Black holes, information, and Hilbert space for quantum gravity

Yasunori Nomura, Jaime Varela, and Sean J. Weinberg

Phys. Rev. D **87**, 084050 — Published 19 April 2013

DOI: 10.1103/PhysRevD.87.084050

MIT-CTP-4405
UCB-PTH-12/17

Black Holes, Information, and Hilbert Space for Quantum Gravity

Yasunori Nomura[a,b], Jaime Varela[a,b], and Sean J. Weinberg[b]

[a] *Center for Theoretical Physics, Laboratory for Nuclear Science, and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

[b] *Berkeley Center for Theoretical Physics, Department of Physics, and Theoretical Physics Group, Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA*

Abstract

A coarse-grained description for the formation and evaporation of a black hole is given within the framework of a unitary theory of quantum gravity preserving locality, without dropping the information that manifests as macroscopic properties of the state at late times. The resulting picture depends strongly on the reference frame one chooses to describe the process. In one description based on a reference frame in which the reference point stays outside the black hole horizon for sufficiently long time, a late black hole state becomes a superposition of black holes in different locations and with different spins, even if the back hole is formed from collapsing matter that had a well-defined classical configuration with no angular momentum. The information about the initial state is partly encoded in relative coefficients—especially phases—of the terms representing macroscopically different geometries. In another description in which the reference point enters into the black hole horizon at late times, an S-matrix description in the asymptotically Minkowski spacetime is not applicable, but it sill allows for an "S-matrix" description in the full quantum gravitational Hilbert space including singularity states. Relations between different descriptions are given by unitary transformations acting on the full Hilbert space, and they in general involve superpositions of "distant" and "infalling" descriptions. Despite the intrinsically quantum mechanical nature of the black hole state, measurements performed by a classical physical observer are consistent with those implied by general relativity. In particular, the recently-considered firewall phenomenon can occur only for an exponentially fine-tuned (and intrinsically quantum mechanical) initial state, analogous to an entropy decreasing process in a system with large degrees of freedom.

1 Introduction

Since its discovery [1], the process of black hole formation and evaporation has contributed tremendously to our understanding of quantum aspects of gravity. Building on earlier ideas, in particular the holographic principle [2, 3] and complementarity [4], one of the authors (Y.N.) has recently proposed an explicit framework for formulating quantum gravity in a way that is consistent with locality at length scales larger than the Planck (or string) length [5]. (Essentially the same idea had been used earlier to describe the eternally inflating multiverse in Ref. [6].) This allows us to describe a system with gravity in such a way that the picture based on conventional quantum mechanics, including the emergence of classical worlds due to amplification in position space [7, 5], persists without any major modification.

In the framework given in Ref. [5], quantum states allowing for spacetime interpretation represent only the limited spacetime region in and on the apparent horizon *as viewed from a fixed (freely falling) reference frame*. Complementarity, as well as the observer dependence of cosmic horizons, can then be understood—and thus precisely formulated—as a reference frame change represented by a unitary transformation acting on the full quantum gravitational Hilbert space, which takes the form

$$\mathcal{H}_{\rm QG} = \left(\bigoplus_{\mathcal{M}} \mathcal{H}_{\mathcal{M}} \right) \oplus \mathcal{H}_{\rm sing}. \tag{1}$$

Here, $\mathcal{H}_{\mathcal{M}}$ is the Hilbert subspace containing states on a fixed semi-classical geometry $\mathcal{M}$ (or more precisely, a set of semi-classical geometries $\mathcal{M} = \{\mathcal{M}_i\}$ having the same horizon $\partial\mathcal{M}$), while $\mathcal{H}_{\rm sing}$ is that containing "intrinsically quantum mechanical" states associated with spacetime singularities.[1] In its minimal implementation—which we assume throughout—the framework of Ref. [5] says that a state that is an element of one of the $\mathcal{H}_{\mathcal{M}}$'s represents a physical configuration on the past light cone of the origin, p, of the reference frame. We call the structure of the Hilbert space in Eq. (1) with this particular implementation the *covariant Hilbert space for quantum gravity*.

The purpose of this paper is to develop a complete picture of black hole formation and evaporation in this framework, based on Eq. (1). Our picture says that:

- The evolution of the full quantum state is unitary.
- The state, however, is in general a superposition of macroscopically different worlds. In particular, the final state of black hole evaporation is a superposition of macroscopically distinguishable terms, even if the initial state forming the black hole is a classical object having a well-defined macroscopic configuration. The information of the initial state is encoded partly in relative coefficients, especially in phases, among these macroscopically different terms.

[1] Here, the geometry means that of a codimension-one hypersurface which the quantum states represent, and *not* that of spacetime.

- No physical observer can recover the initial state forming the black hole by observing final Hawking radiation quanta. This is true even if the measurement is performed with arbitrarily high precision using an arbitrary (in general quantum) measuring device.
- Observations each physical observer makes are well described by the semi-classical picture in the regime it is supposed to be applicable, unless the observer (or measuring device) is in an exponentially rare quantum state in the corresponding Hilbert space.

We note that while some of the considerations here are indeed specific to the present framework, some are more general and apply to other theories of gravity as well, especially to the ones in which the formation and evaporation of a black hole is described as a unitary quantum mechanical process.

There are several key ingredients to understand the features described above, which we now highlight:

- Quantum mechanics has a "dichotomic" nature about locality: while the *dynamics*, encoded in the time evolution operator, is local, a *state* is generically non-local, as is clearly demonstrated in the Einstein-Podolsky-Rosen experiment. In particular, this allows for a state to be a superposition of terms representing macroscopically different spacetime geometries.
- The framework of Ref. [5] says that general covariance implies the quantum states must be defined in a fixed (local Lorentz) reference frame; moreover, to preserve locality of the dynamics (*not* of states), these states represent spacetime regions only in and on the stretched/apparent horizon as viewed from the reference frame.[2] This implies, in particular, that once a reference frame is fixed, the *location* of a black hole (with respect to the reference frame) is a physically meaningful quantity, even if there is no other object.
- The location of a black hole is highly uncertain after long time [8]. In particular, at a timescale of evaporation $\sim M(0)^3$, where $M(0)$ is the initial black hole mass, the uncertainty of the location is of order $M(0)^2$, which is much larger than the Schwarzschild radius of the initial black hole, $R_S = 2M(0)$. (This is also the timescale in which the black hole loses more than a half of its initial Bekenstein-Hawking entropy in the form of Hawking radiation [9].) This implies that a state of a sufficiently old black hole becomes a superposition of terms representing macroscopically distinguishable worlds [10].

As discussed in Refs. [6, 11], these ingredients, especially the first two, are also important to understand the eternally inflating multiverse (or quantum many universes) and to give well-defined probabilistic predictions in such a cosmology.

[2]Note that the apparent horizon, defined as a surface on which the expansion of the past-directed light rays emitted from p turns from positive to negative, is in general *not* the same as the event horizon of the black hole. For example, there is no apparent, or stretched, horizon associated with the black hole when the system is described in an infalling reference frame.

In this paper, we will first discuss the picture presented above in the case that the evolution of a black hole is described in a distant reference frame, i.e. a freely falling frame whose origin p stays outside the black hole horizon for all time. We will, however, also discuss in detail what happens if we describe the system using an infalling reference frame, i.e. a frame in which p enters into the black hole horizon at late times. Following Ref. [5], we treat this problem by performing a unitary transformation on the state representing the black hole evolution as viewed from a distant reference frame. We find that, because of the uncertainty of the black hole location, the resulting state is in general *a superposition of* infalling and distant descriptions of the black hole, and this effect is particularly significant when we try to describe the interior of an old black hole.

The issue of information in black hole evaporation has a long history of extensive research, with earlier proposals including information loss [12, 13], remnants [14, 15], and baby universes [16]; see, e.g., Refs. [17, 18] for reviews. Our picture here is that the evolution of a black hole is unitary [19], as in the cases of the complementarity [4] and fuzzball [20] pictures, with a macroscopically non-local nature of a black hole final state explicitly taken into account. This non-locality is especially important for the explicit realization of complementarity as a unitary reference frame change in the quantum gravitational Hilbert space [5]. We note that some authors have discussed non-locality at a macroscopic level [15, 21] or an intrinsically quantum nature of black hole states [22] but in ways different from the ones considered here. In particular, these proposals lead either to non-locality of the dynamics (not only states), large deviations from the semi-classical picture experienced by a macroscopic physical observer, or intrinsic quantum effects confined only to the microscopic domain, none of which applies to the picture presented here.

The organization of this paper is as follows. In Section 2, we provide a detailed description of the formation and evaporation of a black hole as viewed from a distant reference frame. We elucidate the meaning of information in this context, and find that it is partly in relative coefficients of terms representing different macroscopic configurations of Hawking quanta and/or geometries. We also discuss what a physical observer, who should be a part of the description, will measure and if he/she can reconstruct the initial state based on his/her measurements. In Section 3, we consider descriptions of the system in different reference frames. In particular, we discuss how the spacetime inside the black hole horizon appears in these descriptions and how these descriptions are related to the distant description given in Section 2. We also elaborate on the analysis of Ref. [10], arguing that the firewall paradox recently pointed out in Ref. [23] does not exist (although the firewall "phenomenon" can occur if the initial state is exponentially fine-tuned). In Section 4 we give our final discussion and conclusions. In the appendix, we provide an analysis of the "spontaneous spin-up" phenomenon, which we find to occur for a general Schwarzschild (or very slowly rotating) black hole. This effect makes a Schwarzschild black hole evolve into a superposition of Kerr black holes with distinct angular momenta.

Throughout the paper, we limit our discussions to the case of four spacetime dimensions to avoid unnecessary complications of various expressions; but the extension to other dimensions is straightforward. We also take the unit in which the Planck scale, $G_N^{-1/2} \simeq 1.22 \times 10^{19}$ GeV, is set to unity, so all the quantities appearing can be regarded as dimensionless.

2 Black Holes and Unitarity—A Distant View

In this section we discuss how unitarity of quantum mechanical evolution is preserved in the process in which a black hole forms and evaporates, as viewed from a distant reference frame. We clarify the meaning of the information in this context, and argue that it (partly) lies in relative coefficients—especially phases—of terms representing *macroscopically distinct* configurations in a full quantum state. We also elucidate the fact that a physical observer can never extract complete (quantum) information of the initial state forming the black hole; i.e., observing final-state Hawking radiation does not allow for him/her to infer the initial state, despite the fact that the evolution of the entire quantum state is fully unitary.

2.1 Information paradox—what is the information?

In his famous 1976 paper, Hawking argued, based on semi-classical considerations, that a black hole loses information [12]. Consider two objects having the same energy-momentum, represented by pure quantum states $|A\rangle$ and $|B\rangle$, which later collapse into black holes with the same mass $M(0)$. According to the semi-classical picture, the evolutions of the two states after forming the black holes are identical, leading to the same mixed state ρ_H, obtained by integrating the thermal Hawking radiation states:

$$\begin{array}{rcl} |A\rangle & \to & \rho_H, \\ |B\rangle & \to & \rho_H. \end{array} \tag{2}$$

This phenomenon is referred to as the information loss in black holes.

What is the problem of this picture? The problem is that since the final states are identical, we cannot recover the initial state of the evolution just by knowing the final state, even in principle. This contradicts unitarity of quantum mechanical evolution, which says that time evolution of a state is reversible, i.e. we can always recover the initial state if we know the final state exactly by applying the inverse time evolution operator e^{+iHt}.

Based on various circumstantial evidence, especially AdS/CFT duality [24], we now do not think the above picture is correct. We think that the final states obtained from different initial states differ, and a state obtained by evolving any pure state is always pure even if the evolution

involves formation and evaporation of a black hole. Namely, instead of Eq. (2), we have

$$\begin{array}{rcl} |A\rangle & \to & |\psi_A\rangle\,, \\ |B\rangle & \to & |\psi_B\rangle\,, \end{array} \tag{3}$$

where $|\psi_A\rangle \neq |\psi_B\rangle$ iff $|A\rangle \neq |B\rangle$. In this picture, quantum states representing black holes formed by different initial states are different, even if they have the same mass. (The dimension of the Hilbert space corresponding to a classical black hole of a fixed mass M is $\exp(\mathcal{A}_{\rm BH}/4)$ according to the Bekenstein-Hawking entropy, where $\mathcal{A}_{\rm BH} = 16\pi M^2$ is the area of the black hole horizon.) These states then evolve into different final states $|\psi_A\rangle$ and $|\psi_B\rangle$, representing states for emitted Hawking radiation quanta.

A question is in what form the information is encoded in the final state. On one hand, possible final states of evaporation of a black hole must have a sufficient variety to encode complete information about the initial state forming the black hole. This requires that the dimension of the Hilbert space corresponding to these states must be of order $\exp(\mathcal{A}_{\rm BH}(0)/4)$, where $\mathcal{A}_{\rm BH}(0) = 16\pi M(0)^2$ is the area of the black hole horizon right after the formation. On the other hand, Hawking radiation quanta emitted from the black hole must have the thermal spectrum (with temperature $T_H = 1/8\pi M$ when the black hole mass is M) in the regime where the semi-classical analysis is valid, $M \gg 1$. It is not clear how the state actually realizes these two features [25], although the generalized second law of thermodynamics guarantees that it can be done. Below, we argue that a part of the information that is necessary to recover the initial state is contained in relative coefficients of terms representing different macroscopic worlds, even if the initial state has a well-defined classical configuration.

Our analysis does not prove unitarity of the black hole formation/evaporation process, or address the question of how the complete information of the initial state is encoded in the emitted Hawking quanta at the microscopic level. Rather, we assume that unitarity is preserved at the microscopic level, and study manifestations of this assumption when we describe the process at a semi-classical level. This will provide implications on how such a description must be constructed. For example, in order to preserve all the information in the initial state, the description must not be given on a fixed black hole background in an intermediate stage of the evaporation, since it would correspond to ignoring a part of the information contained in the full quantum state manifested as macroscopic properties of the remaining black hole. Note that we do not claim that these macroscopic properties contain independent information beyond what is in the emitted Hawking quanta—the two are certainly correlated by energy-momentum conservation. The analysis presented here also has implications on the complementarity picture, which will be discussed in Section 3.

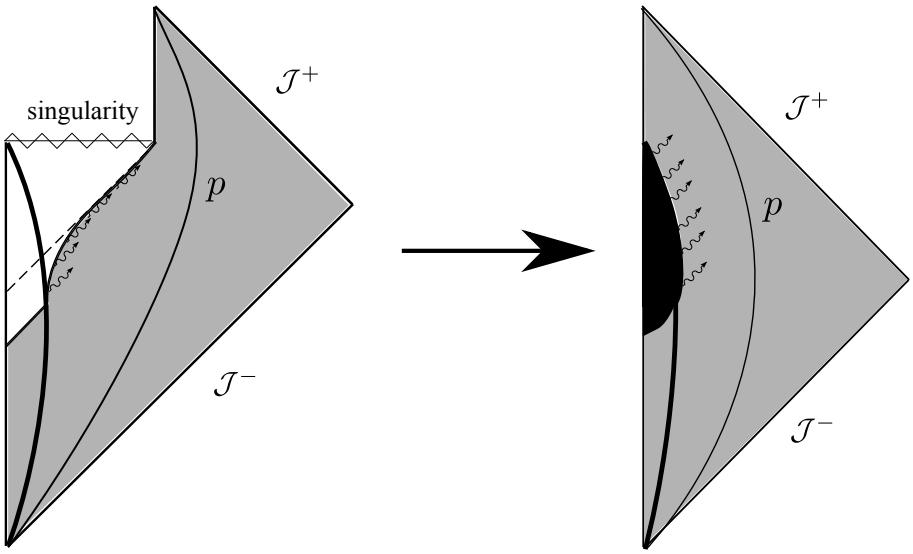


Figure 1: The Penrose diagram representing a black hole formed from a collapsing shell of matter (represented by the thick solid curve) which then evaporates. The left panel shows the standard "global spacetime" picture, in which Hawking radiation (denoted by wavy arrows) comes from the stretched horizon. To obtain a consistent quantum mechanical description, we must fix a reference frame (freely falling frame) and then describe the system from that viewpoint [5]. Quantum states then corresponds to physical configurations in the past light cone of the origin p of that reference frame. Here we choose a "distant" reference frame; the trajectory of its origin p is depicted by a thin solid curve. With this choice, a *complete* description of the evolution of the system is obtained in the shaded region in the panel. In other words, the conformal structure of the *entire* spacetime is as in the right panel, when the system is described in this reference frame.

2.2 Where is the information in the black hole state?

Let us consider a process in which a black hole is formed from a pure state $|A\rangle$ and then evaporates. For simplicity, we assume that the black hole formed does not have a spin or charge. We describe this process in a distant reference frame, i.e. a freely falling (local Lorentz) frame whose origin p is outside the black hole horizon all the time; see the left panel of Fig. 1. In its minimal implementation, the framework of Ref. [5] says that quantum states represent physical configurations on the past light cone of p in and on the stretched/apparent horizon.[3] This description, therefore, represents evolution of the system in the shaded spacetime region in the left panel of Fig. 1.

An important point is that this provides a *complete* description of the *entire* system [19, 4]—it is not that we describe only a part of the system corresponding to the shaded region; physics is complete in that spacetime region. The picture describing the infalling matter inside the horizon

[3]The stretched horizon is defined as a time-like hypersurface on which the local Hawking temperature becomes of order the Planck scale and thus short-distance quantum gravity effects become important (where we have not discriminated between the string and Planck scales). In the Schwarzschild coordinates, it is located at $r-2M \approx 1/M$.

can be obtained only after performing a unitary transformation on the state corresponding to a change of the reference frame to an infalling one [5] (which in general leads to a superposition of infalling and distant views, as will be explained in Section 3). In this sense, the *entire* spacetime is better represented by a Penrose diagram in the right panel of Fig. 1 when the system is described in a distant reference frame. As is clear from the figure, this allows for an S-matrix description of the process in Hilbert space representing Minkowski space $\mathcal{H}_{\rm Minkowski}$, which is a subspace of the whole covariant Hilbert space for quantum gravity: $\mathcal{H}_{\rm Minkowski} \subset \mathcal{H}_{\rm QG}$. This is the case despite the fact that in general quantum mechanics requires only that the evolution of a state is unitary in the whole Hilbert space $\mathcal{H}_{\rm QG}$; see Section 3 for more discussions on this point.

What does the evolution of a quantum state look like in this description? Let us denote the black hole state right after the collapse of the matter by $\left|{\rm BH}^0_A\right\rangle$. Since the subsequent evolution is unitary, the state can be written in the form $\sum_i a^t_i \left|{\rm BH}^t_i\right\rangle \otimes |\psi^t_i\rangle$. Here, $\left|{\rm BH}^t_i\right\rangle$ represent states of the black hole (i.e. the horizon degrees of freedom) when time t is passed since the formation, while $|\psi^t_i\rangle$ those of the rest of the world at the same time, where t is the proper time measured at the origin of the reference frame p. (The dimension of the Hilbert space for $\left|{\rm BH}^t_i\right\rangle$, $\mathcal{H}^t_{\rm BH}$, is $\exp(\mathcal{A}_{\rm BH}(t)/4)$ with $\mathcal{A}_{\rm BH}(t) = 16\pi M(t)^2$, where $M(t)$ is the mass of the black hole at time t; the state $\left|{\rm BH}^0_A\right\rangle$ is an element of $\mathcal{H}^0_{\rm BH}$.) The entire state then evolves into a state representing the final Hawking radiation quanta, which can be written as $\sum_i a^\infty_i |\psi^\infty_i\rangle$. Summarizing, the evolution of the system is described as

$$|A\rangle \longrightarrow \left|{\rm BH}^0_A\right\rangle \longrightarrow \sum_i a^t_i \left|{\rm BH}^t_i\right\rangle \otimes \left|\psi^t_i\right\rangle \longrightarrow \sum_i a^\infty_i \left|\psi^\infty_i\right\rangle . \tag{4}$$

The complete information about the initial state is contained in the state at any time t in the set of complex coefficients when the state is expanded in fixed basis states. In particular, after the evaporation it is contained in $\{a^\infty_i\}$ showing how the radiation states are superposed.

2.3 Black hole drifting: a macroscopic uncertainty of the black hole location after a long time

What actually are the states $|\psi^t_i\rangle$? Namely, what does the intermediate stage of the evaporation look like when it is described from a distant reference frame? Here we argue that $|\psi^t_i\rangle$ for different i span macroscopically different worlds. In particular, the state of the black hole becomes a superposition of macroscopically different geometries (in the sense that they represent different spacetimes as viewed from the reference frame) throughout the course of the evaporation. The analysis here builds upon an earlier suggestion by Page, who noted a large backreaction of Hawking emissions to the location of an evaporating black hole [8].

To analyze the issue, let us take a semi-classical picture of the evaporation but in which the backreaction of the Hawking emission to the black hole energy-momentum is explicitly taken into

account. Specifically, we model it by a process in which the black hole emits a massless quantum with energy $\sim M_{\rm Pl}^2/M$ in a random direction in each time interval $\sim (M/M_{\rm Pl})l_{\rm Pl}$, in the rest frame of the black hole. Here, M is the mass of the black hole at the time of the emission, and we have restored the Planck scale $G_N = 1/M_{\rm Pl}^2 = l_{\rm Pl}^2$, which we keep until our final result in Eq. (11). Suppose that the velocity of the black hole is $\mathbf{v}$ before an emission; then the emission of a Hawking quantum will change the four-momentum of the black hole as

$$p^\mu_{\rm BH} = \begin{pmatrix} M\gamma \\ M\gamma\mathbf{v} \end{pmatrix} \longrightarrow \begin{pmatrix} M\gamma - \frac{\gamma M_{\rm Pl}^2}{M}(1 - \mathbf{n}\cdot\mathbf{v}) \\ M\gamma\mathbf{v} + \frac{M_{\rm Pl}^2}{M}\mathbf{n} - \frac{(1-\gamma)M_{\rm Pl}^2}{M}\frac{\mathbf{n}\cdot\mathbf{v}}{|\mathbf{v}|^2}\mathbf{v} - \frac{\gamma M_{\rm Pl}^2}{M}\mathbf{v} \end{pmatrix}, \tag{5}$$

where $\gamma \equiv 1/\sqrt{1-|\mathbf{v}|^2}$ and $\mathbf{n}$ is a unit vector pointing to a random direction. The mass and the velocity of the black hole, therefore, change by

$$\Delta M \;=\; \sqrt{M^2 - 2M_{\rm Pl}^2} - M \;\approx\; -\frac{M_{\rm Pl}^2}{M}, \tag{6}$$

$$\Delta\mathbf{v} \;=\; \frac{M_{\rm Pl}^2}{\gamma\{M^2 - M_{\rm Pl}^2(1-\mathbf{n}\cdot\mathbf{v})\}}\left\{\mathbf{n} - \left(1 - \frac{1}{\gamma}\right)\frac{\mathbf{n}\cdot\mathbf{v}}{|\mathbf{v}|^2}\mathbf{v}\right\} \;\approx\; \frac{M_{\rm Pl}^2}{M^2}\mathbf{n} - \frac{M_{\rm Pl}^2}{2M^2}(\mathbf{n}\cdot\mathbf{v})\mathbf{v}, \tag{7}$$

in each time interval

$$\Delta t \;=\; \frac{M\gamma}{M_{\rm Pl}}l_{\rm Pl} \;\approx\; \frac{M}{M_{\rm Pl}}l_{\rm Pl}, \tag{8}$$

where we have taken the approximation that $M \gg M_{\rm Pl}$ and $|\mathbf{v}| \ll 1$ in the rightmost expressions. In general, the emission of a Hawking quantum can also change the black hole angular momentum $\mathbf{J}$. We consider this effect in the appendix, where we find that the black hole accumulates macroscopic angular momentum, $|\mathbf{J}| \gg 1$, after long time. This, however, does not affect the essential part of the discussion below, so we will suppress it in most part.

Now, suppose that a (non-spinning) black hole is formed at $t = 0$ with the initial mass $M_0 \equiv M(0)$. Then, in timescales of order $M_0^3/M_{\rm Pl}^4$ or shorter, the black hole mass is still of order M_0 until the very last moment of the evaporation. (For example, at the Page time $t_{\rm Page} \sim M_0^3/M_{\rm Pl}^4$, at which the black hole loses a half of its initial entropy, the black hole mass is still $M \approx M_0/\sqrt{2}$.) The above process, therefore, can be well approximated by a process in which the black hole receives a velocity kick of $|\Delta\mathbf{v}| \approx M_{\rm Pl}^2/M_0^2$ in each time interval $\Delta t \approx (M_0/M_{\rm Pl})l_{\rm Pl}$, which after time t leads to black hole velocity

$$|\mathbf{v}_{\rm BH}| \approx |\Delta\mathbf{v}|\sqrt{\frac{t}{\Delta t}} \sim \frac{M_{\rm Pl}^3}{M_0^{5/2}}\sqrt{t}, \tag{9}$$

whose direction does not change appreciably in each kick (and so is almost constant throughout the process). This implies that after time t ($M_0/M_{\rm Pl}^2 \ll t \lesssim M_0^3/M_{\rm Pl}^4$), the location of the black hole drifts in a random direction by an amount

$$|\mathbf{x}_{\rm BH}| \approx |\mathbf{v}_{\rm BH}|t \sim \frac{M_{\rm Pl}^3}{M_0^{5/2}}t^{3/2}. \tag{10}$$

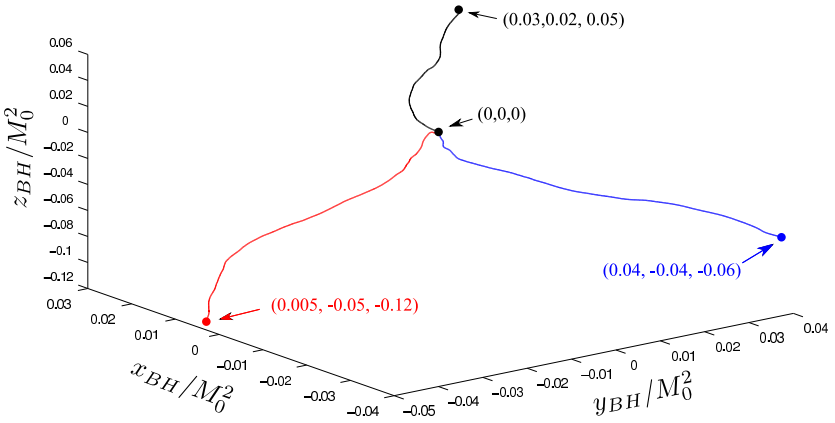


Figure 2: Typical paths of the black hole drifting in the three dimensional space $\mathbf{x}_{\rm BH} = (x_{\rm BH}, y_{\rm BH}, z_{\rm BH})$, normalized by M_0^2.

For $t \sim M_0^3/M_{\rm Pl}^4$, this gives $|\mathbf{v}_{\rm BH}|_{t\sim M_0^3/M_{\rm Pl}^4} \sim M_{\rm Pl}/M_0$ and

$$|\mathbf{x}_{\rm BH}|_{t\sim M_0^3} \sim \left(\frac{M_0}{M_{\rm Pl}}\right)^2 l_{\rm Pl}, \tag{11}$$

which is much larger than the Schwarzschild radius of the initial black hole $R_S = 2(M_0/M_{\rm Pl})l_{\rm Pl}$. By the time of the final evaporation, the velocity is further accelerated to $|\mathbf{v}_{\rm BH}| \sim 1$, but the final displacement is still of the order of Eq. (11).

To appreciate how large the value of Eq. (11) is, consider a black hole whose lifetime is of the order of the current age of the universe, $t_{\rm evap} \sim 10^{10}$ years. It has the initial mass of $M_0 \sim 10^{12}$ kg, implying the initial Schwarzschild radius of $R_S \sim 1$ fm. The result in Eq. (11) says that the displacement of such a black hole is $|\mathbf{x}_{\rm BH}| \sim 100$ km at the time of evaporation! The origin of this surprisingly large effect is the longevity of the black hole lifetime, $t_{\rm evap} \sim M_0^3$. For example, for a black hole of the solar mass $M = M_\odot \sim 10^{30}$ kg (i.e. $R_S \sim 1$ km), the evaporation time is $t_{\rm evap} \sim 10^{62}$ years—52 orders of magnitude longer than the age of the universe.

The probability distribution of $|\mathbf{x}_{\rm BH}|$ has the form

$$dP(|\mathbf{x}_{\rm BH}|) \propto |\mathbf{x}_{\rm BH}|^2 \exp\left(-c\frac{|\mathbf{x}_{\rm BH}|^2}{M_0^4}\right) d|\mathbf{x}_{\rm BH}|, \tag{12}$$

where c is a constant of $O(1)$, as implied by the central limit theorem, i.e. each component of $\mathbf{x}_{\rm BH}$ having the Gaussian distribution centered at zero with a width $\sim M_0^2$. In Fig. 2, we show typical paths of the black hole drift in three spatial dimensions. We see that the direction of the velocity stays nearly constant along a path, as suggested by the general analysis.

Quantum mechanically, the result described above implies that the state of the black hole becomes a superposition of terms in which the black hole exists in macroscopically different locations,

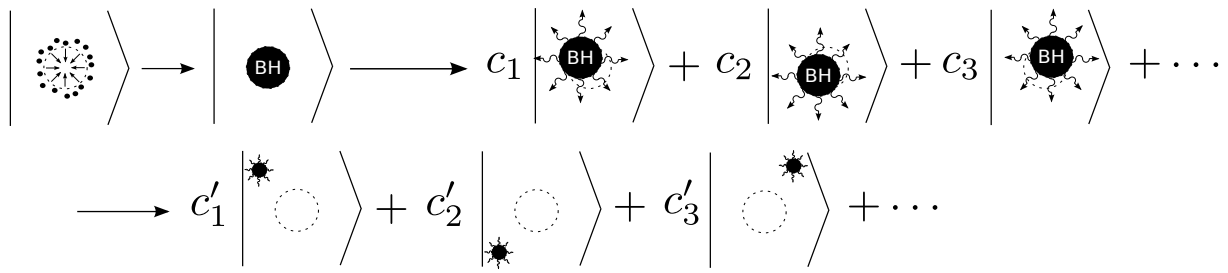


Figure 3: A schematic depiction of the evolution of a black hole state formed by a collapse of matter. After long time, the state will evolve into a superposition of terms representing the black hole to be in macroscopically different locations, even if the initial collapsing matter has a well-defined macroscopic configuration. The variation of the final locations in the evaporation timescale, $t \sim M_0^3$, is of order M_0^2, which is much larger than the Schwarzschild radius of the initial black hole, $R_S = 2M_0$.

even if the initial state forming the black hole is a classical object having a well-defined macroscopic configuration. At time $t \sim M_0^{7/3}$ after the formation (where t is the proper time measured at p), the uncertainty of the black hole location becomes of order M_0, comparable to the Schwarzschild radius of the original black hole. At the timescale of evaporation, $t \sim M_0^3$, the uncertainty is of order M_0^2, much larger than the initial Schwarzschild radius. This is illustrated schematically in Fig. 3. Note that each term in the figure still represents a superposition of terms having different phase space configurations of emitted Hawking quanta. Also, as shown in the appendix, each black hole at a fixed location is a superposition of black holes having macroscopically different angular momenta.

The evolution of the state depicted in Fig. 3 is obviously physical if we consider, for example, a super-Planckian scattering experiment. In this case, we will find that Hawking quanta emitted at the last stage of the evaporation will come from $\sim M_0^2$ away from the interaction point, according to the distribution in Eq. (12); and we can certainly measure this because the wavelengths of these quanta are much smaller than M_0^2, and the interaction point is defined clearly with respect to, e.g., the beam pipe. An important point here, however, is that the superposition nature of the black hole state is physical *even if there is no physical object other than the black hole*, e.g., the beam pipe. This is because the location of an object with respect to the origin p of the reference frame is a physically meaningful quantity in the framework of Ref. [5]. In other words, the superposition nature discussed here is an *intrinsic* property of the black hole state, not one arising only in relation to other physical objects.

While relative values of the moduli of coefficients in front of terms representing different black hole locations, e.g. $|c_1/c_2|$ in Fig. 3, are determined by the statistical analysis leading to Eq. (12), their relative phases are unconstrained by the analysis. Moreover, it is possible that there are higher order corrections to the moduli that are not determined by any semi-classical analysis.

These quantities, therefore, can contain the information about the initial state; i.e., they can reflect the details of the initial configuration of matter that has collapsed into the black hole. (This actually *should* be the case because a particular initial state leads to particular values for the relative phases because the Schrödinger equation is deterministic.) Together with the relative coefficients of terms representing different phase space configurations of emitted Hawking quanta for each black hole location (more precisely, their parts that are not fixed by semi-classical analyses, e.g. the relative phases), these quantities must be able to reproduce the initial state of the evolution by solving the appropriate Schrödinger equation backward in time.

2.4 Evolution in the covariant Hilbert space for quantum gravity

Let us now formulate more precisely how the black hole state, formed by a collapse of matter, evolves in the covariant Hilbert space for quantum gravity, Eq. (1). Recall that a Hilbert subspace $\mathcal{H}_{\mathcal{M}}$ in Eq. (1) corresponds to the states realized on a fixed semi-classical three-geometry $\mathcal{M}$ (more precisely, a set of three-geometries $\mathcal{M} = \{\mathcal{M}_i\}$ having the same boundary $\partial\mathcal{M}$). In our context, the relevant $\mathcal{M}$'s for spacetime with the black hole are specified by the location of the black hole $\mathbf{x}_{\rm BH}$ (which can be parameterized, e.g., by the direction $\{\theta, \phi\}$ and the affine length λ of the past-directed light ray connecting reference point p to the closest point on the stretched horizon) and the size of the black hole (which can be parameterized, e.g., by its mass M or area $\mathcal{A} = 16\pi M^2$). Here and below, we ignore the angular momentum of the black hole, for simplicity. We also need to consider the Hilbert subspace corresponding spacetime without the black hole, $\mathcal{H}_0$.

The part of $\mathcal{H}_{\rm QG}$ relevant to our problem here is then

$$\mathcal{H} = \left(\bigoplus_{\mathbf{x}_{\rm BH},\, M} \mathcal{H}_{\mathbf{x}_{\rm BH},\, M} \right) \oplus \mathcal{H}_0, \tag{13}$$

where $0 < M \leq M_0$, and we have used the notation in which $\mathbf{x}_{\rm BH}$ and M are discretized. The Hilbert subspace $\mathcal{H}_{\mathbf{x}_{\rm BH},\, M}$ consists of the factor associated with the black hole horizon $\mathcal{H}^{\rm horizon}_{\mathbf{x}_{\rm BH},\, M}$ and that with the rest $\mathcal{H}^{\rm bulk}_{\mathbf{x}_{\rm BH},\, M}$ (which represents the region outside the horizon):

$$\mathcal{H}_{\mathbf{x}_{\rm BH},\, M} = \mathcal{H}^{\rm horizon}_{\mathbf{x}_{\rm BH},\, M} \otimes \mathcal{H}^{\rm bulk}_{\mathbf{x}_{\rm BH},\, M}. \tag{14}$$

According to the Bekenstein-Hawking entropy, the size of the horizon factor is given by

$$\dim \mathcal{H}^{\rm horizon}_{\mathbf{x}_{\rm BH},\, M} = e^{\frac{\mathcal{A}_{\rm BH}}{4}} = e^{4\pi M^2}, \tag{15}$$

regardless of $\mathbf{x}_{\rm BH}$. Because of this, Hilbert space factors $\mathcal{H}^{\rm horizon}_{\mathbf{x}_{\rm BH},\, M}$ for different $\mathbf{x}_{\rm BH}$ are all isomorphic with each other, which allows us to view $\mathcal{H}^{\rm horizon}_{\mathbf{x}_{\rm BH},\, M}$ for any fixed $\mathbf{x}_{\rm BH}$ as the intrinsic structure of the black hole.

Now, right after the formation of the black hole, which we assume to have happened at $\mathbf{x}_{\rm BH,0}$ at $t = 0$, the system is in a state that is an element of $\mathcal{H}_{\mathbf{x}_{\rm BH,0},\, M_0}$. In the case of Eq. (4)

$$|\Psi(0)\rangle \equiv \left|{\rm BH}^0_A\right\rangle \in \mathcal{H}_{\mathbf{x}_{\rm BH,0},\, M_0}. \tag{16}$$

This state then evolves into a superposition of states in different $\mathcal{H}_{\mathcal{M}}$'s.[4] At time t, the state of the system can be written as

$$|\Psi(t)\rangle = \sum_{\mathbf{x}_{\rm BH}} \alpha^t_{\mathbf{x}_{\rm BH}} \left|\phi^t_{\mathbf{x}_{\rm BH}}\right\rangle, \tag{17}$$

where $\left|\phi^t_{\mathbf{x}_{\rm BH}}\right\rangle \in \mathcal{H}_{\mathbf{x}_{\rm BH},\, M(t)}$, and we have ignored possible fluctuations of the black hole mass at a fixed time t, for simplicity. (Including this effect is straightforward; we simply have to add terms corresponding to $\mathcal{H}_{\mathbf{x}_{\rm BH},\, M}$ with $M \neq M(t)$.) The state $\left|\phi^t_{\mathbf{x}_{\rm BH}}\right\rangle$ contains the horizon *and* other degrees of freedom, according to Eq. (14). We can expand it in some basis in $\mathcal{H}^{\rm horizon}_{\mathbf{x}_{\rm BH},\, M(t)}$ (e.g. the one spanned by states having well-defined numbers of Hawking quanta emitted afterward) or in some basis in $\mathcal{H}^{\rm bulk}_{\mathbf{x}_{\rm BH},\, M(t)}$ (e.g. the one spanned by states having well-defined phase space configurations of already emitted Hawking quanta). In either case, it takes the form

$$\left|\phi^t_{\mathbf{x}_{\rm BH}}\right\rangle = \sum_n \beta^t_n \left|{\rm BH}^t_{\mathbf{x}_{\rm BH},n}\right\rangle \otimes \left|\psi^t_{\mathbf{x}_{\rm BH},n}\right\rangle, \tag{18}$$

where $\left|{\rm BH}^t_{\mathbf{x}_{\rm BH},n}\right\rangle \in \mathcal{H}^{\rm horizon}_{\mathbf{x}_{\rm BH},\, M(t)}$ and $\left|\psi^t_{\mathbf{x}_{\rm BH},n}\right\rangle \in \mathcal{H}^{\rm bulk}_{\mathbf{x}_{\rm BH},\, M(t)}$. Plugging this into Eq. (17) and defining

$$a^t_i \equiv \alpha^t_{\mathbf{x}_{\rm BH}} \beta^t_n, \tag{19}$$

where $i \equiv \{\mathbf{x}_{\rm BH}, n\}$, we reproduce the third expression in Eq. (4). In this formulation, the statement that the black hole state is a superposition of macroscopically different geometries refers to the fact that coefficients $|\alpha^t_{\mathbf{x}_{\rm BH}}|$ have a significant support in a wide range of $\mathbf{x}_{\rm BH}$ extending beyond the original Schwarzschild radius M_0.

2.5 What does a physical observer actually see?

We have found that a late black hole state is far from a semi-classical state in which the spacetime has a fixed geometry; rather, it involves a superposition of macroscopically different geometries. Does this mean that a physical observer sees something very different from what the usual picture based on general relativity predicts?

The answer is no. To understand this, let us consider a physical observer watching the evaporation process from a distance by measuring (all or parts of) the emitted Hawking quanta. For

[4] This is precisely analogous to the case of e^+e^- scattering, in which the initial state $|e^+e^-\rangle \in \mathcal{H}_2$ evolves into a superposition of states in different $\mathcal{H}_n$'s, e.g. $|e^+e^-\rangle \to c_e\,|e^+e^-\rangle + \cdots + c_{ee}\,|e^+e^-e^+e^-\rangle + \cdots$, where $\mathcal{H}_n$ is the n-particle subspace of the entire Fock space: $\mathcal{H} = \bigoplus_n \mathcal{H}_n$.

simplicity, we consider that he/she does that using usual measuring devices, e.g. by locating photomultipliers around the black hole from which he/she collects the data. This leads to an entanglement between the system and the observer (or his/her brain states). And because the interactions leading to it are local, the observer is entangled with the basis in $\mathcal{H}^{\rm bulk}_{\mathbf{x}_{\rm BH},\,M}$ spanned by the states that have well-defined phase space configurations of emitted Hawking quanta (within the errors dictated by the uncertainty principle) and well-defined locations for the black hole (since the black hole location can be inferred from the momenta of the Hawking quanta) [5]. Namely, the combined state of the black hole and the observer evolves as

$$\left|{\rm BH}^0_A\right\rangle \otimes \left|\,\begin{array}{c}\circ\\ /\!\!\uparrow\!\!\backslash\end{array}\right\rangle \longrightarrow \sum_{\mathbf{x}_{\rm BH},\,n} a^t_{\mathbf{x}_{\rm BH},n}\left|{\rm BH}^t_{\mathbf{x}_{\rm BH},n}\right\rangle \otimes \left|\psi^t_{\mathbf{x}_{\rm BH},n}\right\rangle \otimes \left|\,\begin{array}{c}\circ\\ /\!\!\uparrow\!\!\backslash\end{array}\boxed{\bullet^{\mathbf{x}_{\rm BH}}_{n}}\right\rangle, \tag{20}$$

where $\left|\psi^t_{\mathbf{x}_{\rm BH},n}\right\rangle$ represents the state in which the black hole is in a well-defined location $\mathbf{x}_{\rm BH}$ and Hawking quanta have a well-defined phase space configuration n. The last factor in the right-hand side implies that the observer recognized that the black hole is at $\mathbf{x}_{\rm BH}$ and the configuration of emitted Hawking quanta is n.

Since terms in the right-hand side of Eq. (20) have macroscopically different configurations, e.g. the brain state of the observer differs, their mutual overlaps are exponentially suppressed (e.g. by $\sim \prod_{i=1}^{N}\epsilon_i$, where $\epsilon_i < 1$ is the overlap of each atom and N the total number of atoms). The observer in each term (or branch), therefore, sees his/her own universe; i.e., the interferences between different terms are negligible. For any of these observers, the behavior of the black hole is controlled by semi-classical physics (but with the backreaction of the emission taken into account). For instance, they all see that the black hole keeps emitting Hawking quanta consistent with the thermal spectrum with temperature $T_H(t) = 1/8\pi M(t)$, and that it drifts in a fixed direction as a result of backreactions, eventually evaporating at a location $\sim M_0^2$ away from that of the formation. A single observer cannot predict the direction to which the black hole will drift, reflecting the fact that the entire state is a superposition of terms having different $(\mathbf{x}_{\rm BH} - \mathbf{x}_{\rm BH,0})/|(\mathbf{x}_{\rm BH} - \mathbf{x}_{\rm BH,0})|$, but all these observers find a set of common properties for the black hole, including the relation between T_H and M.[5]

It is these "intrinsic properties" of the black hole that the semi-classical gravity on a fixed Schwarzschild geometry (in which the black hole is located at the "center") really describes. A physical observer watching the evolution does not see anything contradicting what is implied by the semi-classical analysis about these intrinsic properties. This is true despite the fact that the full quantum state obtained by evolving collapsing matter that initially had a well-defined configuration takes the form in Eqs. (17, 18), which involves a superposition of macroscopically

[5]More precisely, there are rare observers who find deviations from these relations, but the probability for that to happen is exponentially suppressed.

different geometries and is very different at late times from a "semi-classical state" having a fixed geometry.

2.6 Can a physical observer recover the information?

The black hole evaporation process is often compared with burning a book in classical physics: if we measure all the details of the emitted Hawking quanta, we can recover the initial state from these data by solving the Schrödinger equation backward in time. Is this correct?

It is true that if we know the coefficients of all the terms in a state when it is expanded in a fixed basis, e.g. $\{a_i^\infty\}$ in Eq. (4), then unitarity must allow us to recover the initial state unambiguously. However, a physical observer measuring Hawking radiation from black hole evaporation can *never* obtain the complete information about these coefficients, even if he/she measures *all* the radiation quanta. In the state in Eq. (20), for example, a physical observer "lives" in *one* of the terms in the right-hand side and, therefore, cannot have the information about the coefficients of the other terms. The other terms are already decohered—or "decoupled"—so that they are other worlds/universes for the observer.

In fact, the situation is exactly the same in usual scattering experiments. Consider two initial states $|e^+e^-\rangle$ and $|\mu^+\mu^-\rangle$ with the same $\sqrt{s}$ ($> 2m_\tau$) and angular momentum. They evolve as

$$\left|e^+e^-\right\rangle \longrightarrow a_1\left|e^+e^-\right\rangle + a_2\left|\mu^+\mu^-\right\rangle + a_3\left|\tau^+\tau^-\right\rangle + \cdots , \tag{21}$$

$$\left|\mu^+\mu^-\right\rangle \longrightarrow b_1\left|e^+e^-\right\rangle + b_2\left|\mu^+\mu^-\right\rangle + b_3\left|\tau^+\tau^-\right\rangle + \cdots , \tag{22}$$

where we have ignored the momenta and spins of the final particles. The information about an initial state is in the *complete* set of coefficients in the final superposition state; i.e., if we know the entire $\{a_i\}$ (or $\{b_i\}$), then we can recover the initial state by solving the evolution equation backward. However, if a physical observer measures a final state, e.g., as $\tau^+\tau^-$, how can he/she know that it has arisen from e^+e^- or $\mu^+\mu^-$ scattering? In general, if an observer measures the final outcome of a process, he/she will be entangled with one of the terms in the final state (in the above case, $|\tau^+\tau^-\rangle$), so there is no way that he/she can learn all the coefficients in the final state.

The situation does not change even if the observer uses a carefully-crafted quantum device which, upon interacting with the radiation, is entangled *not* with a well-defined phase space configuration of the radiation quanta but with a macroscopic superposition of those configurations. In this case, the basis of the final state to which the observer is entangled may be changed, but it still cannot change the fact that he/she will be entangled with *one* of the terms in the final state, i.e., he/she will measure *a* possible outcome among all the possibilities.

Therefore, in quantum mechanics, an observer can never recover the initial state by observing the final state. The statement that the final state of an evolution contains all the information about the initial state is *not* the same as the statement that a physical observer measuring the

final state can recover the initial state if he/she measures a system with high enough (or even infinite) precision. The only way that an observer can test the relation between the initial and final states is to *create* the same initial state many times and perform multiple (including quantum) measurements on the final states. (Note that creating many initial states in this context differs from producing a copy of a generic unknown state, which is prohibited by the quantum no-cloning theorem [26].) A single system does not allow for doing this, no matter how high the precision of the measurement is, and no matter how clever the measurement device is.

3 Complementarity as a Reference Frame Change

So far, we have been describing the formation and evaporation of a black hole from a distant reference frame. In this reference frame, the complete description of the process is obtained in the spacetime region outside and on the (time-like) stretched horizon, where intrinsically quantum gravitational—presumably stringy (such as fuzzball [20])—effects become important. What then is the significance of the interior of the black hole horizon, where we expect to have regular low-curvature spacetime according to general relativity?

As discussed in Ref. [5], and implied by the original complementarity picture [4], a description of the internal spacetime is obtained (only) after changing the reference frame. An important point is that the reference frame change is represented as a unitary transformation acting on a quantum state, so if we want to discuss the precise mapping between the pictures based on different reference frames, then we need to keep all the terms in the state. In this section, we carefully study issues associated with the reference frame changes, especially in describing an old black hole. We also elucidate the analysis in Ref. [10], demonstrating that the framework does not suffer from the firewall paradox discussed in Ref. [23].

3.1 Describing the black hole interior

Suppose collapsing matter, which initially had a well-defined classical configuration, forms a black hole, which then eventually evaporates. In a distant reference frame, this process is described as in Eq. (4), which we denote by $|\Psi(t)\rangle$. How does the process look from a different reference frame?

Since a reference frame can be any freely falling (locally Lorentz) frame, the new description can be obtained by performing a translation, rotation, or boost on a quantum state at fixed t [5]. In general, the state on which these transformations act, however, contains the horizon degrees of freedom as well as the bulk ones. How do they transform under the transformations?

We do not know the microscopic description of the horizon degrees of freedom or their explicit transformations under the reference frame changes. Nevertheless, we can know which spacetime regions are transformed to which horizon degrees of freedom, and vice versa, by *assuming* that

the global spacetime picture in semi-classical gravity is consistent with the one obtained by a succession of these reference frame changes. Here we phrase this in the form of a hypothesis:

> **Complementarity Hypothesis:** The transformation laws of a quantum state under the reference frame changes are consistent with those obtained in the global spacetime picture based on general relativity. In particular, the transformation laws between the horizon and bulk degrees of freedom are constrained by this requirement.

As discussed in Refs. [5, 6], this hypothesis is fully consistent with the holographic principle formulated in the form of the covariant entropy conjecture [3]. Specifically, the dimension of the Hilbert space representing horizon degrees of freedom and that representing the corresponding spacetime region before (or after) a transformation are the same for general spacetimes, including the cosmological ones, as it should be. Alternatively, we can take a view that if we require that the above hypothesis is true in the covariant Hilbert space $\mathcal{H}_{\rm QG}$, then the covariant entropy conjecture is obtained as a consequence.

Let us now consider a reference frame change induced by a boost performed at some early time $t_{\rm boost} < 0$ (before the black hole forms at $t = 0$) in such a way that the origin p of the reference frame enters the black hole horizon at some late time $t_{\rm enter} > 0$. In this subsection, we focus on the case

$$t_{\rm enter} \ll M_0^{7/3}, \tag{23}$$

so that the uncertainty of the black hole location at the time when p enters the horizon is negligible, and we ignore the (exponentially) small probability that p misses the horizon. (The possibility of p missing the horizon becomes important when we discuss the description of the interior of an older black hole.)

Recall that quantum states in the present framework represent physical configurations on the past light cone of p (in and on the apparent horizon) when they allow for spacetime interpretation, i.e. when the curvature at p is smaller than the Planck scale. Therefore, the spacetime region represented by the state of the system after the reference frame change

$$|\Psi'(t)\rangle = e^{-iH(t-t_{\rm boost})} U_{\rm boost} e^{iH(t-t_{\rm boost})} |\Psi(t)\rangle \,, \tag{24}$$

where $U_{\rm boost}$ is the boost operator represented in $\mathcal{H}_{\rm QG}$, corresponds to the shaded region in the left panel of Fig. 4. Specifically, $|\Psi'(t)\rangle$ at

$$t < t_{\rm enter} + t_{\rm fall} \tag{25}$$

describes this region, with $t_{\rm fall} \approx O(M_0)$ being the time needed for p to reach the singularity after it passes the horizon. After $t = t_{\rm enter} + t_{\rm fall}$, the state evolves in the Hilbert subspace $\mathcal{H}_{\rm sing}$, which consists of states that are associated with spacetime singularities and thus do not allow for

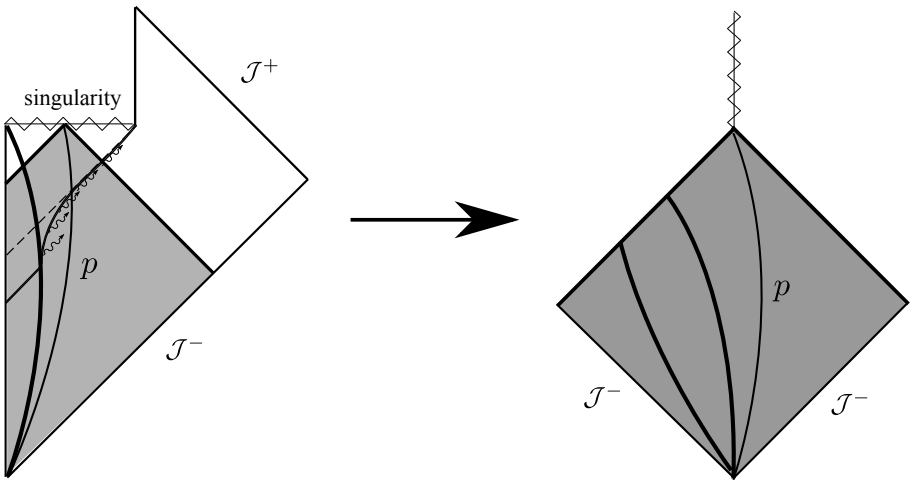


Figure 4: The left panel shows the standard global spacetime picture for the formation and evaporation of a black hole, with the shaded region representing the spacetime region described by an infalling reference frame. (The trajectory of the origin, p, of the reference frame is also depicted.) As discussed in the text, this is the *entire* spacetime when the system is described in this reference frame, so its conformal structure is in fact as in the right panel. Here, the wavy line with a solid core represents singularity states.

spacetime interpretation. The detailed properties of these "intrinsically quantum gravitational" states are unknown, except that $\dim \mathcal{H}_{\rm sing} = \infty$, implying that generic singularity states do not evolve back to the usual spacetime states [5].

In the right panel of Fig. 4, we depict the causal structure of the spacetime as viewed from the new reference frame. Because of the lack of the spherical symmetry, we have depicted the region swept by two past-directed light rays emitted from p in the opposite directions (while in the left panel we have depicted only the region with fixed angular variables with respect to the center of mass of the system). The singularity states are represented by the wavy line with a solid core at the top. Note that, as in the case of the description in the distant reference frame (depicted in Fig. 1), this is the *entire* spacetime region when the system is described in this infalling reference frame—the non-shaded region in the left panel simply does not exist. (Including the non-shaded region, indeed, is overcounting as indicated by the standard argument of information cloning in black hole physics.) A part of the non-shaded region appears if we change the reference frame, but only at the cost of losing some of the shaded region. The global spacetime picture in the left panel appears only if we "patch" the views from different reference frames, which, however, grossly overcounts the correct quantum degrees of freedom.

There are two comments. First, the reference frame change considered here is (obviously) only *a* reference frame change among possible (continuously many) reference frame changes, all of which lead to different descriptions of the same physical process. Second, a unitary transformation

representing this reference frame change,

$$U(t) = e^{-iH(t-t_{\rm boost})} U_{\rm boost} e^{iH(t-t_{\rm boost})}, \tag{26}$$

does not close in the Hilbert space $\mathcal{H}$ in Eq. (13), although it closes in the whole covariant Hilbert space $\mathcal{H}_{\rm QG}$. Before the reference frame change, the evolution of the state is given by a trajectory in $\mathcal{H} = (\oplus_{\mathbf{x}_{\rm BH},\, M} \mathcal{H}_{\mathbf{x}_{\rm BH},\, M}) \oplus \mathcal{H}_0$. The action of $U(t)$ maps this into a trajectory in

$$\mathcal{H}' = \mathcal{H}_0 \oplus \mathcal{H}_{\rm sing}, \tag{27}$$

with $|\Psi'(-\infty)\rangle \in \mathcal{H}_0$ and $|\Psi'(+\infty)\rangle \in \mathcal{H}_{\rm sing}$.[6] As a result, in this new reference frame, the evolution of the system does not allow for an S-matrix description in $\mathcal{H}_0$ (or $\mathcal{H}_{\rm Minkowski}$), although it still allows for an "S-matrix" description in the whole $\mathcal{H}_{\rm QG}$ (or in $\mathcal{H}_{\rm Minkowski} \oplus \mathcal{H}_{\rm sing}$), which contains the singularity states in $\mathcal{H}_{\rm sing}$.

3.2 Complementarity for an old black hole

Let us now try to describe the interior of an older black hole, specifically the spacetime inside the black hole horizon after a time $> O(M_0^{7/3})$ is passed since the formation. To do this, we can consider performing a boost at time $t_{\rm boost} < 0$ on $|\Psi(t)\rangle$ in such a way that p enters the black hole horizon at time $t_{\rm enter} \gg M_0^{7/3}$. What does the resultant state $|\Psi'(t)\rangle$ look like?

As discussed in the previous subsection, this can be done by applying an operator of the form of Eq. (26) on $|\Psi(t)\rangle$, where $U_{\rm boost}$ now represents a different amount of boost than the one considered before. In general, the relation between the states before and after a reference frame change is highly nontrivial. For example, time t is measured by the proper time at p, but relations between the proper times of the two frames depend on the geometries as well as the paths of p therein. Therefore, various terms in $|\Psi'(t)\rangle$ for a fixed t may correspond to terms in $|\Psi(t)\rangle$ of different t's. Without knowing the explicit form of H and $U_{\rm boost}$ represented in the whole $\mathcal{H}_{\rm QG}$, which includes the horizon degrees of freedom, how can we know the form of the state after the transformation?

According to our complementarity hypothesis, the probability of finding a certain history for the evolution of geometry must agree in the two pictures before and after the reference frame change if the geometries are appropriately transformed, i.e. according to the global spacetime picture in general relativity. To elucidate this, let us consider the black hole evolution described in Eqs. (17, 18) in a distant reference frame, and ask what is the probability that the black hole follows a particular path $\mathbf{r}(t)$ in a time interval between t_I and t_F within the error $|\Delta\mathbf{r}| < \epsilon(t)$. For simplicity, we do this by requiring that the black hole satisfies the above conditions at discretized

[6] Note that $\mathcal{H}_0$ contains a set of states that represent three-geometries whose *boundary* (at an infinity) is that of the flat space, i.e. a two-dimensional section of $\mathcal{J}^-$.

times t_i; $i = 0, \cdots, N$ ($\gg 1$), with $t_0 \equiv t_I$ and $t_N \equiv t_F$. The probability is then given by

$$P = \prod_{i=0}^{N} \left(\sum_{|\Delta \mathbf{r}|<\epsilon_i} |\alpha^{t_i}_{\mathbf{r}_i+\Delta\mathbf{r}}|^2 \right), \tag{28}$$

where $\mathbf{r}_i \equiv \mathbf{r}(t_i)$ and $\epsilon_i \equiv \epsilon(t_i)$. This provides the probability of a particular semi-classical history to appear, given the state $|\Psi(t)\rangle$. We can now ask a similar question for the state $|\Psi'(t)\rangle$: what is the probability of having the black hole to follow the trajectory $\mathbf{r}'(t)$ between t'_I and t'_F within the error $\epsilon'(t)$? The resulting probability is

$$P' = \prod_{i=0}^{N} \left(\sum_{|\Delta \mathbf{r}'|<\epsilon'_i} |\alpha^{t_i}_{\mathbf{r}'_i+\Delta\mathbf{r}'}|^2 \right), \tag{29}$$

where $t_0 = t'_I$ and $t_N = t'_F$. The complementarity hypothesis in the previous subsection asserts that the two probabilities are the same

$$P = P', \tag{30}$$

if the relation between $\{\mathbf{r}(t), \epsilon(t), t_I, t_F\}$ and $\{\mathbf{r}'(t), \epsilon'(t), t'_I, t'_F\}$ is the one obtained by performing the corresponding transformation in general relativity on the semi-classical background selected by Eq. (28).

The above analysis implies that when we perform a boost on $|\Psi(t)\rangle$ at an early time $t_{\rm boost} < 0$, trying to describe the interior of an old black hole with $t_{\rm enter} \gg M_0^{7/3}$, then the resultant state can only be a *superposition* of infalling and distant descriptions of the process, since in most of the semi-classical histories represented by $|\Psi(t)\rangle$, the trajectory of p obtained by the boost will miss the black hole horizon because of the large uncertainty of the black hole location. Namely, complementarity obtained by this reference frame change is the one between the distant description and the superposition of the infalling and distant descriptions specified by the state $|\Psi'(t)\rangle$. This is illustrated schematically in Fig. 5.

Is it possible to obtain a direct correspondence between the interior and exterior of an old black hole, without involving a superposition? This can be done if we focus only on a term in $|\Psi(t)\rangle$ in which p just misses the black hole horizon, with the smallest distance between p and the horizon achieved at some time $t_{\rm min} \gg M_0^{7/3}$. We can then evolve this term slightly backward in time, to $t_{\rm boost} = t_{\rm min} - \epsilon$ ($\epsilon \ll M_0^{7/3}$), and perform a boost there so that p enters into the horizon at some time after $t_{\rm boost}$. In this way, the correspondence between the terms representing the interior and exterior can be obtained. An important point, however, is that neither of these terms can be obtained by evolving initial collapsing matter that had a well-defined classical configuration (which would lead to a superposition of the black hole in vastly different locations). Rather, by evolving the state further back beyond $t_{\rm boost}$, we would obtain a superposition of states each of which

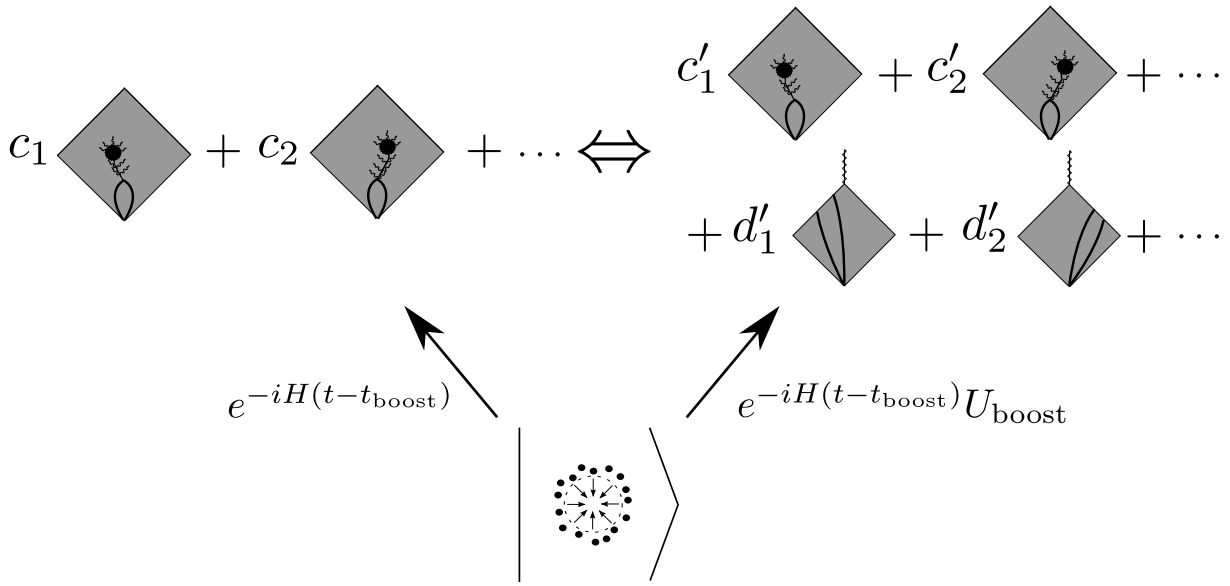


Figure 5: A schematic picture of a relation between the two descriptions based on two different reference frames of an old black hole formed by collapsing matter that initially had a well-defined classical configuration. In one reference frame, the black hole is viewed from outside, and the state becomes a superposition of black holes in different locations at late times (depicted schematically in the left-hand side). In the other reference frame, obtained by acting $U_{\rm boost}$ on the state at $t_{\rm boost}$, the reference point p enters the black hole horizon at late time $t_{\rm enter} \gg M_0^{7/3}$, allowing for a description of internal spacetime (in the right-hand side). This, however, happens only for some of the terms, depicted in the second line, since p misses the horizon in most of the terms because of the large uncertainty of the black hole location, i.e. $\sum_i |d'_i|^2 \ll \sum_i |c'_i|^2$.

represents collapsing matter with a well-defined classical configuration. (This state would have finely-adjusted coefficients so that after evolving to $t_{\rm boost} \sim t_{\rm min}$, the black hole is in a well-defined location with respect to p.) This situation is illustrated in Fig. 6.

The discussion above implies that there is no well-defined complementarity map between the interior and exterior of an old black hole throughout the course of the black hole evolution within the purely semi-classical picture. Such a map must involve a superposition of semi-classical geometries at some point in the evolution. We note that while the state in the intermediate stage of the evolution can be a superposition of elements in $\mathcal{H}_0$, $\mathcal{H}_{\mathbf{x}_{\rm BH},M}$, and $\mathcal{H}_{\rm sing}$, it becomes a superposition of elements in $\mathcal{H}_0$ and $\mathcal{H}_{\rm sing}$ at $t \to \infty$. Therefore, the "S-matrix" description discussed in the previous subsection is still available in this case in the Hilbert space of $\mathcal{H}_{\rm Minkowski} \oplus \mathcal{H}_{\rm sing}$.

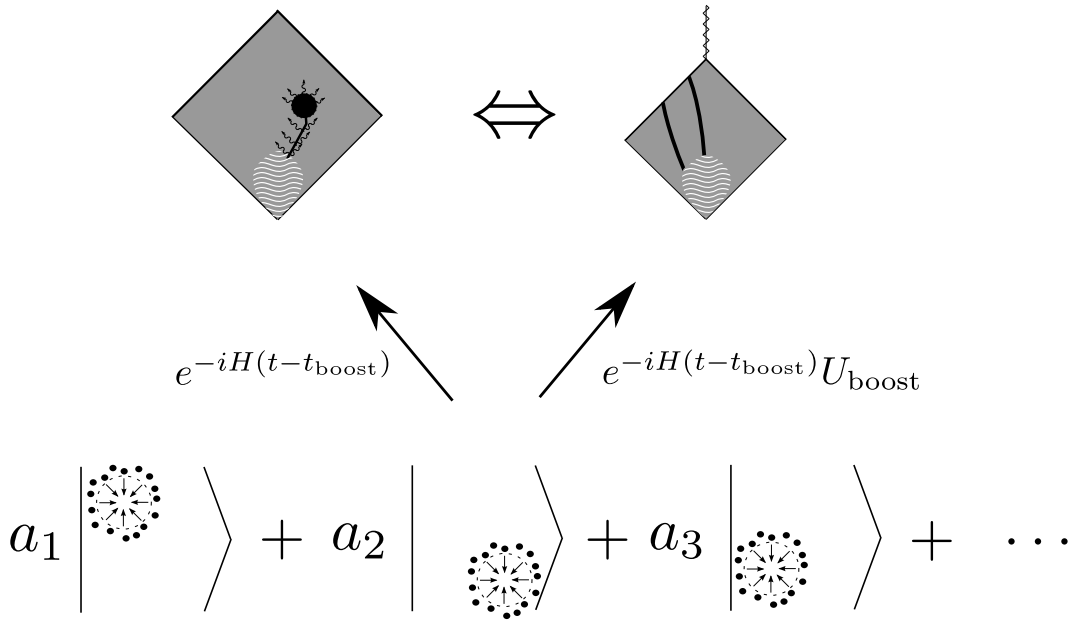


Figure 6: A complementarity relation between the internal and external descriptions of an old black hole can be obtained if we consider a state in which the black hole has a well-defined semi-classical configuration at late time $t_{\rm enter}$. Such a state, however, can arise through evolution only if we consider a special initial state in which coefficients a_i of the terms representing well-defined configurations of collapsing matter are finely-adjusted so that the state represents a black hole in a well-defined location at late time $\sim t_{\rm enter}$. (The regions with wavy white lines indicate superpositions of classical geometries.)

3.3 On firewalls (or firewall as an exponentially unlikely phenomenon)

The complementarity picture has recently been challenged by the "firewall paradox" posed by Almheiri, Marolf, Polchinski, and Sully (AMPS) in Ref. [23]. Here we elucidate the discussion of Ref. [10], refuting one of the AMPS arguments based on measurements of early Hawking radiation. AMPS also has another argument using entropy relations, building on an earlier discussion by Mathur [18]. Here we focus only on the first AMPS argument based on measurements. Addressing the second one requires an additional assumption about the decoherence structure of microscopic degrees of freedom of the horizon, beyond what we have postulated in this paper [27].

In essence, the argument by AMPS goes as follows. Consider an old black hole with $t > t_{\rm Page}$ that has formed from collapse of some pure state. Because of the purity of the state, the system as viewed from a distant reference frame can be written as

$$|\Psi\rangle = \sum_i c_i \, |i\rangle \otimes |\psi_i\rangle \,, \tag{31}$$

where $|i\rangle \in \mathcal{H}_{\rm horizon}$ and $|\psi_i\rangle \in \mathcal{H}_{\rm rad}$ represent degrees of freedom associated with the horizon region and the emitted Hawking quanta. (For simplicity we have suppressed the time index, which is not

essential for the discussion here.) For a black hole older than t_{Page}, the dimensions of the Hilbert space factors satisfy $\dim \mathcal{H}_{\text{horizon}} \ll \dim \mathcal{H}_{\text{rad}}$. Therefore, states $|\psi_i\rangle$ for different i are expected to be nearly orthogonal, and one can construct a projection operator P_i that acts only on $\mathcal{H}_{\text{rad}}$ (not on $\mathcal{H}_{\text{horizon}}$) but selects a term in Eq. (31) corresponding to a specific state $|i\rangle$ in $\mathcal{H}_{\text{horizon}}$ when operated on $|\Psi\rangle$:

$$P_i \, |\Psi\rangle \propto |i\rangle \otimes |\psi_i\rangle \, . \tag{32}$$

The point is that one can construct such an operator for an arbitrary state $|i\rangle$ in $\mathcal{H}_{\text{horizon}}$.

AMPS argue that since the infalling observer can access the early radiation, he/she can select a particular term in Eq. (31) by making a measurement on those degrees of freedom. In particular, they imagine that such a measurement would select a term in which $|i\rangle$ in $\mathcal{H}_{\text{horizon}}$ is an eigenstate, $|\tilde{\imath}\rangle$, of the number operator, $b^\dagger b$, of a Hawking radiation mode that will escape from the horizon region:

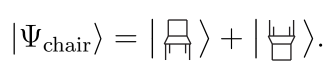
$$b^\dagger b \, |\tilde{\imath}\rangle \propto |\tilde{\imath}\rangle \, . \tag{33}$$

If this were true, then the infalling observer must find physics represented by $|\tilde{\imath}\rangle$ near the horizon, and since an eigenstate of $b^\dagger b$ cannot be a vacuum for the infalling modes a_ω, related to b by

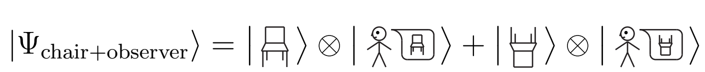
$$b = \int_0^\infty d\omega \left(B(\omega) a_\omega + C(\omega) a_\omega^\dagger \right) \tag{34}$$

with some functions $B(\omega)$ and $C(\omega)$, the infalling observer must experience nontrivial physics at the horizon (i.e. $a_\omega \, |\tilde{\imath}\rangle \neq 0$ for infalling modes with the frequencies much larger than the inverse horizon size). This obviously contradicts what is expected from general relativity.

As discussed in Ref. [10], this argument misses the fact that the emergence of a classical world in the underlying quantum world is *a dynamical process* dictated by unitary evolution of a state, and *not* something we can impose from outside by acting with some projection operator on the state. In particular, the existence of the projection operator P_i for an arbitrary i does not imply that a measurement performed by a *classical* observer, which general relativity is supposed to describe, can pick up the corresponding state $|i\rangle$. To understand this point, consider a state representing a superposition of upward and downward chairs (relative to some other object, e.g. the ground, which we omit):

 (35)

An observer interacting with this system evolves following the unitary, deterministic Schrödinger equation; in particular, the combined chair and observer state becomes

 (36)

This does *not* lead to a classical world in which the chair is in a superposition state but to two different worlds in which the chair is upward and downward, respectively. Namely, the measurement

is performed in the particular basis $\{|\,\text{⊓}\,\rangle, |\,\text{⊔}\,\rangle\}$, which is *determined by the dynamics*—the existence of an operator projecting onto a superposition chair state does not mean that a measurement is performed in that basis. For sufficiently macroscopic object/observer, the appropriate basis for measurements is almost always (see below) the one in which they have well-defined configurations in classical phase space (within some errors, which must exist because of the uncertainty principle). This is because the Hamiltonian has the form that is local in spacetime [5].

In the specific context of the firewall argument, a measurement of the early radiation by an infalling, classical observer will select a state in $\mathcal{H}_{\rm rad}$ that has a well-defined classical configuration of Hawking radiation quanta $|\psi_I\rangle$, because interactions between him/her and the quanta are local. As shown in Ref. [10], the state $|\psi_I\rangle$ selected in this way is expected *not* to be a state that is maximally entangled with $|\tilde{i}\rangle$, i.e. $|\psi_{\tilde{i}}\rangle$. In other words, the basis in $\mathcal{H}_{\rm rad}$ selected by entanglement with the eigenstates of $b^\dagger b$ is different from the one selected by entanglement with the infalling classical observer. This implies that in the world described by the infalling observer, i.e. in a term of the entire state in which the observer has a well-defined classical configuration, the state in $\mathcal{H}_{\rm rad}$ is always in a superposition of $|\psi_{\tilde{i}}\rangle$'s for different $\tilde{i}$'s, so that the corresponding state in $\mathcal{H}_{\rm horizon}$ is *not* an eigenstate of $b^\dagger b$. In particular, there is no contradiction if the state is a simultaneous eigenstate of a_ω's with the eigenvalue zero (up to exponentially small corrections) as implied by general relativity.[7]

One might ask what happens if we prepare a carefully-crafted *quantum* device that will be entangled with one of the $|\psi_{\tilde{i}}\rangle$'s and then send a signal, e.g. a particle, toward the horizon. Wouldn't that particle see a firewall at the horizon? Yes, that particle *does* see a firewall, but it is not a phenomenon described by general relativity, a theory for a classical world. First of all, in order for the device to be entangled with $|\psi_{\tilde{i}}\rangle$, it must collect the information in the Hawking quanta to learn that they are in the $|\psi_{\tilde{i}}\rangle$ state, and since the information is encoded in a highly scrambled form, it must be very large collecting many quanta spread in space without losing their coherence. This implies that the device must have been in an extremely carefully-chosen superposition state of different classical configurations at the beginning of the evolution (which is also clear from the fact that $\dim \mathcal{H}_{\rm horizon} \ll \dim \mathcal{H}_{\rm rad} \lesssim \dim \mathcal{H}_{\rm device}$, where $\mathcal{H}_{\rm device}$ is the Hilbert space factor associated with the device). Now, we know that if the initial state is extremely fine-tuned, an extremely unlikely event can happen. For example, if the initial locations and velocities of the molecules are finely tuned, ink dissolved in a water tank can spontaneously come to a corner. The firewall phenomenon is analogous to this kind of phenomena.

More specifically, we can ask what is the degree of fine-tuning needed to see firewalls. The

[7] Of course, we are only showing here that the argument of AMPS breaks down, i.e. the complementarity picture—or the semi-classical picture in the infalling reference frame—is *consistent*. Proving it, e.g. showing that the state is indeed a simultaneous eigenstate of a_ω's, would require an understanding of the underlying theory of quantum gravity. As stated explicitly in Section 3.1, complementarity is a hypothesis, which we argue is a consistent one [10].

amount of fine-tuning, i.e. the probability for a randomly-chosen device state to see the firewall, is estimated as

$$p_{\rm fw} \sim \dim \mathcal{H}_{\rm horizon} \sqrt{\dim \mathcal{H}_{\rm rad}}\, \epsilon^{\dim \mathcal{H}_{\rm rad}}, \qquad \epsilon < 1, \tag{37}$$

which is obtained by asking the probability of a randomly-chosen basis state for measuring Hawking radiation to agree with one of the $|\psi_{\tilde{i}}\rangle$'s; see the appendix of Ref. [10] for a similar calculation. Since $\dim \mathcal{H}_{\rm rad} \gtrsim e^{2\pi M_0^2}$ after the Page time, this is double-exponentially suppressed in a macroscopic number $M_0^2 \gg 1$:

$$p_{\rm fw} \lesssim \epsilon^{e^{2\pi M_0^2}} \lll 1, \tag{38}$$

analogous to the case of having an entropy decreasing process in a system with large degrees of freedom.

Note that if we ask the amount of fine-tuning needed for ink dissolved in a water tank to come to a corner in the context of classical physics, we would find it to be suppressed single-exponentially by a large number, $\sim \epsilon^{N_A}$ where N_A is Avogadro's number. This is because states having classically well-defined configurations are already exponentially rare in the whole Hilbert space, although they are *dynamically selected*—if we instead ask the fraction of the whole quantum states (including arbitrary superposition states) leading to ink in the corner, we would obtain a number with double-exponential suppression as in Eq. (38). In the context of the firewall, the initial device state is intrinsically quantum mechanical, so we *must* fine-tune the initial condition at the level of Eq. (38), i.e. we cannot use dynamical selection to reduce the fine-tuning. In this sense, the amount of fine-tuning needed to see the firewall phenomenon is even worse than that needed to see an entropy decreasing phenomenon in usual classical systems.

4 Discussion and Conclusions

In this paper, we have described a complete evolution—the formation and evaporation—of a black hole in the framework of quantum gravity preserving locality, given in Ref. [5]. While some of the results obtained are indeed specific to this context, some are more general applying to other theories of gravity as well, especially to the ones in which the formation and evaporation of a black hole is described as a unitary quantum mechanical process. Key ingredients to understand these results are

(i) The system must be described in a fixed reference frame. Moreover, the complete physical description of a process is obtained in the spacetime region in and on the stretched/apparent horizon *as viewed from that reference frame*. In particular, in the minimal implementation of the framework of Ref. [5], quantum states correspond to physical configurations on the past light cone of a fixed reference point p in and on the horizon.

(ii) A quantum state is in general a superposition of terms representing macroscopically different configurations/geometries. This is true despite the fact that the dynamics, represented by the time evolution operator, is local. In fact, interactions between degrees of freedom generically lead to such a superposition because of a distinct feature of quantum mechanics: rapid amplification of information to a more macroscopic level. (Measurements are a special case of this more general phenomenon.)

The element in (i) implies that the global spacetime picture of general relativity is an "illusion" in the sense that the internal and future asymptotically-flat spacetimes do not exist simultaneously. In one description based on a distant reference frame, there is only external spacetime and the process can be described by a unitary S-matrix in $\mathcal{H}_{\rm Minkowski}$, as depicted in Fig. 1. In another description based on an infalling frame, the interior spacetime does exist but there is no future asymptotic Minkowski region, so the process should be described in the whole covariant Hilbert space $\mathcal{H}_{\rm QG}$ containing both $\mathcal{H}_{\rm Minkowski}$ and $\mathcal{H}_{\rm sing}$; see Fig. 4. The global spacetime in general relativity provides a picture in which these and other *equivalent* descriptions are *all* represented at the same time, so it grossly overcounts the true quantum degrees of freedom.

A virtue of the general relativistic description, however, lies in the fact that it correctly represents observations made by a *classical* observer traveling in spacetime. In particular, the equivalence principle correctly captures the fact that the infalling classical observer does not see anything unusual at the horizon. An important point here is that the classical observer—or more generally a classical world—*emerges* through unitary evolution of a full quantum state as a basis state in which the information is amplified [7, 5]. Observations made by any observer, detector, etc. in such a world are well described by general relativity. While a phenomenon contradicting a generic prediction of general relativity could in principle occur [23], such a situation requires an exponentially fine-tuned initial condition [10], analogous to an (exponentially unlikely) entropy decreasing process in a system with large degrees of freedom.

The element in (ii) is important at least for two reasons. First, this provides an additional place in which the information about the initial state is encoded in a black hole final state, in particular in relative phases of the coefficients of the terms representing macroscopically different black hole configurations. Indeed, a late black hole state becomes a superposition of black holes with different spins and in different locations, even if the back hole is formed from collapsing matter that had a well-defined classical configuration without an angular momentum. Second, this aspect of the evolution also affects the way in which a complementarity map between interior and exterior spacetimes works for an old black hole. Because of the branching of the black hole state described above, such a map must involve a superposition of semi-classical geometries at some point in the evolution. Namely, there is no simple correspondence between the two semi-classical regions/geometries for an old black hole that is applicable throughout the entire history of the

evolution (even if we focus on time before the reference point hits the singularity).

In treating a system with gravity, it is often assumed that there is a fixed semi-classical background geometry. For example, this is the case in the original complementarity argument [4] and in most treatments of defining probabilities in the eternally inflating multiverse [28]. At first sight, this does not lead to any problem because the relevant systems are large—we usually do not need to keep track of the whole quantum nature of the state in those cases, including a superposition of possible outcomes, entanglement with the observer, and so on. These intrinsically quantum properties, however, are crucial when we discuss the *fundamental structure* of the theory such as unitarity and information. As discussed in Section 2.6, by focusing on a particular outcome—which is a completely legitimate procedure in discussing the outcome of a particular experiment—we will never be able to see the correct unitary structure of the underlying quantum theory. Committing to a specific semi-classical geometry is precisely such a treatment. An important message is that avoiding these treatments, i.e. keeping the full superposition—or *many worlds*—of the state, is a key to evade many apparent problems/paradoxes in black hole physics [10] and in eternally-inflating multiverse cosmology [6, 5]. Hopefully, the present paper adds further clarifications on this issue, and provides a useful framework for further studies of fundamental issues in black hole physics.

Acknowledgments

This work was supported in part by the Director, Office of Science, Office of High Energy and Nuclear Physics, of the US Department of Energy under Contracts DE-FG02-05ER41360 and DE-AC02-05CH11231, by the National Science Foundation under grants PHY-0855653 and DGE-1106400, and by the Simons Foundation grant 230224.

A Spontaneous Spin-up of a Schwarzschild Black Hole

Just as a black hole accumulates momentum over its lifetime through randomly recoiling from Hawking emissions, we can ask if a black hole also accumulates angular momentum due to the spin and orbital angular momentum of emitted particles. In this appendix, we argue that the answer is yes: non-rotating black holes with initial mass M_0 spontaneously spin up to angular momentum $J \equiv |\mathbf{J}| \sim M_0$ at a time of order M_0^3. This implies that a Schwarzschild black hole evolves into a superposition of Kerr black holes with different values of $\mathbf{J}$, although the resulting angular momenta will be small enough, $J/M^2 \ll 1$, that the geometry of each term is still well approximated by the Schwarzschild one.

To begin with, let us consider how many Hawking quanta are emitted by the time at which an

initial black hole loses some fixed fraction of its mass, e.g. the Page time at which the black hole mass becomes $M = M_0/\sqrt{2}$. The number of emitted quanta is

$$N \sim \frac{M_0}{T_H} \sim M_0^2, \tag{39}$$

where $T_H \sim 1/M_0$ is the Hawking temperature. If the emitted quanta consist of a particle with spin $s > 0$, then each emission changes the angular momentum of the black hole by $\Delta J \sim s$, depending on the direction of the spin. Assuming that the emission is unbiased in the direction of angular momentum (see below), we find that the black hole accumulates the angular momentum

$$J \sim s\sqrt{N} \sim M_0, \tag{40}$$

at a time of order M_0^3, where we have taken $s \sim O(1)$ in the last expression.

If the Hawking quanta consist of a scalar ($s = 0$), then most of the emissions do not affect the black hole angular momentum since the emissions are dominated by s-wave. However, there is a small probability that a quantum is emitted in a higher angular momentum mode. The probability is dominated by p-wave ($l = 1$), which can be calculated for small J/M^2 as $p \simeq 0.002 + O(J/M^2)$, independent of M [29, 30]. Therefore, the number of Hawking quanta that affect the black hole angular momentum is pN, and the accumulated angular momentum of the black hole is

$$J \sim \sqrt{pN} \sim M_0, \tag{41}$$

which is parametrically the same as in the case of a particle with spin.

One might think that once the accumulated angular momentum becomes macroscopic, $J \gg 1$, the black hole becomes a Kerr black hole, so that there is a bias in the Hawking spectrum that preferentially selects emissions that reduce J [1], preventing a further accumulation of J. We now argue, however, that until the time $t \sim M_0^3$ when the mass of the black hole starts decreasing significantly, the evolution of $\mathbf{J}$ is well approximated by a random walk process as described above.

To see this, at a given time t, let us call the direction of $\mathbf{J}$ the z-axis. Suppose an emission of a particle with spin s changes $J = J_z$, which occurs with $O(p)$ and $O(1)$ probabilities for $s = 0$ and $s > 0$, respectively. For small J/M^2, the probability ρ_+ (ρ_-) that the emission increases (reduces) J is [30]:

$$\rho_\pm = \frac{1}{2} \mp c\frac{J}{M^2}, \tag{42}$$

where J and M are the magnitude of angular momentum and the mass before the emission takes place, and c is an $O(1)$ coefficient which depends on the type of a particle emitted and is independent of J and M to first order in J/M^2. Numerical simulations of this process indicate that this bias is not strong enough to prevent a black hole from spinning up to $J \sim M_0$ by the Page

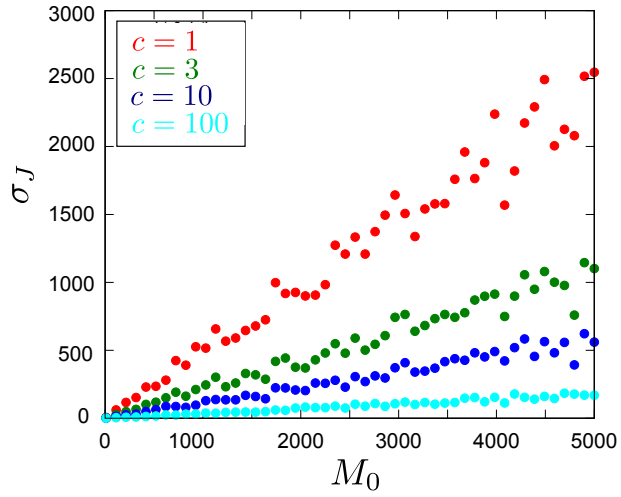


Figure 7: Plot of σ_J, the square root of the variance of J at the Page time, as a function of the initial mass M_0. Each data point represents σ_J obtained by 100 simulations. The different colors correspond to different values of coefficient c in Eq. (42), which measures the strength of the angular momentum emission bias from a Kerr black.

time, $t_{\rm Page}$. Results of these simulations are shown in Fig. 7, where we have assumed a change of J according to Eq. (42) in each time interval M_0. The results indicate that

$$J \sim f(c) M_0 \sim M_0 \tag{43}$$

at $t \sim t_{\rm Page}$, where f is a monotonically decreasing function of c; in fact, our simulations suggest that $f(c) \propto 1/\sqrt{c}$ for $c \gtrsim 1$.

The results obtained above can be understood by the following simple argument. Imagine that at some late time $t \gtrsim M_0^3$, the probability distribution of the black hole angular momentum reaches some "equilibrium" distribution $P(J)$, in which the random walk effect increasing J is balanced with the bias of the emission reducing J. According to Eq. (42), this implies

$$\rho_+ P(J) = \rho_- P(J+1), \tag{44}$$

leading to

$$\frac{P(J+1)}{P(J)} = \frac{1 - 2c\frac{J}{M^2}}{1 + 2c\frac{J+1}{M^2}} \approx 1 - 4c\frac{J}{M^2}. \tag{45}$$

Here, we have used $1 \ll J \ll M^2$ in the last expression. This has the solution

$$P(J) \sim e^{-2c\frac{J^2}{M^2}}. \tag{46}$$

Namely, the black hole angular momentum has a characteristic size

$$J \sim \frac{1}{\sqrt{c}} M, \tag{47}$$

consistent with the result obtained in Eq. (43).

In summary, we conclude that a Schwarzschild black hole with initial mass M_0 will spontaneously spin up to $J \sim M_0$ by a timescale of order M_0^3. When the black hole mass starts decreasing significantly, its angular momentum will also start decreasing, following Eq. (47). The combination J/M^2 keeps increasing as $1/M$ but is still (much) smaller than 1, as long as $M \gg 1$ where our analysis is valid. What happens at the real end of the evaporation is unclear, but we can say that while the evolution of a Schwarzschild black hole leads to a superposition of Kerr black holes with distinct angular momenta, the probability of it becoming a macroscopic extremal black hole ($J = M^2 \gg 1$) is, most likely, exponentially suppressed.

References

[1] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975) [Erratum-ibid. **46**, 206 (1976)].

[2] G. 't Hooft, arXiv:gr-qc/9310026; L. Susskind, J. Math. Phys. **36**, 6377 (1995) [arXiv:hep-th/9409089].

[3] R. Bousso, JHEP **07**, 004 (1999) [arXiv:hep-th/9905177].

[4] L. Susskind, L. Thorlacius and J. Uglum, Phys. Rev. D **48**, 3743 (1993) [arXiv:hep-th/9306069]; C. R. Stephens, G. 't Hooft and B. F. Whiting, Class. Quant. Grav. **11**, 621 (1994) [arXiv:gr-qc/9310006].

[5] Y. Nomura, arXiv:1110.4630 [hep-th].

[6] Y. Nomura, JHEP **11**, 063 (2011) [arXiv:1104.2324 [hep-th]].

[7] H. Ollivier, D. Poulin and W. H. Zurek, Phys. Rev. Lett. **93**, 220401 (2004) [arXiv:quant-ph/0307229]; R. Blume-Kohout and W. H. Zurek, Phys. Rev. A **73**, 062310 (2006) [arXiv:quant-ph/0505031].

[8] D. N. Page, Phys. Rev. Lett. **44**, 301 (1980).

[9] D. N. Page, Phys. Rev. Lett. **71**, 3743 (1993) [hep-th/9306083].

[10] Y. Nomura, J. Varela and S. J. Weinberg, arXiv:1207.6626 [hep-th].

[11] Y. Nomura, Phys. Rev. D **86**, 083505 (2012) [arXiv:1205.5550 [hep-th]].

[12] S. W. Hawking, Phys. Rev. D **14**, 2460 (1976).

[13] R. M. Wald, Phys. Rev. D **21**, 2742 (1980).

[14] Y. Aharonov, A. Casher and S. Nussinov, Phys. Lett. B **191**, 51 (1987); T. Banks, A. Dabholkar, M. R. Douglas and M. O'Loughlin, Phys. Rev. D **45**, 3607 (1992) [hep-th/9201061].

[15] S. B. Giddings, Phys. Rev. D **46**, 1347 (1992) [hep-th/9203059].

[16] F. Dyson, Institute for Advanced Study report (1976), unpublished.

[17] J. Preskill, in *Blackholes, Membranes, Wormholes and Superstrings*, ed. S. Kalara and D. V. Nanopoulos (World Scientific, Singapore, 1993) p. 22 [hep-th/9209058]; S. B. Giddings, in *Particles, Strings and Cosmology*, ed. J. Bagger *et al.* (World Scientific, Singapore, 1996) p. 415 [hep-th/9508151].

[18] S. D. Mathur, Class. Quant. Grav. **26**, 224001 (2009) [arXiv:0909.1038 [hep-th]].

[19] See, e.g., G. 't Hooft, Nucl. Phys. B **335**, 138 (1990), and references therein.

[20] O. Lunin and S. D. Mathur, Nucl. Phys. B **623**, 342 (2002) [hep-th/0109154].

[21] S. B. Giddings, Class. Quant. Grav. **28**, 025002 (2011) [arXiv:0911.3395 [hep-th]]; Phys. Rev. D **85**, 124063 (2012) [arXiv:1201.1037 [hep-th]].

[22] R. Brustein, arXiv:1209.2686 [hep-th].

[23] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, arXiv:1207.3123 [hep-th].

[24] J. M. Maldacena, Adv. Theor. Math. Phys. **2**, 231 (1998) [Int. J. Theor. Phys. **38**, 1113 (1999)] [arXiv:hep-th/9711200].

[25] See, e.g., D. N. Page, hep-th/9305040, and references therein.

[26] W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).

[27] Y. Nomura and J. Varela, arXiv:1211.7033 [hep-th].

[28] For reviews, see e.g. A. H. Guth, Phys. Rept. **333**, 555 (2000) [arXiv:astro-ph/0002156]; A. Vilenkin, J. Phys. A **40**, 6777 (2007) [arXiv:hep-th/0609193]; S. Winitzki, Lect. Notes Phys. **738**, 157 (2008) [arXiv:gr-qc/0612164]; A. Linde, Lect. Notes Phys. **738**, 1 (2008) [arXiv:0705.0164 [hep-th]].

[29] A. A. Starobinskiĭ and S. N. Churilov, Zh. Eksp. Teor. Fiz. **65**, 3 (1973).

[30] D. N. Page, Phys. Rev. D **13**, 198 (1976).