \sum_n e^{S_{\text{soft}}(M) - \beta E_n} = z e^{S_{\text{soft}}(M)}, \quad (6.18)$$

and

$$S_R = \ln [z e^{S_{\text{soft}}(M)}] + \beta \langle E_n \rangle_R + O(1). \quad (6.19)$$

Here,

$$\langle E_n \rangle_R = \frac{1}{z} \sum_n E_n e^{-\beta E_n} \quad (6.20)$$

is the thermal energy of the hard modes in the right exterior, as measured by  $H$ . Thus, up to the thermal contribution of the hard modes (and an unimportant  $O(1)$  term), the entropy  $S_R$  agrees with  $S_{\text{sys}}$  given by Eqs. (3.30) and (3.31), which is the entropy of a single-sided black hole. Namely, for a two-sided black hole in the state of Eq. (6.3), the entropy of the black hole is given at the leading order by the entanglement entropy between the two sides (i.e., two CFTs in the AdS/CFT context).

### 1. Evaporation and the destruction of the wormhole

Let us now consider coupling the two-sided black hole discussed so far to a reservoir so that the black hole can radiate into it. We do this on the right side (i.e., our side) of the black hole and see its effect on an object falling into the black hole from the same side.

Before coupling the two systems, the state is given by the product of the black hole state in Eq. (6.3) and the ground state of the reservoir system  $|\phi_0\rangle$ :

$$|\Psi_0\rangle = \frac{1}{\sqrt{Z}} \sum_m \sum_j e^{-\frac{\beta}{2}(E_m+E_j)} |m\rangle_R |\psi_j\rangle_R |m\rangle_L |\psi_j\rangle_L |\phi_0\rangle, \quad (6.21)$$

where  $Z$  is given by Eq. (6.4). After the coupling of the two systems, this state is no longer an energy eigenstate, so it starts evolving. Specifically, the coupling injects positive energy shock waves into the two systems, after which the

Hawking emission process begins. This process backreacts and produces a superposition of black holes having different masses and momenta (at different locations) in the right-side space. As discussed in Sec. III A, we focus on a branch in which the black hole has a well-defined mass and (vanishing) momentum, within the minimum uncertainty required by quantum mechanics. We can then say that after time  $t$  is passed, the state for each  $(m, j)$  changes as<sup>51</sup>

$$|m\rangle_{\text{R}}|\psi_j\rangle_{\text{R}}|\phi_0\rangle \xrightarrow[\text{+ projection}]{\text{time evolution}} \sum_n \sum_{i_n=1}^{e^{S_{\text{bh}}(M_{\text{R}}-E_n)}} \sum_a c_{ni_n a}^{mj} |n\rangle_{\text{R}} \times |\psi_{i_n}^{(n)}\rangle_{\text{R}}|\phi_a^{(m,j)}\rangle, \quad (6.22)$$

where

$$\sum_n \sum_{i_n=1}^{e^{S_{\text{bh}}(M_{\text{R}}-E_n)}} \sum_a |c_{ni_n a}^{mj}|^2 = 1, \quad (6.23)$$

$M_{\text{R}}$  is the mass of the black hole as viewed from the right side, and  $|\phi_a^{(m,j)}\rangle$  represents the state of the reservoir. Here,  $|n\rangle_{\text{R}}|\psi_{i_n}^{(n)}\rangle_{\text{R}}$  and  $|\phi_a^{(m,j)}\rangle$  carry energies  $M_{\text{R}}$  and  $E_m + E_j - M_{\text{R}}$ , respectively (within the uncertainty), and we have assumed that  $t$  is not too large that the structure of the black hole is dramatically altered (for example, that it is fully evaporated). The index of  $|\psi_{i_n}^{(n)}\rangle_{\text{R}}$  now carries sub-index  $n$  because the energy of  $|\psi_{i_n}^{(n)}\rangle_{\text{R}}$  is correlated with that of  $|n\rangle_{\text{R}}$ .

Substituting Eq. (6.22) into Eq. (6.21), we obtain the state at time  $t$

$$|\Psi_t\rangle = \frac{1}{\sqrt{Z}} \sum_n \sum_m \sum_{i_n=1}^{e^{S_{\text{bh}}(M_{\text{R}}-E_n)}} \sum_j \sum_a e^{-\frac{\beta}{2}(E_m+E_j)} c_{ni_n a}^{mj} \times |n\rangle_{\text{R}}|\psi_{i_n}^{(n)}\rangle_{\text{R}}|m\rangle_{\text{L}}|\psi_j\rangle_{\text{L}}|\phi_a^{(m,j)}\rangle. \quad (6.24)$$

We assume that for  $t > t_{\text{scr}}$ , the coefficients  $c_{ni_n a}^{mj}$  take Gaussian random values across  $(n, i_n, a)$  for each  $(m, j)$ , as in Eq. (3.42), although this assumption is less justified than that for a single-sided black hole. Now, consider the (non-normalized) states multiplying  $|n\rangle_{\text{R}}$  in the sum over  $n$  in Eq. (6.24).

$$\|n\rangle_{\text{R},nn} = \frac{1}{\sqrt{Z}} \sum_m \sum_{i_n=1}^{e^{S_{\text{bh}}(M_{\text{R}}-E_n)}} \sum_j \sum_a e^{-\frac{\beta}{2}(E_m+E_j)} \times c_{ni_n a}^{mj} |\psi_{i_n}^{(n)}\rangle_{\text{R}}|m\rangle_{\text{L}}|\psi_j\rangle_{\text{L}}|\phi_a^{(m,j)}\rangle. \quad (6.25)$$

<sup>51</sup>We regard the left states  $|m\rangle_{\text{L}}|\psi_j\rangle_{\text{L}}$  as (approximate) eigenstates of the left boundary Hamiltonian so that they do not evolve.

With the assumption stated above, the norms of these states are given by

$$\begin{aligned} {}_{\text{R},nn}\langle\langle n||n\rangle\rangle_{\text{R},nn} &= \frac{1}{Z} \sum_m \sum_{i_n=1}^{e^{S_{\text{bh}}(M_{\text{R}}-E_n)}} \sum_j \sum_a e^{-\beta(E_m+E_j)} |c_{ni_n a}^{mj}|^2 \\ &= \frac{1}{(\sum_n e^{-\beta_{\text{R}}E_n})} e^{-\beta_{\text{R}}E_n}, \end{aligned} \quad (6.26)$$

up to corrections exponentially suppressed in  $S_{\text{bh}}(M_{\text{R}})$ . Here, in the last line we have used

$$\langle\langle c_{ni_n a}^{mj}|^2\rangle\rangle = \frac{1}{\sum_n \sum_a e^{S_{\text{bh}}(M_{\text{R}}-E_n)}} \quad (6.27)$$

$$\begin{aligned} \Rightarrow \sum_{i_n=1}^{e^{S_{\text{bh}}(M_{\text{R}}-E_n)}} \sum_a |c_{ni_n a}^{mj}|^2 &= \frac{e^{S_{\text{bh}}(M_{\text{R}}-E_n)}}{(\sum_n e^{S_{\text{bh}}(M_{\text{R}}-E_n)})} \\ &= \frac{1}{(\sum_n e^{-\beta_{\text{R}}E_n})} e^{-\beta_{\text{R}}E_n}, \end{aligned} \quad (6.28)$$

and

$$\beta_{\text{R}} = \left. \frac{\partial S_{\text{bh}}(E)}{\partial E} \right|_{E=M_{\text{R}}} \quad (6.29)$$

is the temperature of the black hole as viewed from the right side, which is different from  $\beta$ . The state  $\|n\rangle_{\text{R},nn}$  is thus related to the corresponding normalized state  $\|n\rangle_{\text{R}}$  by

$$\|n\rangle_{\text{R},nn} = \frac{1}{\sqrt{\sum_n e^{-\beta_{\text{R}}E_n}}} e^{-\frac{\beta_{\text{R}}}{2}E_n} \|n\rangle_{\text{R}}. \quad (6.30)$$

From Eqs. (6.25) and (6.30), we find that the state in Eq. (6.24) can be written in the thermofield double form

$$|\Psi_t\rangle = \frac{1}{\sqrt{\sum_n e^{-\beta_{\text{R}}E_n}}} \sum_n e^{-\frac{\beta_{\text{R}}}{2}E_n} |n\rangle_{\text{R}}\|n\rangle_{\text{R}}. \quad (6.31)$$

This allows us to erect the effective theory of the interior, following the construction described in Sec. V. In particular, it implies that an object located in the zone of the right side at time  $t$  will smoothly pass through the horizon (of the black hole of mass  $M_{\text{R}}$ ) from the right side.<sup>52</sup>

<sup>52</sup>We can also erect an effective theory as viewed from the left exterior, i.e., the opposite to the side from which the black hole evaporates. The construction is analogous to that described here—we identify  $\|m\rangle_{\text{L}}$  as the state that comes with  $|m\rangle_{\text{L}}$ , leading to the state in the effective theory

$$|\Psi'_t\rangle = \frac{1}{\sqrt{\sum_n e^{-\beta E_n}}} \sum_n e^{-\frac{\beta}{2}E_n} |n\rangle_{\text{L}}\|n\rangle_{\text{L}}, \quad (6.32)$$

where  $\beta = \partial S_{\text{bh}}(E)/\partial E|_{E=M}$ . The analysis presented below also applies to this case with similar conclusions.



It is interesting to consider what happens if the initial state has an excitation in the second exterior, i.e., the exterior on the left side. Suppose that the state at the time of the coupling,  $t = 0$ , has an excitation in the zone, e.g.,

$$|\Psi_{\text{init}}\rangle = \prod_{i=1}^N \left( \sum_{\gamma} f_{\gamma}^{(i)} b_{L\gamma}^{\dagger} \right) |\Psi_0\rangle, \quad (6.33)$$

where  $|\Psi_0\rangle$  is given by Eq. (6.21). In this case, the state after  $t > t_{\text{scr}}$  is given by Eq. (6.24) with the replacement

$$e^{-\frac{\beta}{2}E_m} c_{ni,a}^{mj} |\phi_a^{(m,j)}\rangle \rightarrow \sum_p e^{-\frac{\beta}{2}E_p} U_p^m c_{ni,a}^{pj} |\phi_a^{(p,j)}\rangle, \quad (6.34)$$

where  $U_p^m$  is a unitary matrix which has the indices  $p = \{p_{\alpha}\}$  and  $m = \{m_{\alpha}\}$  and depends on  $f_{\gamma}^{(i)}$ . Similarly, if there is an excitation on the left side that fell into the stretched horizon at an earlier time, then the state of the left-side soft modes at  $t = 0$  deviates from that in Eq. (6.21), causing the change of the state at time  $t$

$$e^{-\frac{\beta}{2}E_j} c_{ni,a}^{mj} |\phi_a^{(m,j)}\rangle \rightarrow \sum_k e^{-\frac{\beta}{2}E_k} V_k^j c_{ni,a}^{mk} |\phi_a^{(m,k)}\rangle, \quad (6.35)$$

where  $V_k^j$  is a unitary matrix acting on the space of the left-side soft states. With the changes in Eqs. (6.34) and (6.35), we can define  $\langle\langle n | n \rangle\rangle_{\text{R},nm}$  analogously to Eq. (6.25) and calculate its norm

$$\begin{aligned} & \text{R},nm \langle\langle n | n \rangle\rangle_{\text{R},nm} \\ &= \frac{1}{Z} \sum_m e^{\text{Sbh}(M_{\text{R}} - E_n)} \sum_{i_n=1} \sum_j \sum_{p,p'} \sum_{k,k'} \sum_{a,a'} e^{-\frac{\beta}{2}(E_p + E_{p'} + E_k + E_{k'})} \\ & \quad \times U_p^{m*} U_{p'}^m V_k^{j*} V_{k'}^j c_{ni,a}^{pk*} c_{ni,a'}^{p'k'} \langle \phi_a^{(p,k)} | \phi_{a'}^{(p',k')} \rangle \\ &= \frac{1}{(\sum_n e^{-\beta_{\text{R}} E_n})} e^{-\beta_{\text{R}} E_n}, \end{aligned} \quad (6.36)$$

which we find is the same as Eq. (6.26) up to exponentially suppressed corrections. Here, we have assumed that there is no intricate (and unexpected) cancellation between  $V_k^{j*} V_{k'}^j$  and  $c_{ni,a}^{pk*} c_{ni,a'}^{p'k'}$  in the sum over  $(k, k')$ , which would jeopardize the scaling in the last line.<sup>53</sup>

With this assumption, we thus find that the state of the effective theory is still given by Eq. (6.31). This implies that the existence of an excitation on the left exterior at  $t = 0$  cannot affect the effective theory erected at  $t > t_{\text{scr}}$  after evaporation began on the right side. In other words, the Einstein-Rosen bridge between the two sides is broken by

<sup>53</sup>A similar assumption is not needed for  $U_p^{m*} U_{p'}^m$ , since the number of degrees of freedom of semiclassical excitations is too small to jeopardize this scaling anyway.

Hawking radiation for  $t > t_{\text{scr}}$ , although an object falling into the black hole from either side sees smooth spacetime when it crosses the horizon.

At the technical level, this occurs because the energy constraint in Eq. (6.22) (i.e., the condition that the mass of the black hole as viewed from the right side is  $M_{\text{R}}$ ) breaks entanglement between the right and left side modes necessary to have a bridge [21]. To see this, we can trace out the soft and far modes in the state of Eq. (6.24)

$$\begin{aligned} \text{Tr}_{\text{soft+far}} |\Psi_t\rangle \langle \Psi_t| &= \frac{1}{(\sum_{m'} e^{-\beta_{\text{R}} E_{m'}}) (\sum_{n'} e^{-\beta E_{n'}})} \\ & \quad \times \sum_m \sum_n e^{-\beta_{\text{R}} E_m} e^{-\beta E_n} |m\rangle_{\text{R}} \langle n|_{\text{LR}} \langle m|_{\text{L}} \langle n| \end{aligned} \quad (6.37)$$

and find that it takes a different form than

$$\begin{aligned} \text{Tr}_{\text{soft+far}} |\Psi_0\rangle \langle \Psi_0| &= \frac{1}{\sum_{n'} e^{-\beta E_{n'}}} \sum_m \sum_n e^{-\frac{\beta}{2}(E_m + E_n)} \\ & \quad \times |m\rangle_{\text{R}} \langle m|_{\text{LR}} \langle n|_{\text{L}} \langle n|, \end{aligned} \quad (6.38)$$

obtained from the thermofield double state which has a connected Einstein-Rosen bridge. In particular, we see that the hard modes on the right side are mostly entangled with the right-side soft modes and far modes in  $|\Psi_t\rangle$ , and not with modes on the left side as in  $|\Psi_0\rangle$ .

Physically, this breaking of the Einstein-Rosen bridge occurs due to decoherence caused by Hawking emission from the right side. After the Hawking emission, black holes of different masses (as viewed from the right side) can be semiclassically discriminated, so that they should be viewed as living in different branches of the wave function. Thus, unless a falling object somehow preserves coherence among these different branches,<sup>54</sup> it cannot see any signal sent from the left exterior inside the black hole. In general, the Einstein-Rosen bridge—or wormhole—prepared by the thermofield double state in Eq. (6.3) is fragile under a realistic physical process occurring to it.<sup>55</sup>

## 2. Entangled black holes

For a reason similar to the case above, entangled black holes of the form

<sup>54</sup>It is not clear to what extent this is possible because the falling object necessarily feels gravity generated by the black hole in question.

<sup>55</sup>If the decoherence is imperfect, the two sides can be “weakly connected,” e.g., connected in the interior but with a physical domain wall between the two horizons.

$$|\Psi(M_1, M_2)\rangle = \sum_{A=1}^{e^{S_{\text{sys}}(M_1)}} \sum_{B=1}^{e^{S_{\text{sys}}(M_2)}} \eta_{AB} |\Psi_A(M_1)\rangle |\Psi_B(M_2)\rangle \quad (6.39)$$

generically do not have an Einstein-Rosen bridge connecting them; an object falling into one of the black holes sees a smooth horizon but cannot receive any signal sent in from the other black hole [21]. The same comment applies to more than two entangled black holes.

### 3. Interior holography?

We now discuss the possibility of formulating a “holographic theory of the interior.” In the discussion so far, the fundamental degrees of freedom are assumed to live on boundaries (e.g., holographic screens or AdS boundaries) outside the black hole horizon. However, discussion in Sec. IV suggests that a holographic theory may be able to live on an apparent horizon, the surface on which one of the (quantum) expansions vanishes. We may, therefore, speculate that a holographic theory for the interior can be formulated by picking out the degrees of freedom relevant for describing the interior and distributing them on the horizon; see Fig. 16.

As seen in Fig. 16, the time evolution of this theory occurs toward the future on both arms of the horizon, so that the holographic space at a given time consists of two

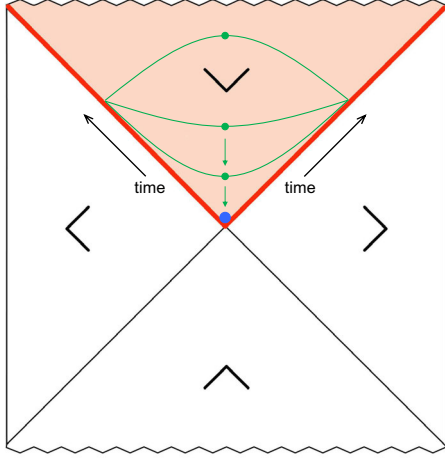


FIG. 16. Holographic theory of the interior may be constructed by distributing the interior degrees of freedom over the horizon  $H(t)$ , which consists of two marginally trapped surfaces as indicated by the wedges. Time  $t$  evolves toward the future on both components of the horizon. Entanglement entropy between the two sides is computed by finding the maximin surface, obtained by minimizing the area of a surface homologous to a component of the horizon on a Cauchy surface anchored to  $H(t)$  (green dots) and then maximizing it over possible Cauchy surfaces, which leads to the bifurcation surface (blue dot). The entanglement entropy is thus given by the Bekenstein-Hawking entropy of the black hole.

disconnected two-spheres. One can then calculate entanglement entropy between the left and right sides of the black hole by finding a maximin surface [152] homologous to one of the two-spheres, which turns out to be the bifurcation surface (see Fig. 16). The entanglement entropy is thus given by

$$S = \frac{\mathcal{A}_{\text{bh}}}{4G_N}, \quad (6.40)$$

where  $\mathcal{A}_{\text{bh}}$  is the area of the horizon. The proposed theory, therefore, passes one of the simplest checks for consistency.

It is not clear if the theory described here is fully consistent or useful. For example, the continuous renormalization procedure in Refs. [13,14] cannot be used here to describe the interior, since the procedure requires the bulk to be normal (while it is trapped here). We have discussed this nonetheless, since it is related to similar proposals in de Sitter spacetime [45–48], which we will address later.

### B. Global de Sitter spacetime

We now discuss a possible description of global de Sitter spacetime analogous to a static, two-sided black hole. As in the case of a two-sided black hole in asymptotically flat spacetime, we consider it to be an approximate description of a sufficiently long-lived metastable de Sitter spacetime.

Specifically, we assume that global de Sitter spacetime at  $t = 0$  can be described by a state of a thermofield doublelike form between two holographic theories each of which describes a static patch. Following the case of a two-sided black hole, we may divide modes in each theory into “zone” and horizon modes. We then take a subset of the zone modes to be hard modes while leaving all the other modes as soft modes. In the case of de Sitter spacetime, however, we expect that gravity is not decoupling in the holographic theories (because the boundary is not in an asymptotic region; see Sec. IV B), so that the two theories are interacting through it.<sup>56</sup> We thus denote by  $|n\rangle_N$ ,  $|\psi_i\rangle_N$ ,  $|n\rangle_S$ , and  $|\psi_i\rangle_S$  the states of the hard and soft modes in the limit that gravity is turned off in these holographic theories. Here, the subscripts N and S specify the theory under consideration (referring to the north and south hemispheres, respectively).

Assuming that the two theories have the same microscopic structure, the relevant thermofield double state can be written as

$$\begin{aligned} |\Psi'(E)\rangle &\approx \frac{1}{\sqrt{Z}} \sum_n \sum_i e^{-\frac{\beta}{2}(E_n + E_i)} |n\rangle_N |\psi_i\rangle_N |n\rangle_S |\psi_i\rangle_S, \\ Z &= \sum_n \sum_i e^{-\beta(E_n + E_i)}, \end{aligned} \quad (6.41)$$

<sup>56</sup>By gravity we mean the full dynamics associated with gravity at short and long distances. Note that in  $2 + 1$  dimensions, there is no massless propagating graviton.

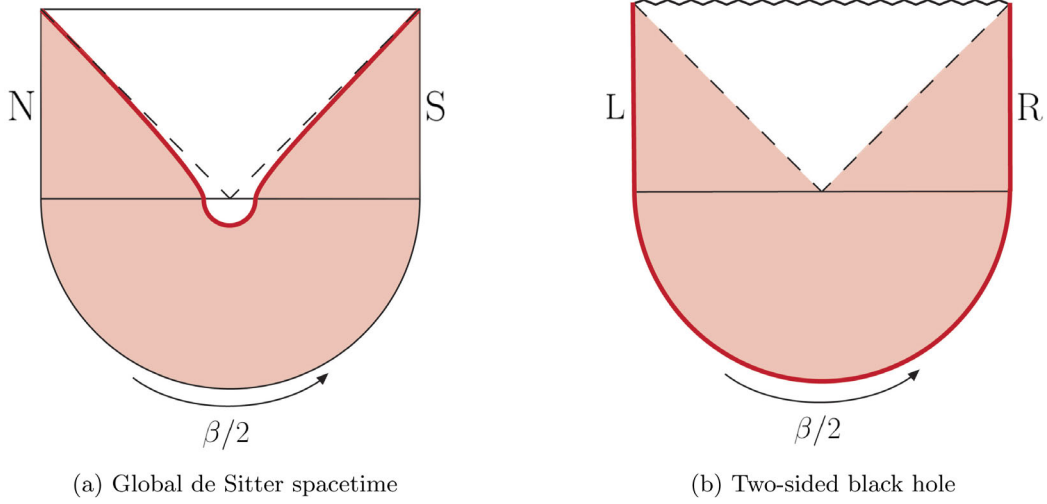


FIG. 17. Preparation of a holographic state of the thermofield double form by Euclidean path integral. In each panel, the thick red line represents the location at which the holographic theory resides, while the shaded region corresponds to the bulk region that can be simply reconstructed, e.g., by the procedure of continuously pulling in the boundary.

in the limit that gravity is turned off in holographic theories. Here,  $E = \alpha/l_p^2$ , and  $E_n$  and  $E_i$  are the energies carried by the hard and soft modes in the north hemisphere defined with respect to the Hamiltonian  $H = H_N - H_S$ . We expect that the state in Eq. (6.41) can be prepared by the Euclidean path integral over a half of the time compactified on a circle of length  $\beta$  (in the limit that gravity is turned off); see Fig. 17(a). While the location of the boundary is now reversed compared with the case of a black hole, depicted in Fig. 17(b), we assume that the periodicity  $\beta$  of the time direction is still related to the temperature of de Sitter spacetime as  $\beta = 1/T_H = 2\pi\alpha$ , as in Ref. [179].

When gravity is turned on in holographic theories, the Euclidean path integral preparing the state is expected to receive extra contributions suppressed by the gravitational coupling which are not diagonal in the space spanned by  $|n\rangle|\psi_i\rangle$ . This gives the correction to the state of the form

$$\begin{aligned} |\Psi'(E)\rangle &\rightarrow |\Psi(E)\rangle \\ &= |\Psi(E)\rangle + \sum_{m,n} \sum_{j,i} \varepsilon_{m,n;j,i} |m\rangle_N |\psi_j\rangle_N |n\rangle_S |\psi_i\rangle_S, \end{aligned} \quad (6.42)$$

where  $\varepsilon_{m,n;j,i}$  are coefficients of order  $l_p^2/\alpha^2$ . We suspect that these off-diagonal parts are related to the fact that, unlike the case of a black hole, a positive energy shockwave in de Sitter spacetime gives a traversable “wormhole” between the two hemispheres [173,176].

Excited states in which there are objects in the north and south hemispheres of the global de Sitter spacetime can be obtained by acting annihilation/creation operators, given by Eqs. (6.7)–(6.10) with  $R \rightarrow N$  and  $L \rightarrow S$ , on the vacuum state in Eq. (6.42). Describing the region outside the horizons of the two static patches requires time evolution

operator other than that generated by  $H = H_N - H_S$ . The appropriate generator  $\tilde{H}$  can be constructed as

$$\tilde{H} = \sum_{\xi} \Omega_{\xi} a_{\xi}^{\dagger} a_{\xi} + \tilde{H}_{\text{int}}(\{a_{\xi}\}, \{a_{\xi}^{\dagger}\}), \quad (6.43)$$

where

$$a_{\xi} = \sum_{\gamma} (\alpha_{\xi\gamma} b_{R\gamma} + \beta_{\xi\gamma} b_{R\gamma}^{\dagger} + \zeta_{\xi\gamma} b_{L\gamma} + \eta_{\xi\gamma} b_{L\gamma}^{\dagger}), \quad (6.44)$$

$$a_{\xi}^{\dagger} = \sum_{\gamma} (\beta_{\xi\gamma}^* b_{R\gamma} + \alpha_{\xi\gamma}^* b_{R\gamma}^{\dagger} + \eta_{\xi\gamma}^* b_{L\gamma} + \zeta_{\xi\gamma}^* b_{L\gamma}^{\dagger}) \quad (6.45)$$

are mode operators with the coefficients  $\alpha_{\xi\gamma}$ ,  $\beta_{\xi\gamma}$ ,  $\zeta_{\xi\gamma}$ , and  $\eta_{\xi\gamma}$  determined by semiclassical calculation.

Since the  $t = 0$  hypersurface is a Cauchy surface, this effective theory describes the entire global de Sitter spacetime. The theory, however, is intrinsically semiclassical. The choice of the vacuum cannot be derived from the first principle, although we expect that the correct choice corresponding to typical soft-mode states is the Bunch–Davies vacuum [182]. The description of field fluctuations on the de Sitter background is also only statistical. To go beyond this, e.g., to describe the details of a state depending also on the horizon microstate, we must resort to the microscopic theory.

In the  $|\Psi(E)\rangle$  state, the von Neumann entropy of the north hemisphere is

$$S_N = \sum_n \sum_i \frac{e^{-\beta(E_n + E_i)}}{Z} [\beta(E_n + E_i) + \ln Z] + O\left(\frac{l_p^2}{\alpha^2}\right). \quad (6.46)$$

Assuming that the density of states of the soft modes does not depend on  $E_i$  and denoting it by  $S_{\text{soft}}(E)$ , we have

$$S_N = \ln [z e^{S_{\text{soft}}(E)}] + \beta \langle E_n \rangle_N + O(1), \quad (6.47)$$

where  $z = \sum_n e^{-\beta E_n}$ , and  $\langle E_n \rangle_N = \sum_n E_n e^{-\beta E_n} / z$  is the thermal energy of the hard modes in the north hemisphere, as measured by  $H$ . Thus, the entropy  $S_N$  agrees with  $S_{\text{sys}}$  given by Eqs. (3.30) and (3.31) at the leading order. Namely, in the global state of Eq. (6.42), the entropy of de Sitter spacetime can be interpreted as the entanglement entropy between the two hemispheres, at the leading order.

### 1. Relation to the DS/dS correspondence

The theory we are considering consists of two holographic systems located on the stretched horizons of two static patches, each of which covers one half of the spatial section of global de Sitter spacetime at  $t = 0$ , the global time at which the spatial volume becomes minimal. These two systems, each of which is expected to be strongly coupled, are weakly interacting through gravity. This structure is reminiscent of that in the DS/dS correspondence [43,44], a proposed holographic description of de Sitter spacetime.

While the two theories have similar structures at  $t = 0$ , they can be different at  $t \neq 0$ . In particular, if we evolve our theory with  $H_N + H_S$ , the places where the holographic systems are located move toward the future along the stretched horizons of the two static patches, as indicated by the red drawings in Fig. 18. This makes simply reconstructed regions stay within the static patches. On the other hand, in the DS/dS correspondence, each of the holographic systems is completed itself into de Sitter spacetime of one lower dimensions, leading to the simply reconstructed region called a DS/dS patch, which is depicted by the blue drawings in the figure.

One might speculate that the two theories represent the same system evolved differently in holographic space. It will be interesting to study this possible relation, but we leave it for the future.

### 2. Relation to the Shaghoulian-Susskind proposal

At the end of Sec. VI A, we discussed the possibility of having a holographic theory of the black hole interior by placing the relevant degrees of freedom on the apparent horizon. We can consider an analogous situation for de Sitter spacetime, in which case the degrees of freedom relevant for describing the region outside the horizons of the two static patches are placed on the horizon; see Fig. 19.

As in the case of the black hole, we can test the consistency of this picture by computing the entanglement entropy between the two patches, which can be done using the maximin procedure. By minimizing a surface homologous to one of the horizons on a spacelike surface bounded

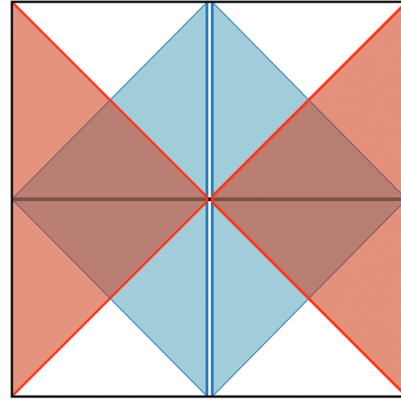


FIG. 18. The holographic theory based on static patches consists of two weakly interacting systems located on the stretched horizons of two static patches, with the simply reconstructed region staying within the two patches (red). In the DS/dS correspondence, each of the holographic systems is located by itself on de Sitter spacetime with one lower dimensions, leading to the simply reconstructed region called the DS/dS patch (blue). The structures of the two theories are similar at  $t = 0$ .

by the two horizons at a fixed time, we find the resulting surface is nothing other than one of the horizons. The subsequent maximization over spacelike surfaces thus gives us the same surface, so the de Sitter horizon is the maximin surface. This gives the entanglement entropy

$$S = \frac{\mathcal{A}_{\text{dS}}}{4G_N}, \quad (6.48)$$

where  $\mathcal{A}_{\text{dS}}$  is the area of the de Sitter horizon, which is consistent with the picture that the evolution of the system is unitary. This seems to be what is proposed in Refs. [45,46]

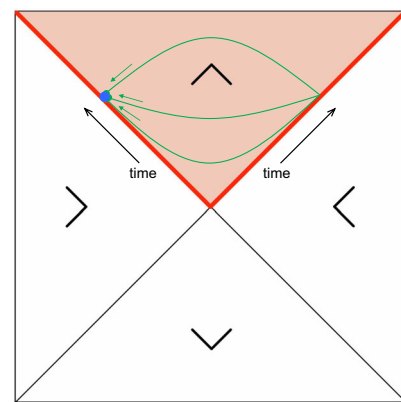


FIG. 19. Holographic theory placed on the de Sitter horizons of two static patches can describe the region outside the horizons (as viewed by polar observers). Entanglement entropy between the two sides is computed by finding the maximin surface, which is nothing other than one of the horizons (blue dot). The entanglement entropy, therefore, is given by the Gibbons-Hawking entropy.



as a holographic theory of de Sitter spacetime (referred to as the monolayer proposal in Ref. [48]), although our interpretation here says that the theory describes only the shaded region in Fig. 19; in particular, it does not describe the regions inside the two static patches.

In our picture, the degrees of freedom used for the theory described above are a subset of the degrees of freedom of holographic theories describing the static patches. Therefore, the former are not really independent of the latter at the microscopic level. However, we may treat them independent for the purpose of describing semiclassical physics in the bulk. If this is the case, we can regard the holographic theory as consisting of two layers of degrees of freedom on the horizons, one describing the outside and the other describing the interiors of the horizons. This seems to be the proposal of Ref. [47], called the bilayer proposal in Ref. [48].

## VII. GRAVITATIONAL PATH INTEGRAL

The approach we have described so far is based on the canonical formalism of quantum mechanics. In particular, we have assumed the existence of a Hilbert space factor representing states of the horizon degrees of freedom, with which we could construct an effective theory describing a spacetime region behind the horizon—the interior in the case of a black hole—using only the low energy input that the dynamics of these degrees of freedom is maximally chaotic (and fast scrambling).

Quantum mechanics, however, can also be formulated using path integrals, and we expect that the same physical conclusions would be obtained from this formalism. In this section, we discuss what the picture looks like in this case. We will see that the corresponding picture is that of Refs. [36,37], in which aspects of unitary evolution can be reproduced by gravitational path integral that fixes the “boundary condition” based on the quantity of interest and then integrates over all possible semiclassical geometries consistent with it, including those with nontrivial topologies (in particular, replica wormholes). This allows us to relate the framework described so far to the treatment based on the quantum extremal surface prescription [33–35]. While we focus on the case of a black hole here, we suspect that a similar story can be developed for de Sitter spacetime as well.

The picture presented here was outlined in Refs. [24,38]. The ensemble nature of the semiclassical description arising from an ensemble of microstates was also discussed in Refs. [60–64]. The understanding of the Page curve presented here is based on the picture developed in Refs. [65–71].

### A. Ensemble from coarse graining

The starting point for the path integral formalism is very different from that of the canonical formalism. Specifically,

in the present context it should start from a collection of classical field configurations on classical geometries, which will then be integrated over. We are interested in a “low energy” framework in which the detailed microscopic knowledge is not necessary to understand the physics. This implies that a black hole must be treated as a (semi) classical object in which the detailed microscopic structure cannot be discriminated.

This treatment, in fact, is required by quantum mechanics. As we have argued in Secs. II and III, microstates of a black hole can be regarded as independent quantum states arising from superposing energy eigenstates in a small energy window, e.g., of order  $T_H$ , around  $M$ . Thus, to discriminate these microstates, one would need an exponentially long measurement time  $\sim e^{S_{\text{bh}}}/T_H$ . The black hole, however, would already have evaporated by the time such a measurement would be completed (or the state of the black hole will be altered significantly for a large AdS black hole thermalized with the environment). One therefore cannot operationally discriminate different microstates by performing a measurement on the black hole.<sup>57</sup>

We thus have to regard a black hole spacetime appearing in gravitational path integral as representing a maximally mixed ensemble of the microstates consistent with the classical specification of the black hole [24,38]. Note that this picture is associated with a single-sided (or a single-sided description) of a black hole. To see this, one can imagine preparing a black hole state using Euclidean path integral in the approximation that the black hole is static. As represented in Fig. 20, the “ $\psi_+ - \psi_-$  component” of the density matrix of the exterior region can then be obtained by the path integral in which the spatial field configurations are fixed to be  $\psi_+$  and  $\psi_-$  above and below the cut corresponding to the exterior region of the black hole.

At the semiclassical level, this simply gives the thermal density matrix which we regard as the semiclassical black hole vacuum state. We can, however, imagine following the same procedure at the microscopic level to get the corresponding density matrix

$$\rho_{\text{micro}} \propto \sum_n e^{-\frac{\beta}{2}\Delta H} |n\rangle \langle n| e^{-\frac{\beta}{2}\Delta H}, \quad (7.1)$$

where  $\{|n\rangle\}$  is a complete set of black hole vacuum microstates consistent with the background used in the semiclassical path integral,  $\beta = 1/T_H$  is the periodicity of Euclidean time in the angular direction, and  $\Delta H$  is the microscopic exterior (boost) Hamiltonian. Here, the zero of  $\Delta H$  is chosen to be a typical energy associated with the space  $\mathcal{H}_{\text{vac}}$  spanned by  $|n\rangle$ 's.<sup>58</sup>

<sup>57</sup>One can, of course, *infer* the microstate of the black hole if we prepare a specific state of initial collapsing matter and then use the microscopic theory to simulate its time evolution.

<sup>58</sup>The index  $n$  here corresponds to the index  $A$  in the notation in Secs. III and V.

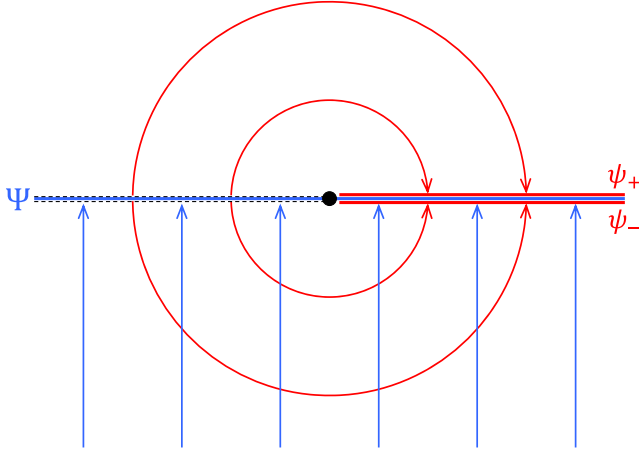


FIG. 20. The density matrix of the exterior region, whose element is specified by the field configurations  $\psi_+$  and  $\psi_-$ , can be computed by Euclidean path integral, which can be interpreted to receive equal contributions from all the black hole microstates. Euclidean path integral can also be used to prepare a two-sided state  $\Psi$ , which is pure and unique.

Since the matrix elements of the  $\beta\Delta H/2$  operator in  $\mathcal{H}_{\text{vac}}$  are of order or smaller than 1, we find that many microstates contribute to  $\rho_{\text{micro}}$ . In particular, there are exponentially many microstates that contribute dominantly and almost equally; these are the microstates among which the matrix elements  $\beta\Delta H/2$  is much smaller than 1. The fact that semiclassical theory cannot resolve the detailed structure of microstates implies that it perceives  $\rho_{\text{micro}}$  to be the maximally mixed state<sup>59</sup>

$$\rho_{\text{sc}} \propto \sum_{n=1}^{e^{S_{\text{bh}}}} |n\rangle\langle n|, \quad (7.2)$$

where the sum runs over a sufficiently large set of vacuum microstates  $|n\rangle$ , whose precise specification is not important as discussed in Sec. III (or as in standard statistical mechanics). The number of elements of this set is of order  $e^{S_{\text{bh}}}$ , which we have already indicated in Eq. (7.2).

This interpretation indeed reproduces many features which are attributed to the ensemble nature of holographic theories in lower dimensional quantum gravity [36] in terms of an ensemble of microscopic states [24] (see also Refs. [60–64]). Incidentally, this should be contrasted with the situation in which a global (two-sided) state is prepared by Euclidean path integral. In this case, the generated state is a pure state

<sup>59</sup>The contributions to  $\rho_{\text{micro}}$  from states with  $|\beta\Delta H| \gg 1$  are negligible either because of the Boltzmann suppression (for  $\Delta H > 0$ ) or because the number of states is too small (for  $\Delta H < 0$ ). The chaotic nature of the black hole dynamics then implies that the remaining contribution can be well approximated to come from the maximally mixed ensemble of microstates with  $|\beta\Delta H| \lesssim 1$ .

$$|\Psi_{\text{micro}}\rangle \propto \lim_{\tau \rightarrow \infty} e^{-\Delta\tilde{H}\tau}|i\rangle \quad (7.3)$$

obtained by evolving some initial state  $|i\rangle$  at  $\tau = -\infty$  to  $\tau = 0$ . Here,  $|i\rangle$  is chosen in the space of black hole (not necessarily vacuum) microstates  $\mathcal{H}_{\text{bh}}$ , and  $\Delta\tilde{H}$  is the microscopic “infalling” (inertial) Hamiltonian with its zero chosen to be a typical energy of vacuum microstates. This leads to the unique, lowest energy eigenstate, determined by the choice of  $\mathcal{H}_{\text{bh}}$ ; see also discussion around Eq. (6.1).

## B. Replica method, entanglement island, and the Page curve

With the understanding that the black hole spacetime appearing in gravitational path integral represents the (maximally mixed) ensemble of black hole microstates, the results of Refs. [36,37] can be understood in a simple manner.

### 1. Euclidean gravitational path integral

Let us first consider a setup similar to Ref. [37] in which a black hole living in a gravitational region is coupled to a nongravitational region.<sup>60</sup> The black hole then radiates into the nongravitational region. Our interest is to compute the von Neumann entropy of the emitted radiation.

Since the radiation lives in a nongravitational region, we must be able to describe its state semiclassically. In particular, assuming that the system can be viewed as quasistatic at each instance of time, we expect that the density matrix of the radiation in a region  $R$  can be calculated using a Euclidean path integral by specifying its element by spatial configurations of the radiation field above and below the cut on  $R$ ; see Fig. 21(a). The obtained density matrix is not normalized, which we denote by  $\hat{\rho}_R$  with the hat over  $\rho_R$  indicating that the density matrix is not normalized.<sup>61</sup>

The path integral performed, however, includes the gravitational region in which a black hole resides. Thus, the obtained density matrix, in fact, involves an ensemble average over the black hole microstates in the sense of Eq. (7.2)

<sup>60</sup>The nongravitational region is a proxy for a weakly gravitational region. The distinction between gravitational and nongravitational regions is necessary to correctly identify semiclassical degrees of freedom in the latter region, in particular Hawking radiation quanta, of which the von Neumann entropy is calculated. Without this separation, the Page curve of our interest cannot be computed using the method adopted here [183], although we expect that there is a way to reproduce the same result without involving an artificial separation of regions, at least in principle.

<sup>61</sup>The density matrix  $\rho_R$  here refers to the fine-grained (reduced) density matrix of fields in  $R$ . In the language of Ref. [165], the radiation here is the “radiation in boldface.”

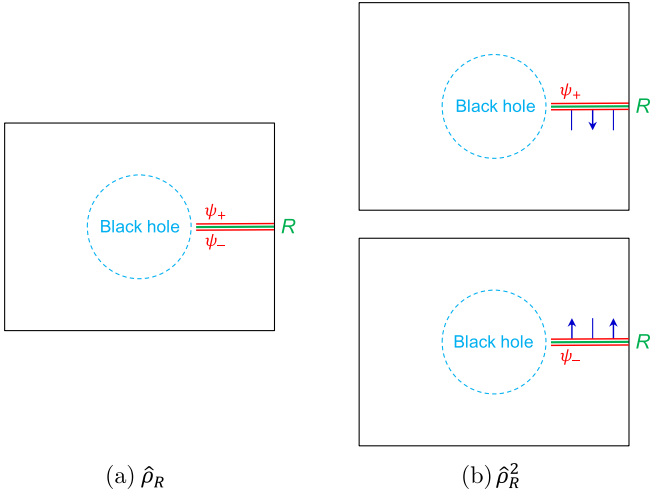


FIG. 21. (a) The (non-normalized) density matrix  $\hat{\rho}_R$  of radiation in nongravitational region  $R$  can be calculated by performing path integral with the boundary condition fixing the spatial field configurations above and below the cut along  $R$ . (b) The square of the density matrix,  $\hat{\rho}_R^2$  is calculated using the replica method, which replicates the spacetime into two and imposes the boundary condition such that one side of the cuts along  $R$  on the two sheets are appropriately sewn together (as represented by blue arrows).

$$\overline{\hat{\rho}_R} = \frac{1}{e^{S_{\text{bh}}}} \sum_{n=1}^{e^{S_{\text{bh}}}} \hat{\rho}_{R,n}, \quad (7.4)$$

where  $\hat{\rho}_{R,n} = \text{Tr}_{R^c} |n\rangle\langle n|$  represents the (non-normalized) density matrix of  $R$  for the microstate  $|n\rangle$  with  $R^c$  being the complement of  $R$ . Defining the normalized density matrix by

$$\overline{\rho_R} = \frac{\overline{\hat{\rho}_R}}{\text{Tr} \overline{\hat{\rho}_R}}, \quad (7.5)$$

we can compute the von Neumann entropy of  $R$  as

$$S_R^{(\text{sc})} = -\text{Tr}[\overline{\rho_R} \ln \overline{\rho_R}]. \quad (7.6)$$

Since this involves coarse-graining, i.e., the ensemble average over microstates, the quantity  $S_R^{(\text{sc})}$  is well approximated by the thermal entropy of the radiation, which increases monotonically in time (until the black hole is fully evaporated, if it is not eternal). This is Hawking's result [1] showing the apparent violation of unitarity in the black hole evaporation process.

However, what we really want to understand is the behavior of

$$S_R^{(\text{micro})} = -\text{Tr}[\rho_{R,n} \ln \rho_{R,n}] \quad (7.7)$$

(for each microstate  $|n\rangle$ ), which must show the Page behavior [20] if the evolution of the system is unitary.

In particular,  $S_R^{(\text{micro})}$  must go down to zero when the black hole is fully evaporated and all the emitted radiation is included in  $R$ . A question is how (or if) we can see this behavior in a semiclassical analysis.

The idea is that while the semiclassical description necessarily involves the ensemble average over black hole microstates, it still allows for calculating the ensemble average of many different quantities. For example, we can calculate the ensemble average of the square of the density matrix for the radiation using the replica method [89,184,185], i.e., by replicating the spacetime into two copies and imposing the boundary condition for the path integral such that (one side of) the cuts along  $R$  on two sheets are appropriately sewn together; see Fig. 21(b). Performing path integral with this boundary condition gives

$$\overline{\hat{\rho}_R^2} = \frac{1}{e^{S_{\text{bh}}}} \sum_{n=1}^{e^{S_{\text{bh}}}} \hat{\rho}_{R,n}^2. \quad (7.8)$$

An important point is that the performed path integral must involve all possible geometries in the gravitational region, including those having a topology that geometrically connects the two sheets (the replica wormhole) [36,37]. This is because the gravitational path integral should not predetermine the geometry on which quantum fields are integrated over.<sup>62</sup> In general, the inclusion of geometries with nontrivial topologies makes Eq. (7.8) different from the square of Eq. (7.4):  $\overline{\hat{\rho}_R^2} \neq (\overline{\hat{\rho}_R})^2$ .

In Fig. 22, we illustrate the calculation of the trace of the square of the non-normalized density matrix  $\hat{\rho}_R$ , which we denote by  $Z_2 = \text{Tr} \hat{\rho}_R^2$ . As found in Refs. [36,37], there are two contributions to this. The first is the one in which the gravitational region is filled separately for two sheets; Fig. 22(a). Restricting the radiation configuration to those resembling the emitted Hawking radiation at a coarse-grained level, the integral of semiclassical fields provides a factor of  $e^{S_{\text{rad}}}$ , since the relevant four spatial configurations of the fields (both sides of the two cuts) are all related by Euclidean evolution as indicated by the red arrows in the figure. On the other hand, the gravitational path integral gives  $e^{S_{\text{bh}}}$  for each sheet, so that this contribution is given by

$$Z_2^{(\text{disconnected})} \sim e^{2S_{\text{bh}} + S_{\text{rad}}}. \quad (7.9)$$

<sup>62</sup>As elucidated, e.g., in Ref. [70], this prescription calculates  $\overline{\hat{\rho}_R^2}$  with exponential accuracy, taking into account the detailed microscopic structure of radiation (radiation in boldface in the language of Ref. [165]). If we instead calculate  $\overline{\hat{\rho}_R^2}$  including only trivial topology, i.e., the disconnected contribution, then we would fail to capture the effect from the exponentially complex microscopic structure, yielding  $\overline{\hat{\rho}_R^2}$  as computed in the semiclassical theory (i.e., the squared density matrix of nonboldface radiation). This latter prescription would lead to Hawking's result [1] when we compute the von Neumann entropy using  $S_R = -\lim_{n \rightarrow 1} \partial(\text{Tr} \overline{\rho_R^n}) / \partial n$ , where  $\overline{\rho_R^n} = \overline{\hat{\rho}_R^n} / (\text{Tr} \overline{\hat{\rho}_R})^n$ .

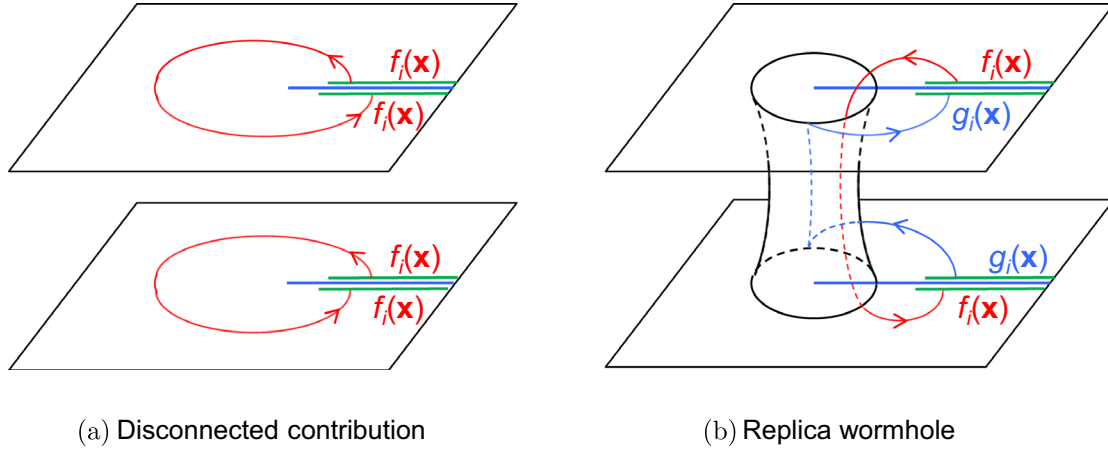


FIG. 22. Two contributions to replica method calculation of  $\overline{\text{Tr}\hat{\rho}_R^2}$ . (a) The normal contribution in which the gravitational region is filled separately on each sheet. Integration of semiclassical matter gives  $e^{S_{\text{rad}}}$  coming from Euclidean evolution relating all 4 field configurations above and below the cuts on two sheets (red arrows), while gravitational path integral gives the  $e^{S_{\text{bh}}}$  factor for each sheet. (b) For the replica wormhole configuration, the gravitational regions on two sheets are geometrically connected. The contribution from gravitation path integral is thus only  $e^{S_{\text{bh}}}$ , while the contribution from matter integration gives  $e^{2S_{\text{rad}}}$  since there are two independent cycles for evolution (red and blue arrows).

The other contribution is the one coming from a replica wormhole; Fig. 22(b). In this case, the radiation contribution is  $e^{2S_{\text{rad}}}$ , since we now have two independent cycles for the evolution, so that the four relevant configurations are related only pairwise (the red and blue arrows). On the other hand, the gravitational contribution is now  $e^{S_{\text{bh}}}$  because it comes from a single, connected component. This therefore gives

$$Z_2^{(\text{wormhole})} \sim e^{S_{\text{bh}}+2S_{\text{rad}}}. \quad (7.10)$$

Adding the two contributions together, we find that  $Z_2 \sim \max\{e^{2S_{\text{rad}}+S_{\text{bh}}}, e^{S_{\text{bh}}+2S_{\text{rad}}}\}$ .

A similar analysis can be performed for  $Z_n$  for  $n \in \mathbb{N}$ . Assuming that the replica symmetric wormhole connecting all  $n$  sheets dominates for  $S_{\text{rad}} > S_{\text{bh}}$ , we find

$$Z_n \sim \begin{cases} e^{nS_{\text{bh}}+S_{\text{rad}}} & \text{for } S_{\text{bh}} > S_{\text{rad}} \\ e^{S_{\text{bh}}+nS_{\text{rad}}} & \text{for } S_{\text{rad}} > S_{\text{bh}}. \end{cases} \quad (7.11)$$

This is illustrated in Fig. 23. Since the density matrix  $\rho_R$  is given by  $\hat{\rho}_R/\text{Tr}\hat{\rho}_R$ , we thus find

$$\begin{aligned} \overline{\text{Tr}\rho_R^n} &= \frac{\overline{\text{Tr}\hat{\rho}_R^n}}{(\overline{\text{Tr}\hat{\rho}_R})^n} \approx \frac{\overline{\text{Tr}\hat{\rho}_R^n}}{(\overline{\text{Tr}\hat{\rho}_R})^n} = \frac{Z_n}{Z_1^n} \\ &\sim \begin{cases} e^{(1-n)S_{\text{rad}}} & \text{for } S_{\text{bh}} > S_{\text{rad}} \\ e^{(1-n)S_{\text{bh}}} & \text{for } S_{\text{rad}} > S_{\text{bh}}, \end{cases} \end{aligned} \quad (7.12)$$

where in the second equation we have used the fact that the standard deviation of  $\text{Tr}\hat{\rho}_R^n$  is smaller than its typical size in the ensemble, which we can easily be convinced. By analytically continuing this result in  $n$ , we can now obtain

the ensemble average of the von Neumann entropy of the radiation

$$\overline{S_R} = -\lim_{n \rightarrow 1} \frac{\partial}{\partial n} \overline{\text{Tr}\rho_R^n} \sim \begin{cases} S_{\text{rad}} & \text{for } S_{\text{bh}} > S_{\text{rad}} \\ S_{\text{bh}} & \text{for } S_{\text{rad}} > S_{\text{bh}}, \end{cases} \quad (7.13)$$

which reproduces the Page curve. This is because we have calculated the ensemble average of the von Neumann entropy (which obeys the Page curve for all members of the ensemble), and not the von Neumann entropy of the averaged density matrix as in Eq. (7.6).

Note that this observation is similar to that in Ref. [65], but here we do not consider any ensemble of holographic theories. Instead, the relevant ensemble arises from the coarse graining of black hole microstates, which is *forced* on us if we adopt any formalism involving the semiclassical picture of the black hole, such as gravitational path integral. See also Refs. [70,71] for related discussion.

## 2. Entanglement island and the Lorentzian picture

It was shown in Ref. [37], building on the technique developed in Refs. [89,92,186], that the prescription using the replica method in gravitational path integral is equivalent to the quantum extremal surface prescription [91] in the original unreplicated spacetime.<sup>63</sup> Going to the Lorentzian signature, this therefore leads to the following picture [33–35].

<sup>63</sup>Strictly speaking, this was shown only in the limit that the contribution from the graviton to the semiclassical field integration is negligible compared with those from other quantum fields. Below we assume that the equivalence of the two prescriptions persists beyond this limit.



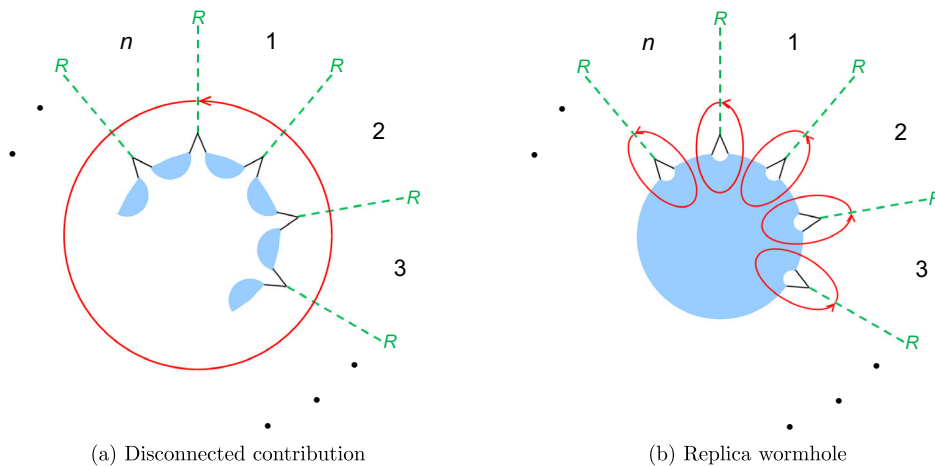


FIG. 23. (a) The contribution to  $Z_n$  with the gravitational region filled separately (light blue) on each sheet. All semiclassical matter configurations above and below the cuts  $R$  (green dashed) are related by Euclidean evolution (red arrow), giving the factor of  $e^{S_{\text{rad}}}$ . (b) For the replica symmetric wormhole configuration, the gravitational regions on  $n$  sheets are geometrically connected. This allows for  $n$  independent cycles for evolution (red arrows), giving the contribution from matter integration of order  $e^{nS_{\text{rad}}}$ .

In order to calculate the von Neumann entropy of radiation  $S_{\text{rad}}$  in some region  $R$ , we need to find an entanglement island  $I$  which extremizes the following quantity<sup>64</sup>

$$S_{\text{gen}}(R, I) = \frac{\mathcal{A}(\partial I)}{4G_N} - \text{Tr}[\rho_{\text{sc}}(R \cup I) \ln \rho_{\text{sc}}(R \cup I)], \quad (7.14)$$

where  $\mathcal{A}(\partial I)$  is the area of the boundary of  $I$ , and  $\rho_{\text{sc}}(X)$  is the reduced density matrix of the region  $X$  in the semiclassical theory. Note that  $I$  can be a null region. In general, there can be multiple such  $I$ 's, and the entropy  $S_{\text{rad}}$  is given by the minimum of  $S_{\text{gen}}(R, I)$ 's associated with all such  $I$ 's:

$$S_{\text{rad}} = \min_I \text{ext} S_{\text{gen}}(R, I). \quad (7.15)$$

This entropy is the same as that calculated by the replica method in path integral

$$S_{\text{rad}} = \overline{S_R}, \quad (7.16)$$

so that it obeys the Page curve.

It is important that this extremization procedure is performed on a global spacetime of general relativity. In particular, for a black hole spacetime, it must be performed on the whole spacetime including the interior of the black hole. This, therefore, gives a complementary picture to that described in Secs. II–VI. In the picture described here, the existence of the black hole interior is obvious—in fact, the framework assumes it—while to understand the unitary nature of black hole evolution, one needs to resort to a

method that appropriately incorporates nonperturbative effects of quantum gravity, such as replica wormholes. On the other hand, in the framework of Secs. II–VI, the unitarity of the evolution is an assumption and, as a consequence, the existence of the interior is not manifest—the interior emerges only effectively as a collective phenomenon involving horizon degrees of freedom, which are subject to universally chaotic and fast-scrambling dynamics.

### 3. Manifestly unitary vs global spacetime descriptions

We can, in fact, understand the existence of the two frameworks discussed above—based on the manifestly unitary and global spacetime pictures, respectively—from the viewpoint of gauge symmetries of the underlying theory [24,38]. As emphasized in Ref. [187], a theory of quantum gravity has nonperturbative gauge redundancies much larger than the standard diffeomorphism. A particular manifestation of this is the apparent violation of the Bekenstein-Hawking entropy bound in the semiclassical description of a black hole [24,38,188]. This violation arises from huge spatial volume inside the black hole [189,190] (including the so-called bags-of-gold configurations), which leads to the number of independent quantum states exceeding the Bekenstein-Hawking bound. Many of these semiclassically independent states, however, are equivalent under the nonperturbative gauge symmetries, making the number of truly independent states satisfy the Bekenstein-Hawking bound.

The two frameworks describe this phenomenon in very different, though equivalent, ways. In the framework discussed in Secs. II–VI, the nonperturbative gauge redundancies (as well as a part of the standard diffeomorphism) are explicitly fixed by employing a Schwarzschild-like time foliation, which is motivated by holography. In this framework, making unitarity manifest, the number of

<sup>64</sup>The regions  $R$  and  $I$  can be viewed either as spatial regions or causal regions associated with them.

independent black hole states is simply  $e^{S_{\text{BH}}}$ , and the apparently much larger number of independent interior states in semiclassical theory comes from the fact that the Hilbert space spanned by the  $e^{S_{\text{BH}}}$  states admit  $e^{e^{S_{\text{BH}}}}$  approximately orthogonal states, between which inner products are of order  $e^{-S_{\text{BH}}/2}$  or smaller. Since the semiclassical theory cannot detect such small inner products, it appears to accommodate independent quantum states larger than the Bekenstein-Hawking bound in the interior of a black hole.

On the other hand, in the framework making the interior manifest, the argument runs in the other direction. In this case, the starting point is the global spacetime picture, which is highly redundant. In the canonical formulation of quantum mechanics, this implies the existence of overentropic semiclassical interior states  $|\psi_i\rangle$ . After including nonperturbative effects of gravity, however, these semiclassically orthogonal states develop small overlaps,  $\langle\psi_i|\psi_j\rangle \sim e^{-S_{\text{BH}}/2}$ , in a such way that the rank of the matrix  $M_{ij} \equiv \langle\psi_i|\psi_j\rangle$  is reduced drastically to  $e^{S_{\text{BH}}}$ . Such a large number of null states are a manifestation of the large nonperturbative gauge redundancies [187,191], which relate even spaces with different topologies [187,192]. In the path integral formulation, this reduction of the semiclassical Hilbert space to the physical one is achieved by including additional contributions to the path integral (e.g., replica wormholes) which projects states onto those invariant under the relevant gauge symmetries as we have seen in this section.

## VIII. CONCLUSION

In quantum gravity, there has been a difficulty in reconciling fundamental principles of physics in the presence of a black hole (or other) horizon, particularly the unitarity of quantum mechanics and the equivalence principle of general relativity [1]. This difficulty is, in fact, an avatar of the conventional difficulty of describing UV physics at the Planck scale, albeit in a redshifted form. Because of a huge gravitational redshift between the horizon and asymptotic regions, the degrees of freedom represented by the Bekenstein-Hawking entropy [4,5], which obey an intrinsically “stringy” dynamics, appear to have exponentially degenerate states that cannot be discriminated operationally by performing a measurement on the black hole. A low energy description of quantum gravity treats these states in a thermal way, so that unitarity appears to be lost.

While a semiclassical theory, as a low energy description of quantum gravity, cannot describe all the microscopic dynamics of the fundamental theory, it can still be used to obtain a coarse-grained understanding of how unitarity and the equivalence principle can coexist in a black hole system, with only a few inputs from the UV theory. This is what we have explored in this paper.

One way of doing this is simply to postulate that the evolution of a black hole is unitary when viewed from the exterior [17–20]. This is a view motivated by holography [6]. In this case, the existence of the black hole interior appears to be in jeopardy [3]. We have seen, however, that with the assumption that the dynamics of horizon degrees of freedom is chaotic and fast scrambling across all low energy species, the interior emerges at a semiclassical level as a collective phenomenon involving the horizon, and possibly other, degrees of freedom [21–25]. An important point is that for a black hole with minimal uncertainties (i.e., a specific black hole at the semiclassical level), the assumption of Gaussian randomness of the coefficients for microstates allows us to learn many features associated with the construction of the effective theory of the interior. We do not need any further details about the microscopic theory to figure these things out. A similar construction applies to de Sitter spacetime, which we have also elaborated in this paper.

The compatibility between unitarity and the equivalence principle can alternatively be seen by starting from the global spacetime of general relativity. In this case, the existence of the interior is manifest, but at the apparent expense of unitarity as Hawking’s original calculation indicates [1]. However, while a semiclassical picture necessarily involves an average over microstates, we can directly calculate the average of the quantity we are interested in, e.g., the von Neumann entropy of emitted Hawking radiation, rather than the quantity in the averaged state. This is what is done in Refs. [33–37] by employing the replica method in gravitational path integral, or equivalently the quantum extremal surface prescription. In this way, one can see the unitarity of the underlying theory without knowing its detailed dynamics.

Despite the fact that the two frameworks described above appear very different, they give the same physical conclusions. In particular, a black hole evolves unitarily and has a smooth horizon. Viewed in this way, one can conclude that historical confusions about black hole physics come from the fact that only one of these features can be made manifest in a given low energy description; the other one appears in a highly nontrivial manner. We have seen that the description making unitarity (quantum mechanics) manifest comes more naturally with the canonical/Hamiltonian formulation of quantum mechanics, while the one making the interior (general relativity) manifest is associated with the path integral/Lagrangian formulation. It remains to be seen if there is a microscopic formulation of quantum gravity in which both these features are manifest at the same time.

## ACKNOWLEDGMENTS

We would like to thank Chris Akers, Sinya Aoki, Vijay Balasubramanian, Arjun Kar, Lampros Lamprou, Adam Levine, Juan Maldacena, Masamichi Miyaji, Dominik

Neuenfeld, Edgar Shaghoulian, Tomonori Ugajin, Shreya Vardhan, Zixia Wei, and Zhenbin Yang for useful discussions on this and related subjects. This work was supported in part by the Department of Energy, Office

of Science, Office of High Energy Physics under QuantISED Award No. DE-SC0019380 and Contract No. DE-AC02-05CH11231 and in part by MEXT KAKENHI Grants No. JP20H05850, No. JP20H05860.

- 
- [1] S. W. Hawking, Breakdown of predictability in gravitational collapse, *Phys. Rev. D* **14**, 2460 (1976).
- [2] S. D. Mathur, The information paradox: A pedagogical introduction, *Classical Quantum Gravity* **26**, 224001 (2009).
- [3] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, Black holes: Complementarity or firewalls?, *J. High Energy Phys.* **02** (2013) 062.
- [4] J. D. Bekenstein, Black holes and entropy, *Phys. Rev. D* **7**, 2333 (1973).
- [5] S. W. Hawking, Particle creation by black holes, *Commun. Math. Phys.* **43**, 199 (1975); **46**, 206(E) (1976).
- [6] J. M. Maldacena, The large N limit of superconformal field theories and supergravity, *Adv. Theor. Math. Phys.* **2**, 231 (1998).
- [7] G. 't Hooft, Dimensional reduction in quantum gravity, *Conf. Proc. C* **930308**, 284 (1993).
- [8] L. Susskind, The world as a hologram, *J. Math. Phys. (N.Y.)* **36**, 6377 (1995).
- [9] T. Banks, W. Fischler, S. H. Shenker, and L. Susskind, M theory as a matrix model: A conjecture, *Phys. Rev. D* **55**, 5112 (1997).
- [10] R. Bousso, The holographic principle, *Rev. Mod. Phys.* **74**, 825 (2002).
- [11] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, Local bulk operators in AdS/CFT: A Boundary view of horizons and locality, *Phys. Rev. D* **73**, 086003 (2006).
- [12] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, Holographic representation of local bulk operators, *Phys. Rev. D* **74**, 066009 (2006).
- [13] Y. Nomura, P. Rath, and N. Salzetta, Pulling the boundary into the bulk, *Phys. Rev. D* **98**, 026010 (2018).
- [14] C. Murdia, Y. Nomura, and P. Rath, Coarse-graining holographic states: A semiclassical flow in general spacetimes, *Phys. Rev. D* **102**, 086001 (2020).
- [15] A. R. Brown, H. Gharibyan, G. Penington, and L. Susskind, The python's lunch: Geometric obstructions to decoding Hawking radiation, *J. High Energy Phys.* **08** (2020) 121.
- [16] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, A world without pythons would be so simple, *Classical Quantum Gravity* **38**, 234001 (2021).
- [17] G. 't Hooft, On the quantum structure of a black hole, *Nucl. Phys.* **B256**, 727 (1985).
- [18] G. 't Hooft, The black hole interpretation of string theory, *Nucl. Phys.* **B335**, 138 (1990).
- [19] L. Susskind, L. Thorlacius, and J. Uglum, The stretched horizon and black hole complementarity, *Phys. Rev. D* **48**, 3743 (1993).
- [20] D. N. Page, Information in Black Hole Radiation, *Phys. Rev. Lett.* **71**, 3743 (1993).
- [21] Y. Nomura, Reanalyzing an evaporating black hole, *Phys. Rev. D* **99**, 086004 (2019).
- [22] Y. Nomura, Spacetime and universal soft modes—black holes and beyond, *Phys. Rev. D* **101**, 066024 (2020).
- [23] Y. Nomura, The interior of a unitarily evaporating black hole, *Phys. Rev. D* **102**, 026001 (2020).
- [24] K. Langhoff and Y. Nomura, Ensemble from coarse graining: Reconstructing the interior of an evaporating black hole, *Phys. Rev. D* **102**, 086021 (2020).
- [25] Y. Nomura, Black hole interior in unitary gauge construction, *Phys. Rev. D* **103**, 066011 (2021).
- [26] K. Papadodimas and S. Raju, An infalling observer in AdS/CFT, *J. High Energy Phys.* **10** (2013) 212.
- [27] E. Verlinde and H. Verlinde, Black hole entanglement and quantum error correction, *J. High Energy Phys.* **10** (2013) 107.
- [28] Y. Nomura and J. Varela, A note on (no) firewalls: The entropy argument, *J. High Energy Phys.* **07** (2013) 124.
- [29] J. Maldacena and L. Susskind, Cool horizons for entangled black holes, *Fortschr. Phys.* **61**, 781 (2013).
- [30] P. Hayden and G. Penington, Learning the alpha-bits of black holes, *J. High Energy Phys.* **12** (2019) 007.
- [31] K. Papadodimas and S. Raju, State-dependent bulk-boundary maps and black hole complementarity, *Phys. Rev. D* **89**, 086010 (2014).
- [32] K. Papadodimas and S. Raju, Remarks on the necessity and implications of state-dependence in the black hole interior, *Phys. Rev. D* **93**, 084049 (2016).
- [33] G. Penington, Entanglement wedge reconstruction and the information paradox, *J. High Energy Phys.* **09** (2020) 002.
- [34] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole, *J. High Energy Phys.* **12** (2019) 063.
- [35] A. Almheiri, R. Mahajan, J. Maldacena, and Y. Zhao, The Page curve of Hawking radiation from semiclassical geometry, *J. High Energy Phys.* **03** (2020) 149.
- [36] G. Penington, S. H. Shenker, D. Stanford, and Z. Yang, Replica wormholes and the black hole interior, *J. High Energy Phys.* **03** (2022) 205.
- [37] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, Replica wormholes and the entropy of Hawking radiation, *J. High Energy Phys.* **05** (2020) 013.
- [38] Y. Nomura, From the black hole conundrum to the structure of quantum gravity, *Mod. Phys. Lett. A* **36**, 2130007 (2021).



- [39] G. W. Gibbons and S. W. Hawking, Cosmological event horizons, thermodynamics, and particle creation, *Phys. Rev. D* **15**, 2738 (1977).
- [40] Y. Nomura, P. Rath, and N. Salzetta, Spacetime from unentanglement, *Phys. Rev. D* **97**, 106010 (2018).
- [41] F. Sanches and S. J. Weinberg, Holographic entanglement entropy conjecture for general spacetimes, *Phys. Rev. D* **94**, 084034 (2016).
- [42] Y. Nomura, N. Salzetta, F. Sanches, and S. J. Weinberg, Toward a holographic theory for general spacetimes, *Phys. Rev. D* **95**, 086002 (2017).
- [43] M. Alishahiha, A. Karch, E. Silverstein, and D. Tong, The dS/dS correspondence, *AIP Conf. Proc.* **743**, 393 (2004).
- [44] M. Alishahiha, A. Karch, and E. Silverstein, Hologravity, *J. High Energy Phys.* **06** (2005) 028.
- [45] L. Susskind, Black holes hint towards de Sitter-matrix theory, [arXiv:2109.01322](https://arxiv.org/abs/2109.01322).
- [46] L. Susskind, Entanglement and chaos in de Sitter space holography: An SYK example, *J. Hologr. Appl. Phys.* **1**, 1 (2021).
- [47] E. Shaghoulian, The central dogma and cosmological horizons, *J. High Energy Phys.* **01** (2022) 132.
- [48] E. Shaghoulian and L. Susskind, Entanglement in de Sitter space, *J. High Energy Phys.* **08** (2022) 198.
- [49] J. M. Bardeen, B. Carter, and S. W. Hawking, The four laws of black hole mechanics, *Commun. Math. Phys.* **31**, 161 (1973).
- [50] J. D. Bekenstein, Generalized second law of thermodynamics in black hole physics, *Phys. Rev. D* **9**, 3292 (1974).
- [51] J. Maldacena, S. H. Shenker, and D. Stanford, A bound on chaos, *J. High Energy Phys.* **08** (2016) 106.
- [52] P. Hayden and J. Preskill, Black holes as mirrors: Quantum information in random subsystems, *J. High Energy Phys.* **09** (2007) 120.
- [53] Y. Sekino and L. Susskind, Fast scramblers, *J. High Energy Phys.* **10** (2008) 065.
- [54] W. G. Unruh, Notes on black hole evaporation, *Phys. Rev. D* **14**, 870 (1976).
- [55] W. Israel, Thermo field dynamics of black holes, *Phys. Lett.* **57A**, 107 (1976).
- [56] T. Banks, Cosmological breaking of supersymmetry?, *Int. J. Mod. Phys. A* **16**, 910 (2001).
- [57] E. Witten, Quantum gravity in de Sitter space, in *Strings 2001*, edited by A. Dabholkar, S. Mukhi, and S. R. Wadia (Clay Mathematics Proceedings, 2002), 1, p. 351.
- [58] L. Susskind, Addendum to fast scramblers, [arXiv:1101.6048](https://arxiv.org/abs/1101.6048).
- [59] N. Goheer, M. Kleban, and L. Susskind, The trouble with de Sitter space, *J. High Energy Phys.* **07** (2003) 056.
- [60] J. Pollack, M. Rozali, J. Sully, and D. Wakeham, Eigenstate Thermalization and Disorder Averaging in Gravity, *Phys. Rev. Lett.* **125**, 021601 (2020).
- [61] A. Belin and J. de Boer, Random statistics of OPE coefficients and Euclidean wormholes, *Classical Quantum Gravity* **38**, 164001 (2021).
- [62] H. Liu and S. Vardhan, Entanglement entropies of equilibrated pure states in quantum many-body systems and gravity, *PRX Quantum* **2**, 010344 (2021).
- [63] B. Freivogel, D. Nikolakopoulou, and A. F. Rotundo, Wormholes from averaging over states, [arXiv:2105.12771](https://arxiv.org/abs/2105.12771).
- [64] J. Cotler and K. Jensen, A precision test of averaging in AdS/CFT, *J. High Energy Phys.* **11** (2022) 070.
- [65] R. Bousso and E. Wildenhain, Gravity/ensemble duality, *Phys. Rev. D* **102**, 066005 (2020).
- [66] N. Engelhardt, S. Fischetti, and A. Maloney, Free energy from replica wormholes, *Phys. Rev. D* **103**, 046021 (2021).
- [67] R. Renner and J. Wang, The black hole information puzzle and the quantum de Finetti theorem, [arXiv:2110.14653](https://arxiv.org/abs/2110.14653).
- [68] X.-L. Qi, Z. Shangnan, and Z. Yang, Holevo information and ensemble theory of gravity, *J. High Energy Phys.* **02** (2022) 056.
- [69] A. Almheiri and H. W. Lin, The entanglement wedge of unknown couplings, *J. High Energy Phys.* **08** (2022) 062.
- [70] A. Blommaert, L. V. Iliesiu, and J. Kruthoff, Gravity factorized, *J. High Energy Phys.* **09** (2022) 080.
- [71] J. Chandra and T. Hartman, Coarse graining pure states in AdS/CFT, [arXiv:2206.03414](https://arxiv.org/abs/2206.03414).
- [72] H. Maxfield and G. J. Turiaci, The path integral of 3D gravity near extremality; or, JT gravity with defects as a matrix integral, *J. High Energy Phys.* **01** (2021) 118.
- [73] M. Heydemann, L. V. Iliesiu, G. J. Turiaci, and W. Zhao, The statistical mechanics of near-BPS black holes, *J. Phys. A* **55**, 014004 (2022).
- [74] T. Regge and J. A. Wheeler, Stability of a Schwarzschild singularity, *Phys. Rev.* **108**, 1063 (1957).
- [75] R. H. Price, Nonspherical perturbations of relativistic gravitational collapse. I. Scalar and gravitational perturbations, *Phys. Rev. D* **5**, 2419 (1972).
- [76] S. A. Teukolsky, Perturbations of a rotating black hole. I. Fundamental equations for gravitational electromagnetic and neutrino field perturbations, *Astrophys. J.* **185**, 635 (1973).
- [77] Y. Hatsuda, An alternative to the Teukolsky equation, *Gen. Relativ. Gravit.* **53**, 93 (2021).
- [78] J. D. Bekenstein, Nonexistence of baryon number for static black holes, *Phys. Rev. D* **5**, 1239 (1972).
- [79] Y. B. Zeldovich, A novel type of radioactive decay: Gravitational baryon annihilation, *Zh. Eksp. Teor. Fiz.* **72**, 18 (1977).
- [80] T. Banks and N. Seiberg, Symmetries and strings in field theory and gravity, *Phys. Rev. D* **83**, 084019 (2011).
- [81] D. Harlow and H. Ooguri, Constraints on Symmetries from Holography, *Phys. Rev. Lett.* **122**, 191601 (2019).
- [82] G. Dvali, Black holes and large N species solution to the hierarchy problem, *Fortschr. Phys.* **58**, 528 (2010).
- [83] D. N. Page, Is Black Hole Evaporation Predictable?, *Phys. Rev. Lett.* **44**, 301 (1980).
- [84] Y. Nomura, J. Varela, and S. J. Weinberg, Black holes, information, and Hilbert space for quantum gravity, *Phys. Rev. D* **87**, 084050 (2013).
- [85] S. Aoki, T. Onogi, and S. Yokoyama, Conserved charges in general relativity, *Int. J. Mod. Phys. A* **36**, 2150098 (2021).
- [86] Y. Nomura and S. J. Weinberg, Entropy of a vacuum: What does the covariant entropy count?, *Phys. Rev. D* **90**, 104003 (2014).
- [87] S. Ryu and T. Takayanagi, Holographic Derivation of Entanglement Entropy from AdS/CFT, *Phys. Rev. Lett.* **96**, 181602 (2006).



- [88] V. E. Hubeny, M. Rangamani, and T. Takayanagi, A covariant holographic entanglement entropy proposal, *J. High Energy Phys.* **07** (2007) 062.
- [89] A. Lewkowycz and J. Maldacena, Generalized gravitational entropy, *J. High Energy Phys.* **08** (2013) 090.
- [90] T. Faulkner, A. Lewkowycz, and J. Maldacena, Quantum corrections to holographic entanglement entropy, *J. High Energy Phys.* **11** (2013) 074.
- [91] N. Engelhardt and A. C. Wall, Quantum extremal surfaces: Holographic entanglement entropy beyond the classical regime, *J. High Energy Phys.* **01** (2015) 073.
- [92] X. Dong, A. Lewkowycz, and M. Rangamani, Deriving covariant holographic entanglement, *J. High Energy Phys.* **11** (2016) 028.
- [93] A. Almheiri, X. Dong, and B. Swingle, Linearity of holographic entanglement entropy, *J. High Energy Phys.* **02** (2017) 074.
- [94] W. H. Zurek, Entropy Evaporated by a Black Hole, *Phys. Rev. Lett.* **49**, 1683 (1982).
- [95] D. N. Page, Comment on ‘Entropy Evaporated by a Black Hole’, *Phys. Rev. Lett.* **50**, 1013 (1983).
- [96] J. B. Hartle and S. W. Hawking, Path integral derivation of black hole radiance, *Phys. Rev. D* **13**, 2188 (1976).
- [97] Y. Nomura, F. Sanches, and S. J. Weinberg, Black Hole Interior in Quantum Gravity, *Phys. Rev. Lett.* **114**, 201301 (2015).
- [98] Y. Nomura, F. Sanches, and S. J. Weinberg, Relativeness in quantum gravity: Limitations and frame dependence of semiclassical descriptions, *J. High Energy Phys.* **04** (2015) 158.
- [99] W. G. Unruh, Origin of the particles in black hole evaporation, *Phys. Rev. D* **15**, 365 (1977).
- [100] W. Israel, Shenanigans at the black hole horizon: Pair creation or Boulware accretion?, [arXiv:1504.02419](https://arxiv.org/abs/1504.02419).
- [101] S. B. Giddings, Hawking radiation, the Stefan–Boltzmann law, and unitarization, *Phys. Lett. B* **754**, 39 (2016).
- [102] W. G. Unruh and R. M. Wald, Acceleration radiation and generalized second law of thermodynamics, *Phys. Rev. D* **25**, 942 (1982).
- [103] A. R. Brown, Tensile Strength and the Mining of Black Holes, *Phys. Rev. Lett.* **111**, 211301 (2013).
- [104] D. N. Page, Particle emission rates from a black hole: Massless particles from an uncharged, nonrotating hole, *Phys. Rev. D* **13**, 198 (1976).
- [105] S. Ghosh and S. Raju, Quantum information measures for restricted sets of observables, *Phys. Rev. D* **98**, 046005 (2018).
- [106] A. Almheiri, X. Dong, and D. Harlow, Bulk locality and quantum error correction in AdS/CFT, *J. High Energy Phys.* **04** (2015) 163.
- [107] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, Holographic quantum error-correcting codes: toy models for the bulk/boundary correspondence, *J. High Energy Phys.* **06** (2015) 149.
- [108] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter, and Z. Yang, Holographic duality from random tensor networks, *J. High Energy Phys.* **11** (2016) 009.
- [109] S. R. Coleman and F. De Luccia, Gravitational effects on and of vacuum decay, *Phys. Rev. D* **21**, 3305 (1980).
- [110] N. Bao, G. Penington, J. Sorce, and A. C. Wall, Beyond toy models: Distilling tensor networks in full AdS/CFT, *J. High Energy Phys.* **11** (2019) 069.
- [111] B. Swingle, Entanglement renormalization and holography, *Phys. Rev. D* **86**, 065007 (2012).
- [112] M. Miyaji and T. Takayanagi, Surface/state correspondence as a generalized holography, *Prog. Theor. Exp. Phys.* **2015**, 073B03 (2015).
- [113] N. Engelhardt and A. C. Wall, Extremal surface barriers, *J. High Energy Phys.* **03** (2014) 068.
- [114] I. Heemskerk, D. Marolf, J. Polchinski, and J. Sully, Bulk and transhorizon measurements in AdS/CFT, *J. High Energy Phys.* **10** (2012) 165.
- [115] R. Bousso, Z. Fisher, S. Leichenauer, and A. C. Wall, Quantum focusing conjecture, *Phys. Rev. D* **93**, 064044 (2016).
- [116] R. Bousso, A covariant entropy conjecture, *J. High Energy Phys.* **07** (1999) 004.
- [117] R. Bousso, Holography in general space-times, *J. High Energy Phys.* **06** (1999) 028.
- [118] R. Bousso and N. Engelhardt, Generalized second law for cosmology, *Phys. Rev. D* **93**, 024025 (2016).
- [119] C. Fefferman and C. R. Graham, Conformal invariants, in *The Mathematical Heritage of Élie Cartan (Lyon, 1984)* (1985), Astérisque, no. S131, 95.
- [120] E. T. Akhmedov, A remark on the AdS/CFT correspondence and the renormalization group flow, *Phys. Lett. B* **442**, 152 (1998).
- [121] E. Álvarez and C. Gómez, Geometric holography, the renormalization group and the c-theorem, *Nucl. Phys.* **B541**, 441 (1999).
- [122] V. Balasubramanian and P. Kraus, Space-Time and the Holographic Renormalization Group, *Phys. Rev. Lett.* **83**, 3605 (1999).
- [123] K. Skenderis and P. K. Townsend, Gravitational stability and renormalization group flow, *Phys. Lett. B* **468**, 46 (1999).
- [124] J. de Boer, E. Verlinde, and H. Verlinde, On the holographic renormalization group, *J. High Energy Phys.* **08** (2000) 003.
- [125] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Gravitational waves in general relativity VII. Waves from axisymmetric isolated systems, *Proc. R. Soc. A* **269**, 21 (1962).
- [126] R. K. Sachs, Gravitational waves in general relativity VIII. Waves in asymptotically flat space-times, *Proc. R. Soc. A* **270**, 103 (1962).
- [127] J. de Boer and S. N. Solodukhin, A holographic reduction of Minkowski space-time, *Nucl. Phys.* **B665**, 545 (2003).
- [128] A. Strominger, On BMS invariance of gravitational scattering, *J. High Energy Phys.* **07** (2014) 152.
- [129] C. Cheung, A. de la Fuente, and R. Sundrum, 4D scattering amplitudes and asymptotic symmetries from 2D CFT, *J. High Energy Phys.* **01** (2017) 112.
- [130] S. Pasterski, S.-H. Shao, and A. Strominger, Flat space amplitudes and conformal symmetry of the celestial sphere, *Phys. Rev. D* **96**, 065026 (2017).
- [131] A. Laddha, S. G. Prabhu, S. Raju, and P. Shrivastava, The holographic nature of null infinity, *SciPost Phys.* **10**, 041 (2021).

- [132] L. Dyson, M. Kleban, and L. Susskind, Disturbing implications of a cosmological constant, *J. High Energy Phys.* **10** (2002) 011.
- [133] S. Kachru, R. Kallosh, A. D. Linde, and S. P. Trivedi, de Sitter vacua in string theory, *Phys. Rev. D* **68**, 046005 (2003).
- [134] L. F. Abbott and S. Deser, Stability of gravity with a cosmological constant, *Nucl. Phys.* **B195**, 76 (1982).
- [135] Y. Nomura and S. J. Weinberg, Black holes, entropies, and semiclassical spacetime in quantum gravity, *J. High Energy Phys.* **10** (2014) 185.
- [136] P. V. Buividovich and M. I. Polikarpov, Entanglement entropy in gauge theories and the holographic principle for electric strings, *Phys. Lett. B* **670**, 141 (2008).
- [137] W. Donnelly, Decomposition of entanglement entropy in lattice gauge theory, *Phys. Rev. D* **85**, 085004 (2012).
- [138] H. Casini, M. Huerta, and J. A. Rosabal, Remarks on entanglement entropy for gauge fields, *Phys. Rev. D* **89**, 085012 (2014).
- [139] S. Ghosh, R. M. Soni, and S. P. Trivedi, On the entanglement entropy for gauge theories, *J. High Energy Phys.* **09** (2015) 069.
- [140] W. Donnelly and L. Freidel, Local subsystems in gauge theory and gravity, *J. High Energy Phys.* **09** (2016) 102.
- [141] T. Takayanagi and K. Tamaoka, Gravity edges modes and Hayward term, *J. High Energy Phys.* **02** (2020) 167.
- [142] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghi, Canonical Typicality, *Phys. Rev. Lett.* **96**, 050403 (2006).
- [143] S. Popescu, A. J. Short, and A. Winter, The foundations of statistical mechanics from entanglement: individual states vs averages, *Nat. Phys.* **2**, 754 (2006).
- [144] D. N. Page, Average Entropy of a Subsystem, *Phys. Rev. Lett.* **71**, 1291 (1993).
- [145] J. Maltz and L. Susskind, De Sitter Space as a Resonance, *Phys. Rev. Lett.* **118**, 101602 (2017).
- [146] R. Bousso, Violations of the Equivalence Principle by a Nonlocally Reconstructed Vacuum at the Black Hole Horizon, *Phys. Rev. Lett.* **112**, 041102 (2014).
- [147] D. L. Jafferis and L. Lamprou, Inside the hologram: Reconstructing the bulk observer's experience, *J. High Energy Phys.* **03** (2022) 084.
- [148] P. Gao and L. Lamprou, Seeing behind black hole horizons in SYK, *J. High Energy Phys.* **06** (2022) 143.
- [149] D. Petz, Sufficient subalgebras and the relative entropy of states of a von Neumann algebra, *Commun. Math. Phys.* **105**, 123 (1986).
- [150] D. Petz, Sufficiency of channels over von Neumann algebras, *Q. J. Math. Oxford Ser.* **39**, 97 (1988).
- [151] B. Czech, J. L. Karczmarek, F. Nogueira, and M. Van Raamsdonk, The gravity dual of a density matrix, *Classical Quantum Gravity* **29**, 155009 (2012).
- [152] A. C. Wall, Maximin surfaces, and the strong subadditivity of the covariant holographic entanglement entropy, *Classical Quantum Gravity* **31**, 225007 (2014).
- [153] M. Headrick, V. E. Hubeny, A. Lawrence, and M. Rangamani, Causality & holographic entanglement entropy, *J. High Energy Phys.* **12** (2014) 162.
- [154] D. L. Jafferis, A. Lewkowycz, J. Maldacena, and S. J. Suh, Relative entropy equals bulk relative entropy, *J. High Energy Phys.* **06** (2016) 004.
- [155] X. Dong, D. Harlow, and A. C. Wall, Reconstruction of Bulk Operators Within the Entanglement Wedge in Gauge-Gravity Duality, *Phys. Rev. Lett.* **117**, 021601 (2016).
- [156] J. Cotler, P. Hayden, G. Penington, G. Salton, B. Swingle, and M. Walter, Entanglement Wedge Reconstruction via Universal Recovery Channels, *Phys. Rev. X* **9**, 031011 (2019).
- [157] J. Polchinski, L. Susskind, and N. Toumbas, Negative energy, superluminosity and holography, *Phys. Rev. D* **60**, 084006 (1999).
- [158] S. B. Giddings and M. Lippert, Precursors, black holes, and a locality bound, *Phys. Rev. D* **65**, 024006 (2002).
- [159] I. Kim, E. Tang, and J. Preskill, The ghost in the radiation: Robust encodings of the black hole interior, *J. High Energy Phys.* **06** (2020) 031.
- [160] J. M. Maldacena, Eternal black holes in anti-de Sitter, *J. High Energy Phys.* **04** (2003) 021.
- [161] S. H. Shenker and D. Stanford, Black holes and the butterfly effect, *J. High Energy Phys.* **03** (2014) 067.
- [162] L. Susskind, Computational complexity and black hole horizons, *Fortschr. Phys.* **64**, 24 (2016).
- [163] K. Papadodimas and S. Raju, Local Operators in the Eternal Black Hole, *Phys. Rev. Lett.* **115**, 211601 (2015).
- [164] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao, Holographic Complexity Equals Bulk Action?, *Phys. Rev. Lett.* **116**, 191301 (2016).
- [165] A. Almheiri, R. Mahajan, and J. Maldacena, Islands outside the horizon, [arXiv:1910.11077](https://arxiv.org/abs/1910.11077).
- [166] A. Bouland, B. Fefferman, and U. Vazirani, Computational pseudorandomness, the wormhole growth paradox, and constraints on the AdS/CFT duality, [arXiv:1910.14646](https://arxiv.org/abs/1910.14646).
- [167] D. Neuenfeld, Homology conditions for RT surfaces in double holography, *Classical Quantum Gravity* **39**, 075009 (2022).
- [168] H. Verlinde, Deconstructing the wormhole: Factorization, entanglement and decoherence, [arXiv:2105.02142](https://arxiv.org/abs/2105.02142).
- [169] S. Leutheusser and H. Liu, Causal connectability between quantum systems and the black hole interior in holographic duality, [arXiv:2110.05497](https://arxiv.org/abs/2110.05497).
- [170] S. Leutheusser and H. Liu, Emergent times in holographic duality, [arXiv:2112.12156](https://arxiv.org/abs/2112.12156).
- [171] E. Witten, Gravity and the crossed product, *J. High Energy Phys.* **10** (2022) 008.
- [172] A. Strominger, The dS/CFT correspondence, *J. High Energy Phys.* **10** (2001) 034.
- [173] L. Aalsma and G. Shiu, Chaos and complementarity in de Sitter space, *J. High Energy Phys.* **05** (2020) 152.
- [174] H. Geng, Non-local entanglement and fast scrambling in de-Sitter holography, *Ann. Phys. (Amsterdam)* **426**, 168402 (2021).
- [175] H. Geng, Y. Nomura, and H.-Y. Sun, Information paradox and its resolution in de Sitter holography, *Phys. Rev. D* **103**, 126004 (2021).
- [176] L. Aalsma, A. Cole, E. Morvan, J. P. van der Schaar, and G. Shiu, Shocks and information exchange in de Sitter space, *J. High Energy Phys.* **10** (2021) 104.
- [177] J. Kames-King, E. M. H. Verheijden, and E. P. Verlinde, No Page curves for the de Sitter horizon, *J. High Energy Phys.* **03** (2022) 040.

- [178] V. Balasubramanian, P. Hayden, A. Maloney, D. Marolf, and S. F. Ross, Multiboundary wormholes and holographic entanglement, *Classical Quantum Gravity* **31**, 185015 (2014).
- [179] G. W. Gibbons and S. W. Hawking, Action integrals and partition functions in quantum gravity, *Phys. Rev. D* **15**, 2752 (1977).
- [180] D. Harlow and E. Shaghoulian, Global symmetry, Euclidean gravity, and the black hole information problem, *J. High Energy Phys.* 04 (2021) 175.
- [181] L. V. Iliesiu, M. Mezei, and G. Sárosi, The volume of the black hole interior at late times, *J. High Energy Phys.* 07 (2022) 073.
- [182] T. S. Bunch and P. C. W. Davies, Quantum field theory in de Sitter space: Renormalization by point splitting, *Proc. R. Soc. A* **360**, 117 (1978).
- [183] H. Geng, A. Karch, C. Perez-Pardavila, S. Raju, L. Randall, M. Riojas, and S. Shashi, Information transfer with a gravitating bath, *SciPost Phys.* **10**, 103 (2021).
- [184] C. Callan and F. Wilczek, On geometric entropy, *Phys. Lett. B* **333**, 55 (1994).
- [185] P. Calabrese and J. Cardy, Entanglement entropy and quantum field theory, *J. Stat. Mech.* (2004) P06002.
- [186] X. Dong and A. Lewkowycz, Entropy, extremality, Euclidean variations, and the equations of motion, *J. High Energy Phys.* 01 (2018) 081.
- [187] D. Marolf and H. Maxfield, Transcending the ensemble: Baby universes, spacetime wormholes, and the order and disorder of black hole information, *J. High Energy Phys.* 08 (2020) 044.
- [188] J. Chakravarty, Overcounting of interior excitations: A resolution to the bags of gold paradox in AdS, *J. High Energy Phys.* 02 (2021) 027.
- [189] M. Christodoulou and C. Rovelli, How big is a black hole?, *Phys. Rev. D* **91**, 064046 (2015).
- [190] M. Christodoulou and T. De Lorenzo, Volume inside old black holes, *Phys. Rev. D* **94**, 104002 (2016).
- [191] J. McNamara and C. Vafa, Baby universes, holography, and the swampland, [arXiv:2004.06738](https://arxiv.org/abs/2004.06738).
- [192] D. L. Jafferis, Bulk reconstruction and the Hartle-Hawking wavefunction, [arXiv:1703.01519](https://arxiv.org/abs/1703.01519).