

RECEIVED: June 19, 2014

REVISED: September 17, 2014

ACCEPTED: October 13, 2014

PUBLISHED: October 31, 2014

Black holes, entropies, and semiclassical spacetime in quantum gravity

Yasunori Nomura and Sean J. Weinberg

Berkeley Center for Theoretical Physics, Department of Physics, University of California,
Berkeley, CA 94720, U.S.A.

Theoretical Physics Group, Lawrence Berkeley National Laboratory,
Berkeley, CA 94720, U.S.A.

E-mail: ynomura@berkeley.edu, sjweinberg@berkeley.edu

ABSTRACT: We present a coherent picture of the quantum mechanics of black holes. The picture does not require the introduction of any drastically new physical effect beyond what is already known; it arises mostly from synthesizing and (re)interpreting existing results in appropriate manners. We identify the Bekenstein-Hawking entropy as the entropy associated with coarse-graining performed to obtain semiclassical field theory from a fundamental microscopic theory of quantum gravity. This clarifies the issues around the unitary evolution, the existence of the interior spacetime, and the thermodynamic nature in black hole physics — any result in semiclassical field theory is a statement about the maximally mixed ensemble of microscopic quantum states consistent with the specified background, within the precision allowed by quantum mechanics. We present a detailed analysis of information transfer in Hawking emission and black hole mining processes, clarifying what aspects of the underlying dynamics are (not) visible in semiclassical field theory. We also discuss relations between the black hole entropy and the entanglement entropy across the horizon. We then extend our discussions to more general contexts in quantum gravity. The subjects include extensions to de Sitter and Minkowski spaces and implications for complementarity and cosmology, especially the eternally inflating multiverse.

KEYWORDS: Black Holes in String Theory, Black Holes, Models of Quantum Gravity

ARXIV EPRINT: [1406.1505](https://arxiv.org/abs/1406.1505)

Contents

1	Introduction	1
2	How to interpret the black hole entropy?	3
2.1	Bekenstein-Hawking entropy from spatial localization	4
2.2	Where does the black hole entropy reside?	6
3	What is the semiclassical approximation?	9
3.1	Thermal nature	10
3.2	Interior spacetime	12
3.3	UV/IR correspondence and relation to the entanglement entropy	16
4	Nature of microscopic dynamics	18
4.1	Response of a static detector	19
4.2	Effect from acceleration	20
4.3	Effect from spacetime curvature	23
4.4	Horizon of an evaporating black hole	25
5	Discussions	26

1 Introduction

One of the major cornerstones in the pursuit of a quantum theory of gravity was the discovery of the finite nonzero entropy of a black hole [1] and the associated thermal radiation from it [2]. The precise interpretation of this result, however, has not been entirely clear. What does this entropy represent? Where does it reside? The fact that a black hole emits radiation at the quantum level allowed us to contemplate the possibility that a complete description of the system is obtained by referring only to the spacetime region outside the horizon; in particular, the evolution of a black hole can be unitary when described from a distance [3, 4]. The existence of the interior spacetime then becomes manifest only after changing the description to an infalling one, which is supposed to provide a description complementary to the distant one [5–7]. This last picture, however, has recently been challenged [8]: if the emission process is indeed unitary, then an infalling observer is claimed to encounter something drastic at the horizon, so that there is no such thing as the interior spacetime, at least for an old black hole.

The main purpose of this paper is to present a coherent picture of the quantum mechanics of black holes and to clarify the issues around their unitary evolution and the existence of the interior spacetime, although we also extend our discussions to more general contexts in quantum gravity. We adopt the hypothesis that, as postulated in refs. [3, 4], there is a unitary description of a black hole which involves only the region outside the horizon.

We discuss how this hypothesis is consistent with the existence of the interior spacetime, in the sense of complementarity in refs. [5, 6]. Our picture does not involve any drastically new physical effect beyond what is already known or postulated, and yet it requires a certain revision of the applicability of semiclassical theory in discussing physics around the black hole. We feel that this provides a significant advancement toward understanding the structure of quantum gravity and its relation to the emergent semiclassical picture of spacetime.

Our discussion begins with studying the origin of the Bekenstein-Hawking entropy. In general, the concept of entropy is associated with coarse-graining. What is the coarse-graining responsible for the Bekenstein-Hawking entropy? Since the Bekenstein-Hawking entropy arises in semiclassical field theory without doing any coarse-graining in the theory, it must be associated with the coarse-graining performed *to obtain the theory* from a fundamental theory of quantum gravity. This picture is indeed consonant with the fact that in quantum mechanics, having a well-defined geometry of spacetime, e.g. a black hole formed in a well-defined spacetime location, requires taking a superposition of an enormous number of energy-momentum eigenstates, so we expect that there are many different ways to arrive at the same background for a semiclassical field theory within the precision allowed by quantum mechanics. In particular, this implies that detailed microscopic physics occurring within the fine-grained degrees of freedom, including the flow of microscopic information in the Hawking emission process, cannot be seen in semiclassical field theory. This is the origin of the apparent violation of unitarity [9] in the Hawking evaporation process.

The fact that the microstates of a black hole cannot be resolved in semiclassical field theory implies that for the purpose of describing physics of excitations in semiclassical field theory — including an ordinary observer falling into the black hole — we may ignore the dynamics within the fine-grained microscopic degrees of freedom, which is responsible for the fundamental unitarity of the Hawking process. Specifically, we may view that a semiclassical field theory vacuum has a hidden “vacuum index” k labeling the microstates, and can construct field theory operators that act in the same way on each of these exponentially many vacuum states: $k = 1, \dots, e^S$, where S represents the Bekenstein-Hawking entropy. This allows us to construct operators describing the interior spacetime following the standard method in semiclassical field theory [10]. Note that this does not contradict the hypothesis that the evaporation of a black hole *when described in the fundamental theory* is a unitary process.

Since emissions of later Hawking quanta occur within the causal future of the region in which earlier Hawking emissions have occurred, there is no reason to expect that the information recovery from a black hole must violate locality, or causality, in spacetime as described in a distant reference frame. We perform a detailed analysis of the Hawking emission process, as well as black hole mining, and argue that the modes of a black hole relevant for the information transfer are not visible to semiclassical field theory — they are “too soft” to be resolved. In semiclassical field theory, Hawking emission occurs because of an intrinsic quantum mechanical ambiguity of defining particles in curved spacetime, and the information about the microstate of a black hole it carries away is viewed as being stored nonlocally in states representing the field theory vacuum, which extends into the

whole zone region. Note that this does not contradict the locality *of the dynamics*; in quantum mechanics, the information about a state can in general be delocalized even if the dynamics is local. This also implies that the purifiers of the emitted Hawking quanta are what semiclassical gravity describes as vacuum states; in particular, they are not outgoing field theory modes as envisioned in refs. [8, 11, 12].

The picture described in this paper has been developed over the past years in refs. [13–15], although its final form presented here does not agree in all respects with those given in these earlier works. Following the inspirational work in ref. [8], there appeared numerous attempts to reconcile the unitary evolution of a black hole with the existence of the interior, some of which introduce dramatic new physics; see, e.g., refs. [16–48]. To give examples, Giddings [16, 17] studied the possibility that the dynamics in the exterior region may be nonlocal. Papadodimas and Raju [18–20] and Verlinde and Verlinde [21–23] considered “state-dependent” maps to describe the interior region. Maldacena and Susskind [24] considered the possibility that a part of the degrees of freedom describing the interior region comes from Hawking radiation emitted earlier. Our picture does not require the introduction of such new physics.

The organization of this paper is as follows. In section 2, we discuss how to interpret the Bekenstein-Hawking entropy, in particular the way it may manifest when the black hole interacts with other systems. In section 3, we discuss how the microscopic structure of quantum gravity is related to the semiclassical view of the world. We argue that semiclassical field theory corresponds to the description arising after coarse-graining the degrees of freedom associated with the Bekenstein-Hawking entropy, specifically after taking the maximally mixed ensemble of microstates consistent with the given background within some precision. This explains various confusing aspects of black hole physics regarding fundamental unitarity and the thermodynamic nature of semiclassical gravity. We also discuss relations between the black hole entropy and the entanglement entropy across the horizon. In section 4, we present a detailed analysis of the information transfer in the Hawking emission and black hole mining processes, especially focusing on what aspects of the underlying dynamics can be captured in semiclassical field theory. Finally, in section 5, we extend our discussions to more general contexts in quantum gravity, including discussions of de Sitter and Minkowski spaces and implications for complementarity and cosmology, especially the eternally inflating quantum multiverse.

Throughout the paper, we adopt the Schrödinger picture for quantum evolution, and adopt natural units in which $\hbar = c = 1$, unless otherwise stated. $l_P \simeq 1.62 \times 10^{-35}$ m denotes the Planck length. In our discussion of black hole physics, we focus on a black hole that is well approximated by a Schwarzschild black hole in 4-dimensional asymptotically flat spacetime. We do not expect difficulty in extending it to other more general cases.

2 How to interpret the black hole entropy?

In this section, we discuss an interpretation of the black hole entropy, including how it manifests when the black hole interacts with other systems. Understanding this issue correctly is important in the development of our picture in later sections.

2.1 Bekenstein-Hawking entropy from spatial localization

We first remind the readers that the Bekenstein-Hawking entropy is an *entropy density*. It is a finite function of the black hole mass M because we restrict our considerations to a black hole in a specific spacetime region.

Let us begin our discussion by considering a simple quantum system consisting of massless particles that do not have a conserved charge. The quantum states representing a single particle can be labeled by three continuous numbers corresponding to the momentum of the particle $\mathbf{k} = (k_x, k_y, k_z)$, so that the states with N (weakly interacting) particles are labeled by $3N$ continuous numbers. Consider now a set of quantum states $|\psi_i\rangle$ in which all the particles in the system are confined in a spatial region with volume V at some time $t = t_0$. These states are obtained (only) by taking appropriate superpositions of states in which particles have well-defined momenta. How many such states exist? Assuming that the system has energy between E and $E + \delta E$, quantum mechanics tells us that the number of independent $|\psi_i\rangle$ states is finite and of order $\mathcal{N} \approx e^{O(\rho^{3/4}V)}\delta E/E$, where $\rho \equiv E/V$. Namely, the entropy density is $S = \ln(\mathcal{N}E/\delta E) \approx O(\rho^{3/4}V)$, which does not depend on how δE is chosen as long as $\delta E/E \gg e^{-S}$. Note that physics described here is intrinsically quantum mechanical in the sense that classical mechanics would allow the specified spatial region to support a continuously infinite number of states with energy between E and $E + \delta E$. This can be seen from the fact that S can be written as $O(\rho^{3/4}V/\hbar^{3/4}c^{3/4})$ when \hbar and c are restored, so that $S \rightarrow \infty$ for $\hbar \rightarrow 0$.

The finiteness of the Bekenstein-Hawking entropy suggests that the situation for a black hole is similar. Let us consider forming a black hole of mass M by collapsing matter. In order to have such a process in which the black hole is formed at a well-defined spacetime location, the initial matter state must involve a superposition of energy-momentum eigenstates. Suppose we want to identify the spacetime location of the black hole with precision comparable to the quantum stretching of the horizon $\Delta r \approx O(1/M)$, i.e. $\Delta d \approx O(l_P)$, and the timescale of Hawking emission $\Delta t \approx O(Ml_P^2)$, where r and t are the Schwarzschild radial and time coordinates, respectively, and d is the proper length. In this case the superposition must involve momenta with spread $\Delta p \gtrsim O(1/Ml_P^2)$ and energy with $\Delta E \gtrsim O(1/Ml_P^2)$, where Δp and ΔE are both measured in the asymptotic region, and these uncertainties must exist shortly before the formation of the black hole in the branch we are interested in. For an older black hole, the stochastic nature of Hawking radiation introduces macroscopic uncertainties for the black hole mass, location, and spin [49, 50]. This effect, however, can be easily separated by focusing on appropriate branches in the full quantum state. A consideration similar to the one above then applies in the limit that the induced spin of the black hole is neglected, which is a reasonable approximation for a large black hole.

How many black hole states do there exist in which the black hole is at a specific location within an uncertainty of $\Delta r \approx O(1/M)$ and $\Delta t \approx O(Ml_P^2)$ measured in the asymptotic region? Because of the required uncertainty in energy-momentum, we must consider the number of black hole states of mass (which we may identify with the energy) in the range between M and $M + \delta M$ where $\delta M \gtrsim O(1/Ml_P^2)$. As in the case before, quantum mechanics makes the number of independent such localized states finite. Indeed,

the logarithm of this number — the entropy density — is given by the Bekenstein-Hawking formula

$$S = \frac{\mathcal{A}}{4l_P^2}, \quad (2.1)$$

at the leading order in expansion in inverse powers of \mathcal{A}/l_P^2 , where $\mathcal{A} = 16\pi M^2 l_P^4$ is the area of the horizon. Here we identify the expression of eq. (2.1) to represent the density of black hole *vacuum* states, i.e. states representing a black hole that does not have an excitation in the interior or the exterior region. This is reasonable because the Bekenstein-Hawking considerations apply to a black hole after the horizon is classically stabilized. Possible higher order corrections to eq. (2.1) do not affect our argument, so we will ignore them. The states in which there are excitations in the interior or the near exterior region will be discussed below.

As can be easily seen, the expression in eq. (2.1) is insensitive to the precise choice of δM as long as $\delta M/M \gg e^{-S}$. For definiteness, we will mostly take the smallest possible value of δM :

$$\delta M \approx O\left(\frac{1}{Ml_P^2}\right). \quad (2.2)$$

The corresponding uncertainty in timescale is then $\Delta t \approx 1/\delta M \approx O(Ml_P^2)$. Only an $O(1)$ number of Hawking quanta with energy $E \approx O(1/Ml_P^2)$ are emitted within Δt , so a black hole does not change its mass more than $O(\delta M)$ in this timescale.

The discussion above demonstrates that the finite entropy of a black hole can be understood in a similar manner to that of other quantum systems confined in a finite spatial region, although there are important differences including the fact that the black hole entropy scales as the area unlike those of usual (much) lower density materials which scale as the volume V for fixed intensive quantities, e.g., ρ .¹ (This leads to the conjecture for the universal entropy bound associated with the area of a codimension-2 surface [51, 52]; a precise formulation of it must involve null hypersurfaces rather than spatial regions [53, 54].) In particular, it is not appropriate to consider that quantum mechanics introduces an exponentially large number of degeneracies for the microstates that do not exist in the corresponding classical black hole. In classical general relativity, a set of Schwarzschild black holes located at some place at rest are parameterized by a continuous mass parameter M ; i.e., there are a continuously infinite number of black hole states in the energy interval between M and $M + \delta M$ for any M and small δM . Quantum mechanics *reduces* this to a finite number $\approx e^S \delta M/M$ with S given by eq. (2.1).² This can also be seen from the fact that S is written as $\mathcal{A}c^3/4l_P^2\hbar$ when \hbar and c are restored, which becomes infinite for $\hbar \rightarrow 0$.

We now discuss black hole states having excitations in the interior or near exterior region. (In a distant reference frame, an excitation in the interior region is described as an excitation of the horizon.) Consider the set C_I of the states which have specified excitations

¹Following standard convention, here and below we use the term entropy to also mean entropy density.

²Of course, quantum mechanics allows for a superposition of these finite number of independent states, so the number of possible (not necessarily independent) states is continuously infinite. The statement here applies to the number of independent states, regarding classical black holes with different M as independent states.

I in a background of a black hole of mass M (more precisely between M and $M + \delta M$). As suggested, e.g., by a representative estimate in ref. [51], the entropy associated with all these sets is expected to be written roughly as

$$S = \frac{\mathcal{A}}{4l_P^2} + \sum_I S_I; \quad \sum_I |S_I| \approx O\left(\frac{\mathcal{A}^q}{l_P^{2q}}; q < 1\right), \quad (2.3)$$

where $\mathcal{A} = 16\pi M^2 l_P^4$, and S_I is the entropy associated with the class of excitations I . The total entropy S , therefore, is simply $\mathcal{A}/4l_P^2$ up to fractional corrections in inverse powers of \mathcal{A}/l_P^2 . Namely, the existence of excitations provides only small perturbation in terms of the entropy counting. (This point was particularly emphasized in refs. [15, 55].) Note that being defined as fluctuations with respect to a fixed background spacetime, the energy E_I and entropy S_I of excitations can be either positive or negative.

The approximate nature of eq. (2.3) needs to be emphasized. First, dividing a physical configuration into excitations and background is artificial. In the expression in eq. (2.3), we have simply added the entropies of the excitations to that of the black hole. This is appropriate since we can view semiclassical physics occurring in a fixed black hole background to correspond to excitations on *each* of the black hole vacuum states (see section 3.2). On the other hand, it is clear that this treatment involves an approximation — because the existence of excitations must affect geometry, the “on-shell” Hilbert space (the space of states after the equations of motion are imposed) must take the form more complicated than the one implied by eq. (2.3). Nevertheless, eq. (2.3) represents the way semiclassical field theory treats physical systems, which, with some care, may capture certain aspects of quantum gravitational physics.

2.2 Where does the black hole entropy reside?

Consider the set of all the independent black hole vacuum states $|\Psi_k(M)\rangle$ of mass between M and $M + \delta M$, with the black hole localized in a specified spatial region with precision $\Delta r \approx O(1/M)$. As discussed in the previous subsection, the index k runs over

$$k = 1, \dots, e^S = e^{4\pi M^2 l_P^2} \equiv n(M), \quad (2.4)$$

representing the Bekenstein-Hawking entropy. Where does this entropy reside?

Let us assume that physics is local in the region outside the stretched horizon, in particular we can use appropriately quantized Einstein gravity to calculate physical quantities within its regime of validity. It is important to remind ourselves that locality is a property of dynamics, in particular the Hamiltonian, and not that of states. In fact, in quantum mechanics the information about a state is quite generally delocalized in space. Consider, for instance, states $|1\rangle$ and $|2\rangle$ representing, respectively, a particle of species 1 and 2 in a momentum eigenstate, located in a finite box with periodic boundary conditions. Suppose the system is in one of these two states. Where does the information about the state exist? The answer is: everywhere. We may find out if the system is in $|1\rangle$ or $|2\rangle$ anywhere in the

box, as long as we are equipped with an appropriate detector and have enough time for the measurement.³

Quite analogously, we assume that a part of the information about the index k of a black hole vacuum state is delocalized over a large spatial region. In particular, we consider that this occurs mostly in the region $r \lesssim 3Ml_P^2$, which, in a distant view, consists of the stretched horizon at $r = r_s = 2Ml_P^2 + O(1/M)$ and near exterior “zone” $r_s < r \lesssim 3Ml_P^2$, the region inside the effective gravitational potential barrier separating the near and far exterior regions. To isolate the part of the state that carries a meaningful amount of information, let us separate the Hilbert space into two factors: one representing states in the region $r \leq R$, $\mathcal{H}_<$, and the other $r > R$, $\mathcal{H}_>$. Here, the separation radius R may be conveniently chosen to be somewhere in the outer side of the potential barrier, e.g. in the range $(3\text{--}5)Ml_P^2$. The black hole vacuum state $|\Psi_k(M)\rangle$ can then be written approximately as

$$|\Psi_k(M)\rangle = \sum_a c_a |\psi_{ka}(M)\rangle |\phi_a(M)\rangle, \quad (2.5)$$

where $|\psi_{ka}(M)\rangle$ and $|\phi_a(M)\rangle$ are elements of $\mathcal{H}_<$ and $\mathcal{H}_>$, respectively. The statement that the information is spread essentially only in the region $r_s \leq r \lesssim 3Ml_P^2$ corresponds to the absence of the index k for c_a and $|\phi_a(M)\rangle$.

We note that by the information delocalization above, we do not mean that the information is spread *uniformly* over the region $r_s \leq r \lesssim 3Ml_P^2$. In fact, we expect that most of the information is localized in the region close to the horizon. Specifically, when we split the Hilbert space at $r = R' < R$, we have the expression

$$|\Psi_k(M)\rangle = \sum_a c'_a |\psi'_{ia}(M)\rangle |\phi'_{ja}(M)\rangle, \quad (2.6)$$

where $k = \{i, j\}$, and i and j take values $i = 1, \dots, I$ and $j = 1, \dots, J$ with $IJ = n(M)$. We expect that $s(R') \equiv \ln J$, which is a function of R' , is given by the thermal entropy contained in the region $r > R'$, calculated in semiclassical field theory using the blueshifted local Hawking temperature. In particular,

$$s(R') \ll \ln n(M), \quad (2.7)$$

unless $R' - 2Ml_P^2 \lesssim O(1/M)$. Note that taking $J = 1$, which corresponds to $R' \rightarrow R$, eq. (2.6) is reduced to eq. (2.5). An important point here is that *some* information about k can be extracted by a physical process occurring as far as $r \simeq 3Ml_P^2$ in timescale of order $1/\delta M$ (more precisely, without directly interacting with the stretched horizon at the field theory level; see section 4 for further discussions).

³This does not mean that there is no physical information flow associated with the measurement process. It is simply that to discuss the information flow, we need to consider the process preparing the state. For example, if the particle is created at some point (e.g. in the form of a wavepacket, which later broadens), the information must be regarded as being transferred from the creation point to the detector location in the branch in which the detector has responded. If the dynamics is local, this information transfer does not violate causality in spacetime.

From now on, we suppress the trivial entanglement in eq. (2.5) that does not depend on k for the simplicity of the notation. This allows us to write the black hole vacuum state $|\Psi_k(M)\rangle$ (very roughly) as

$$|\Psi_k(M)\rangle \approx |\psi_k(M)\rangle |0_{\text{ext}}\rangle, \quad (2.8)$$

where $|0_{\text{ext}}\rangle$ is the vacuum state in the region $r > R$, which does not depend on k . While $|0_{\text{ext}}\rangle$ has some dependence on M , for our purposes here we will ignore it. We have also assumed the absence of Hawking quanta emitted at earlier times. If there are excitations in the interior region (or of the stretched horizon in a distant view) or in the zone, then $|\psi_k(M)\rangle$ in eq. (2.8) must be replaced with the corresponding excited states $|\tilde{\psi}_{kn}(M)\rangle$, where n labels the excitations. The dimension of the relevant Hilbert spaces satisfies

$$\ln \dim \mathcal{H}_{\tilde{\psi}_k} \approx O(M^{2q} l_P^{2q}; q < 1), \quad (2.9)$$

where $\mathcal{H}_{\tilde{\psi}_k}$ is the Hilbert space spanned by $|\psi_k(M)\rangle$ and $|\tilde{\psi}_{kn}(M)\rangle$ for a fixed k , and all the $\mathcal{H}_{\tilde{\psi}_k}$'s with different k (as well as the structure of field theory operators defined for each of them) are isomorphic with each other; see section 3.2. If the far exterior region is not in the vacuum, the state $|0_{\text{ext}}\rangle$ in eq. (2.8) must be replaced accordingly.

How can we probe the microscopic information associated with a black hole vacuum state? Consider a physical detector located somewhere in the region $r_s < r \lesssim 3Ml_P^2$ whose ground state $|d_0\rangle$ represents the “ready” state while excited states $|d_i\rangle$ ($i = 1, 2, \dots$) are the pointer states. Suppose the proper energies needed to excite $|d_0\rangle$ to $|d_i\rangle$ are given by $E_{d,i}$. The state representing the region $r \leq R$ then evolves as

$$|\psi_k(M)\rangle |d_0\rangle \rightarrow \sum_i \sum_{k_i=1}^{n(M_i)} \alpha_{k_i i}^k |\psi_{k_i}(M_i)\rangle |d_i\rangle; \quad M_i = M - E_{d,i} \sqrt{1 - \frac{2Ml_P^2}{r_d}}, \quad (2.10)$$

where the function $n(M)$ is defined in eq. (2.4). The coefficients $\alpha_{k_i i}^k$ in general depend on the location of the detector, and M_i for different i may belong to the same mass within the precision δM , i.e. $M_i = M_{i'}$ for $i \neq i'$. Here, we have separated the detector state from the rest of the system, although in a complete treatment the detector itself may be better viewed as an excitation over $|\psi_k(M)\rangle$.

The process in eq. (2.10) implies that (a part of) the information encoded in the index k can be probed by the detector. In fact, the detailed microscopic process leading to eq. (2.10) is somewhat more subtle. In particular, in eq. (2.10) we have assumed that the black hole states appearing in the right-hand side are vacuum states, but this is the case only after multiple elementary processes have occurred. We will discuss these processes in section 4.

Emission of Hawking quanta to the asymptotic region also carries information away from the black hole. Since the effective gravitational potential is damped in the region $r \gtrsim R$, Hawking quanta emitted from the region $r \sim R$ propagate essentially freely to the asymptotic region (except that they receive a small residual gravitational redshift of a factor of about 1.5). The elementary emission process may thus be written as

$$|\psi_k(M)\rangle |0_{\text{ext}}\rangle \rightarrow \sum_i \sum_{k_i=1}^{n(M-E_i)} \beta_{k_i i}^k |\psi_{k_i}(M-E_i)\rangle |i_{\text{ext}}\rangle, \quad (2.11)$$

where $|i_{\text{ext}}\rangle$ represents a state with energy E_i in which outgoing radiation modes are excited. Again, we have put black hole vacuum states after the evolution in the right-hand side; the eligibility of this will be discussed in section 4. Since the final state depends on k , this evolution can be unitary, which we assume to be the case.

We stress that the emission process in eq. (2.11) can be viewed as occurring locally in the potential barrier region because of the information delocalization discussed above. To elucidate this point, we consider the tortoise coordinate

$$r^* = r + 2Ml_P^2 \ln \frac{r - 2Ml_P^2}{2Ml_P^2}, \quad (2.12)$$

in which the region outside the Schwarzschild horizon $r \in (2Ml_P^2, \infty)$ is mapped into $r^* \in (-\infty, \infty)$. This coordinate is useful in that the kinetic term of an appropriately redefined field takes the canonical form, so that its propagation can be analyzed as in flat space. In this coordinate, the stretched horizon, located at $r = 2Ml_P^2 + O(l_s^2/Ml_P^2)$, is at

$$r_s^* \simeq -4Ml_P^2 \ln \frac{Ml_P^2}{l_s}, \quad (2.13)$$

where l_s is the string (cutoff) scale, which we take to be within a couple of orders of magnitude of l_P . This implies that there is a large distance between the stretched horizon and the potential barrier region when measured in r^* : $\Delta r^* \approx 4Ml_P^2 \ln(Ml_P) \gg O(Ml_P^2)$ for $\ln(Ml_P) \gg 1$. On the other hand, a localized Hawking quantum is represented by a wavepacket with width of $O(Ml_P^2)$ in r^* , since it has an energy of order $T_H = 1/8\pi Ml_P^2$ defined in the asymptotic region.

The point is that, given the black hole state $|\psi_k(M)\rangle$, the process in eq. (2.11) occurs in the region $|r^*| \approx O(Ml_P^2)$ without involving deep interior of the zone $-r^* \gg Ml_P^2$. In this region, the information stored (nonlocally) in the vacuum state is converted into that of a particle state outside the zone, where the concept of particles with frequencies $\omega \lesssim 1/Ml_P^2$ is well defined. A corollary of this statement is that if we evolve the system backward in time, an originally outgoing Hawking quantum does not become a highly blueshifted ingoing quantum in a region deep in the zone. Instead, it becomes “vacuum degrees of freedom”, k , due to interactions with spacetime caused by the curvature. Since the concept of particles with $\omega \lesssim 1/Ml_P^2$ is not well defined in the zone region, we need not view this as new physics contradicting conventional low energy Einstein gravity. It is simply that semiclassical field theory cannot be used to analyze these quanta within the zone, where the scale of curvature length is of order Ml_P^2 .

3 What is the semiclassical approximation?

In this section, we discuss relations between the picture presented in the previous section and the standard semiclassical treatment of a black hole. This will clarify the meaning of the Bekenstein-Hawking entropy further, and shed important light on the nature of the semiclassical approximation in quantum gravity.

It is often stated that in the semiclassical approximation, in particular quantum field theory on a fixed spacetime background, particle excitations are treated quantum mechanically while the background spacetime classically. The interpretation of this statement, however, needs care. As was seen in the previous section, if a black hole is treated “fully classically”, its entropy is infinite, and so it has zero temperature (and hence does not radiate). On the other hand, Hawking’s calculation finds a nonzero temperature in semiclassical gravity [2]. This implies that, as in any statistical mechanical understanding of entropy (which requires, e.g., quantization of a phase space), semiclassical gravity is capturing certain, though not all, quantum aspects of the relevant physical system, in this case the background spacetime.

So, what is the semiclassical approximation really? We assert that any result in the semiclassical approximation is a statement about the (maximally mixed) *ensemble* of microscopic quantum states consistent with the specified spacetime, within the precision allowed by quantum mechanics. In the case of a black hole at a fixed location, the semiclassical approximation deals with the mixed state

$$\rho(M) = \frac{1}{n(M)} \sum_{k=1}^{n(M)} |\Psi_k(M)\rangle\langle\Psi_k(M)|. \quad (3.1)$$

Because the relevant density matrix ρ in general takes the form of a maximally mixed state, results in semiclassical gravity do not depend on the basis of the microscopic pure states chosen, in the black hole case of the $|\Psi_k(M)\rangle$ ’s. Below we will elucidate the above assertion by considering the thermal nature (in section 3.1) and the interior spacetime (in section 3.2) of a semiclassical black hole.

By construction, the semiclassical approximation cannot capture unitarity of detailed microscopic processes associated with spacetime because it involves coarse-graining in the sense of eq. (3.1). This is why Hawking’s calculation found apparent violation of unitarity in the black hole evaporation process [9]. In this section, we limit our discussion to the level in which the details of these microscopic processes are ignored. The underlying microscopic dynamics of quantum gravity and its implications for the semiclassical approximation will be discussed further in section 4.

3.1 Thermal nature

According to the semiclassical calculation, a black hole of mass M emits black-body radiation, corrected by gray-body factors, of temperature $T_H = 1/8\pi M l_P^2$. In particular, if a detector is located at $r = r_d$, then it will see blueshifted Hawking radiation with temperature

$$T(r_d) = \frac{1}{8\pi M l_P^2 \sqrt{1 - \frac{2Ml_P^2}{r_d}}}. \quad (3.2)$$

What does this mean at the microscopic level?

Suppose we put the model detector considered in section 2.2 somewhere in the zone, $r_s < r_d \lesssim 3Ml_P^2$. Specifically, the detector has the ready state $|d_0\rangle$ and pointer states $|d_i\rangle$ ($i = 1, 2, \dots$) with the proper energies needed to excite $|d_0\rangle$ to $|d_i\rangle$ given by $E_{d,i}$. Applying

eq. (2.10) (with the understanding that we are focusing on the branch in which the detector has responded, and that the right-hand side is normalized), the interaction between the detector and the mixed state in eq. (3.1) is given by

$$\rho(M) \otimes |d_0\rangle\langle d_0| \rightarrow \frac{1}{n(M)} \sum_{k=1}^{n(M)} \sum_{i,i'} \sum_{k_i=1}^{n(M_i)} \sum_{k'_{i'}=1}^{n(M_{i'})} \alpha_{k_i i}^k \alpha_{k'_{i'} i'}^{k*} |\psi_{k_i}(M_i)\rangle |d_i\rangle \langle \psi_{k'_{i'}}(M_{i'})| \langle d_{i'}|. \quad (3.3)$$

This leads to the density matrix describing the detector state after the interaction

$$\rho_d = \sum_{i,i'} \gamma_{ii'} |d_i\rangle\langle d_{i'}|; \quad \gamma_{ii'} = \frac{1}{n(M)} \sum_{k=1}^{n(M)} \sum_{k_i=1}^{n(M_i)} \alpha_{k_i i}^k \alpha_{k_i i}^{k*} \delta_{ii'}, \quad (3.4)$$

where we have assumed $M_i \neq M_{i'}$ for $i \neq i'$ for simplicity. The result of the semiclassical calculation implies that this density matrix takes the form as if the detector is immersed in the thermal bath of temperature in eq. (3.2). If the sensitivity of the detector does not depend on the excitation level i ,

$$\gamma_{ii'} \approx \frac{1}{Z} e^{-\frac{E_{d,i}}{T(r_d)}} \delta_{ii'}, \quad (3.5)$$

where $Z = \sum_i e^{-E_{d,i}/T(r_d)}$. Note that while the fundamental process in eq. (2.10) leads to the correlation between the detector and the microstate of the black hole k , the density matrix ρ_d in eq. (3.4) does not reflect it, because of the maximally mixed nature of the state $\rho(M)$.

The situation in Hawking emission, in this respect, is similar. As discussed around eq. (2.11), the fundamental emission process, occurring in a time interval of order $1/\delta M \approx M l_P^2$, can be written as

$$|\psi_k(M)\rangle |\phi_a\rangle \rightarrow \sum_i \sum_{k_i=1}^{n(M-E_i)} \beta_{k_i i}^k |\psi_{k_i}(M-E_i)\rangle |\phi_{a+i}\rangle, \quad (3.6)$$

where $|\phi_a\rangle$ is a general far exterior state representing the region $r > R$, and $|\phi_{a+i}\rangle$ is the state in which newly emitted Hawking quanta, labeled by i and having total energy E_i , are added to the appropriately time evolved $|\phi_a\rangle$. Applying this to the “semiclassical state” in eq. (3.1), i.e.

$$\rho(M) = \frac{1}{n(M)} \sum_{k=1}^{n(M)} |\psi_k(M)\rangle |\phi_a\rangle \langle \psi_k(M)| \langle \phi_a|, \quad (3.7)$$

its evolution is given by

$$\rho(M) \rightarrow \frac{1}{n(M)} \sum_{k=1}^{n(M)} \sum_{i,i'} \sum_{k_i=1}^{n(M-E_i)} \sum_{k'_{i'}=1}^{n(M-E_{i'})} \beta_{k_i i}^k \beta_{k'_{i'} i'}^{k*} |\psi_{k_i}(M-E_i)\rangle |\phi_{a+i}\rangle \langle \psi_{k'_{i'}}(M-E_{i'})| \langle \phi_{a+i'}|. \quad (3.8)$$

The result of the semiclassical calculation implies that

$$\frac{1}{n(M)} \sum_{k=1}^{n(M)} \beta_{k,i}^k \beta_{k',i'}^{k*} \approx \frac{1}{\bar{Z}} g_i e^{-\frac{E_i}{T_H}} \delta_{k,k'} \delta_{i,i'}, \quad (3.9)$$

where $\bar{Z} = \sum_i n(M - E_i) g_i e^{-E_i/T_H}$, and g_i is the gray-body factor calculable in the semi-classical analysis [56]. The evolution in eq. (3.8) is then approximated by

$$\rho(M) \rightarrow \frac{1}{\bar{Z}} \sum_i n(M - E_i) g_i e^{-\frac{E_i}{T_H}} \rho_{\text{fin},i} \approx \frac{1}{Z} \sum_i g_i e^{-\frac{E_i}{T_H}} \rho_{\text{fin},i}, \quad (3.10)$$

where $Z = \sum_i g_i e^{-E_i/T_H}$, and we have ignored the i dependence of $n(M - E_i)$ in the last expression. The density matrix

$$\rho_{\text{fin},i} = \frac{1}{n(M - E_i)} \sum_{k_i=1}^{n(M-E_i)} |\psi_{k_i}(M - E_i)\rangle |\phi_{a+i}\rangle \langle \psi_{k_i}(M - E_i)| \langle \phi_{a+i}|, \quad (3.11)$$

represents the coarse-grained state in which the newly created Hawking quanta are in state i , so that eq. (3.10) corresponds to the well-known result in semiclassical gravity. Again, the final expression in eq. (3.10) does not depend on microstates of the black hole, despite the fact that the elementary process in eq. (3.6) is unitary, so that the coefficients $\beta_{k,i}^k$ depend on k . This elucidates why the semiclassical calculation sees apparent violation of unitarity in the Hawking emission process — it deals with the mixed state, eq. (3.7), from the beginning.

3.2 Interior spacetime

The fact that the semiclassical approximation is built on the maximally mixed state, eq. (3.1), implies that it cannot probe the index k labeling the microstates for the spacetime. In particular, a structure seen in semiclassical gravity is either factored from k or arises as a result of taking the maximal mixture. Here we elucidate how this picture works for the emergence of the interior spacetime of a black hole.

Consider a black hole formed by a gravitational collapse. The spacetime region relevant for a distant description is then the region outside the event horizon, which we identify as the Schwarzschild horizon ignoring a small difference (inessential here) between the two: $r > 2Ml_P^2 \equiv R_S$. We may split the degrees of freedom in this region into two classes: those in $R_S < r \leq r_s$ and in $r > r_s$. The former is the stretched horizon degrees of freedom, which are intrinsically quantum gravitational, while the latter is well described by a field theory at low energies. Now, consider Hilbert space $\mathcal{H}_{\tilde{\psi}_k}$ spanned by $|\psi_k(M)\rangle$ and all the excited states of it, $|\tilde{\psi}_{kn}(M)\rangle$, for fixed k . Assuming that the number of excitations is sufficiently small that eq. (2.3) is satisfied, a basis vector of this Hilbert space is effectively specified by the state of the stretched horizon, labeled by \tilde{i} , and the number of excitations n_σ in each field theory mode σ in the zone region:

$$|\tilde{\psi}_{k;\tilde{i}\{n_\sigma\}}(M)\rangle = |\tilde{i};k\rangle |\{n_\sigma\};k\rangle. \quad (3.12)$$

Here, we assume that the field theory modes σ are defined using the Schwarzschild time t at a sufficiently late time when the geometry in the region $r > R_S$ is well approximated by the Schwarzschild spacetime.

Note that in eq. (3.12), we have kept the index k in the stretched horizon and zone states to remind ourselves that they represent excitations on the black hole vacuum state $|\psi_k(M)\rangle$. This, however, does not mean that the states $|\tilde{i}; k\rangle$ and $|\tilde{i}; k'\rangle$ for $k \neq k'$, or $|\{n_\sigma\}; k\rangle$ and $|\{n_\sigma\}; k'\rangle$ for $k \neq k'$, are all independent. In fact, to see the structure of the independence of these states, it is better to write eq. (3.12) in a form analogous to eq. (2.6):

$$|\tilde{\psi}_{k;\tilde{i}\{n_\sigma\}}(M)\rangle = |\tilde{i}, g\rangle |\{n_\sigma\}, h\rangle, \quad (3.13)$$

where $k = \{g, h\}$ with $g = 1, \dots, G$ and $h = 1, \dots, H$. The states $|\tilde{i}, g\rangle$ and $|\{n_\sigma\}, h\rangle$ may then be viewed as all independent:

$$\langle \tilde{i}, g | \tilde{i}, g' \rangle = \delta_{gg'}, \quad \langle \{n_\sigma\}, h | \{n_\sigma\}, h' \rangle = \delta_{hh'}, \quad (3.14)$$

with

$$GH = n(M). \quad (3.15)$$

We expect that the dimensions of the Hilbert space factors for the stretched horizon and zone states are roughly comparable, $\ln G \approx \ln H \approx O(M^2 l_P^2)$; their precise ratio depends on how we divide the degrees of freedom between the stretched horizon and the zone for the modes around $r \approx r_s$. In the rest of the paper we adopt the notation in eq. (3.12), which makes the connection to the vacuum states clearer. It should, however, be understood that the index k represents the information that is *shared* between (and not possessed by both) the stretched horizon and zone states, in the sense of eqs. (3.13), (3.15).

Now, the analysis in semiclassical gravity tells us that to describe the interior of a black hole, we need quantum field theory modes corresponding to the *second* exterior region of a maximally extended — or two-sided — black hole, in addition to the “original” exterior modes σ [10, 57]. Where do such modes come from? We assume that the required modes arise from excitations of the stretched horizon degrees of freedom. In particular, the intrinsically quantum gravitational dynamics at the stretched horizon organize these modes such that they “mirror” the structure of the near exterior modes, which can be interpreted as being located in the region outside the stretched horizon of the second exterior region [15]. In order to reproduce the relevant interior spacetime region, the stretched horizon has to mirror the modes in the region $r_s^* < r^* < 0$ (corresponding to the causal past of a $t = 0$ point on singularities, assuming the reflection symmetry between the two exterior regions), but it need not do more; in particular, we expect that the stretched horizon does not produce modes mirroring the far exterior modes.

General near horizon black hole states built on a vacuum state $|\psi_k(M)\rangle$ may then be written as

$$|\tilde{\psi}_k(M)\rangle = \sum_{\{\tilde{n}_\sigma\}, \{n_\sigma\}} d_{\{\tilde{n}_\sigma\} \{n_\sigma\}} |\{\tilde{n}_\sigma\}; k\rangle |\{n_\sigma\}; k\rangle. \quad (3.16)$$

The stretched horizon states are now labeled by the set of occupation numbers, \tilde{n}_σ , for all the mirror modes, which we have labeled using the same symbol σ as the original modes.⁴ We can now define mode operators $\tilde{b}_\sigma^{(k)}$ and $\tilde{b}_\sigma^{(k)\dagger}$ acting on the states for each k as

$$\tilde{b}_\sigma^{(k)} |\{\tilde{n}_{\sigma'}\}; k'\rangle |\{n_{\sigma''}\}; k'\rangle = \delta_{kk'} \sqrt{\tilde{n}_\sigma} |\{\tilde{n}_{\sigma'} - \delta_{\sigma\sigma'}\}; k\rangle |\{n_{\sigma''}\}; k\rangle, \quad (3.17)$$

$$\tilde{b}_\sigma^{(k)\dagger} |\{\tilde{n}_{\sigma'}\}; k'\rangle |\{n_{\sigma''}\}; k'\rangle = \delta_{kk'} \sqrt{\tilde{n}_\sigma + 1} |\{\tilde{n}_{\sigma'} + \delta_{\sigma\sigma'}\}; k\rangle |\{n_{\sigma''}\}; k\rangle, \quad (3.18)$$

analogously to the annihilation-creation operators $b_\sigma^{(k)}$ and $b_\sigma^{(k)\dagger}$ for the near exterior modes:

$$b_\sigma^{(k)} |\{\tilde{n}_{\sigma''}\}; k'\rangle |\{n_{\sigma'}\}; k'\rangle = \delta_{kk'} \sqrt{n_\sigma} |\{\tilde{n}_{\sigma''}\}; k\rangle |\{n_{\sigma'} - \delta_{\sigma\sigma'}\}; k\rangle, \quad (3.19)$$

$$b_\sigma^{(k)\dagger} |\{\tilde{n}_{\sigma''}\}; k'\rangle |\{n_{\sigma'}\}; k'\rangle = \delta_{kk'} \sqrt{n_\sigma + 1} |\{\tilde{n}_{\sigma''}\}; k\rangle |\{n_{\sigma'} + \delta_{\sigma\sigma'}\}; k\rangle. \quad (3.20)$$

These operators then satisfy the commutation relations

$$[\tilde{b}_\sigma^{(k)}, \tilde{b}_{\sigma'}^{(k')\dagger}] = \delta_{\sigma\sigma'} \delta_{kk'} P_k, \quad [\tilde{b}_\sigma^{(k)}, \tilde{b}_{\sigma'}^{(k')}] = [\tilde{b}_\sigma^{(k)\dagger}, \tilde{b}_{\sigma'}^{(k')\dagger}] = 0, \quad (3.21)$$

$$[b_\sigma^{(k)}, b_{\sigma'}^{(k')\dagger}] = \delta_{\sigma\sigma'} \delta_{kk'} P_k, \quad [b_\sigma^{(k)}, b_{\sigma'}^{(k')}] = [b_\sigma^{(k)\dagger}, b_{\sigma'}^{(k')\dagger}] = 0, \quad (3.22)$$

which imply that we can interpret $\tilde{b}_\sigma^{(k)}$ and $\tilde{b}_\sigma^{(k)\dagger}$ as the annihilation-creation operators for “mirror quanta” arising as collective excitation modes of the stretched horizon degrees of freedom. Here, P_k is the projection operator on $\mathcal{H}_{\tilde{\psi}_k}$, i.e. $P_k |\tilde{\psi}_{k'}(M)\rangle = \delta_{kk'} |\tilde{\psi}_k(M)\rangle$, which appears because $\tilde{b}_\sigma^{(k)}$ and $b_\sigma^{(k)}$ involve projection on $\mathcal{H}_{\tilde{\psi}_k}$, i.e. $\tilde{b}_\sigma^{(k)} = \tilde{b}_\sigma^{(k)} P_k$ and $b_\sigma^{(k)} = b_\sigma^{(k)} P_k$. Note that this is not a local operator in the usual sense of field theory, since the information about k is spread over the stretched horizon as well as the zone regions.⁵

The construction of the operators relevant for describing the black hole spacetime can now go as in the semiclassical analysis [10]. Specifically, we can split the modes σ into two classes: outgoing, σ^+ , and ingoing, σ^- . For the outgoing modes, we introduce the mode operators associated with the Killing vector on the past horizon, $\partial/\partial U$ where $U = -Ml_P^2 e^{(r^*-t)/4Ml_P^2}$, in the analytically extended black hole background:

$$a_\xi^{(k)} = \sum_{\sigma^+} (\alpha_{\xi\sigma^+} b_{\sigma^+}^{(k)} + \gamma_{\xi\sigma^+} b_{\sigma^+}^{(k)\dagger} + \zeta_{\xi\sigma^+} \tilde{b}_{\sigma^+}^{(k)} + \eta_{\xi\sigma^+} \tilde{b}_{\sigma^+}^{(k)\dagger}), \quad (3.23)$$

where ξ represents the quantum numbers labeling the modes, and $\alpha_{\xi\sigma^+}$, $\gamma_{\xi\sigma^+}$, $\zeta_{\xi\sigma^+}$, and $\eta_{\xi\sigma^+}$ are the Bogoliubov coefficients, calculable using the standard quantum field theory method. These operators satisfy the commutation relations for annihilation-creation operators

$$[a_\xi^{(k)}, a_{\xi'}^{(k')\dagger}] = \delta_{\xi\xi'} \delta_{kk'} P_k, \quad [a_\xi^{(k)}, a_{\xi'}^{(k')}] = [a_\xi^{(k)\dagger}, a_{\xi'}^{(k')\dagger}] = 0. \quad (3.24)$$

⁴In a distant description, a stretched horizon mode specified by σ , being intrinsically quantum gravitational degrees of freedom, need not have the same localization property in the angular directions as the corresponding mode in the near exterior region.

⁵We consider that the states obtained by acting operators $b_\sigma^{(k)}$, $b_\sigma^{(k)\dagger}$, $\tilde{b}_\sigma^{(k)}$, and $\tilde{b}_\sigma^{(k)\dagger}$ on a black hole vacuum state are not all physically realized states, so that the actual physical Hilbert space is smaller than the Fock space implied by the construction here. In particular, the dimension of the physical Hilbert space for the states in which there are negative energy excitations (as defined in the asymptotic region) must be smaller than that of the naive Fock space described here, implying that the entropy S_I associated with a negative energy excitation I is negative in eq. (2.3). For more discussions on this point, see section 4.2.

The near horizon black hole vacuum state $|\psi_k(M)\rangle$ is then well approximated by the condition

$$\forall \xi, \sigma^- \quad a_\xi^{(k)} |\psi_k(M)\rangle = b_{\sigma^-}^{(k)} |\psi_k(M)\rangle = \tilde{b}_{\sigma^-}^{(k)} |\psi_k(M)\rangle = 0. \quad (3.25)$$

Ignoring any excitations, this state takes approximately the thermofield double form:

$$|\psi_k(M)\rangle = \frac{1}{\sqrt{Z}} \sum_{\{n_{\sigma^+}\}} e^{-\frac{E_{\{n_{\sigma^+}\}}}{2T_H}} |\{\tilde{n}_{\sigma^+} = n_{\sigma^+}\}; k\rangle |\{n_{\sigma^+}\}; k\rangle; \quad Z = \sum_{\{n_{\sigma^+}\}} e^{-\frac{E_{\{n_{\sigma^+}\}}}{T_H}}, \quad (3.26)$$

where $E_{\{n_{\sigma^+}\}}$ is the energy of the state $|\{n_{\sigma^+}\}; k\rangle$ as measured in the asymptotic region, and the occupation numbers for the ingoing modes in $|\{\tilde{n}_{\sigma^+} = n_{\sigma^+}\}; k\rangle$ and $|\{n_{\sigma^+}\}; k\rangle$ are zero.

In general, the microscopic Hamiltonian in quantum gravity may depend in a complicated way on the sets of operators $a_\xi^{(k)}$, $b_{\sigma^-}^{(k)}$, and $\tilde{b}_{\sigma^-}^{(k)}$ as well as other operators, including the ones that act nontrivially on the index k :

$$H(M) = H_{\text{QG}}\left(\{a_\xi^{(k)}\}, \{b_{\sigma^-}^{(k)}\}, \{\tilde{b}_{\sigma^-}^{(k)}\}, \dots; M\right), \quad (3.27)$$

even if we restrict our considerations to the states having spacetime with a black hole at a fixed location of mass between M and $M+\delta M$. (If we consider processes in which M varies, the total Hamiltonian must be taken as the sum of the $H(M)$'s with different M 's, with terms allowing hopping between different M 's added. Of course, the segmentation of the black hole mass into different discrete M 's with some widths is completely artificial, and thus can be done in any arbitrary way.) On the other hand, the semiclassical approximation is built on the maximally mixed vacuum state, eq. (3.1). What is the precise meaning of this statement?

Consider a general semiclassical state

$$\tilde{\rho}(M) = f(a_\xi^\dagger, b_{\sigma^-}^\dagger, \tilde{b}_{\sigma^-}^\dagger) \rho(M) f(a_\xi^\dagger, b_{\sigma^-}^\dagger, \tilde{b}_{\sigma^-}^\dagger)^\dagger, \quad (3.28)$$

where $\rho(M)$ is the semiclassical vacuum state in eq. (3.1), and f represents the excitation. Here, a_ξ , b_{σ^-} , and \tilde{b}_{σ^-} are (semiclassical) quantum field theory operators defined by

$$a_\xi = \sum_k a_\xi^{(k)}, \quad b_{\sigma^-} = \sum_k b_{\sigma^-}^{(k)}, \quad \tilde{b}_{\sigma^-} = \sum_k \tilde{b}_{\sigma^-}^{(k)}, \quad (3.29)$$

which satisfy the commutation relations

$$[a_\xi, a_{\xi'}^\dagger] = \delta_{\xi\xi'} \sum_k P_k, \quad [b_{\sigma^-}, b_{\sigma'^-}^\dagger] = [\tilde{b}_{\sigma^-}, \tilde{b}_{\sigma'^-}^\dagger] = \delta_{\sigma^-\sigma'^-} \sum_k P_k, \quad \text{others} = 0. \quad (3.30)$$

Note that the operator $\sum_k P_k$ becomes unity when acting on any normalized (pure or mixed) state in $\bigoplus_k \mathcal{H}_{\tilde{\psi}_k}$, so that eq. (3.30) indeed represents the standard commutation relations for the annihilation-creation operators.

The semiclassical approximation asserts that the evolution of the state in eq. (3.28), which at the microscopic level is generated by the Hamiltonian in eq. (3.27), is well approximated by the evolution caused by the standard quantum field theory Hamiltonian:

$$\tilde{\rho}(M) \rightarrow e^{-iH(M)t} \tilde{\rho}(M) e^{iH(M)t}, \quad H(M) \approx H_{\text{QFT}}(a_\xi, b_{\sigma^-}, \tilde{b}_{\sigma^-}; M), \quad (3.31)$$

where the dependence of H_{QFT} on the creation, as well as annihilation, operators is implied. Note that here we have used the a_ξ operators to describe the outgoing modes in the semiclassical field theory, but we may instead use b_{σ^+} and \tilde{b}_{σ^+} operators defined analogously to eq. (3.29):

$$b_{\sigma^+} = \sum_k b_{\sigma^+}^{(k)}, \quad \tilde{b}_{\sigma^+} = \sum_k \tilde{b}_{\sigma^+}^{(k)}. \quad (3.32)$$

These two descriptions correspond, respectively, to seeing the system from an infalling and static (or distant) observer's viewpoints.⁶

The expression in eq. (3.31) makes it clear that semiclassical gravity cannot describe detailed microscopic physics associated with the index k . In fact, at this level of approximation, the Hamiltonian can be written as

$$H(M) \approx H_{\text{QFT}}(a_\xi, b_{\sigma^-}, \tilde{b}_{\sigma^-}; M) = \sum_k H_{\text{QFT}}(a_\xi^{(k)}, b_{\sigma^-}^{(k)}, \tilde{b}_{\sigma^-}^{(k)}; M), \quad (3.33)$$

since $a_\xi^{(k)}$, $b_{\sigma^-}^{(k)}$, and $\tilde{b}_{\sigma^-}^{(k)}$ (as well as $b_{\sigma^+}^{(k)}$ and $\tilde{b}_{\sigma^+}^{(k)}$) involve projection on \mathcal{H}_{ψ_k} . This implies that the dynamics involving (only) excitations on a fixed black hole background is decomposed into $n(M) = e^{4\pi M^2 l_P^2}$ decoupled, identical copies. Since this statement also applies to any fixed spatial geometry arising for a sufficiently short time interval (with the appropriate replacement of the number of microstates), we may simply drop the index k in describing the dynamics of the excitations, and write the evolution of the system as

$$|\tilde{\Psi}\rangle \rightarrow e^{-iH_{\text{QFT}}t} |\tilde{\Psi}\rangle. \quad (3.34)$$

Here, $|\tilde{\Psi}\rangle$ is a field theoretical state of the system built on a field theoretical vacuum $|\Psi\rangle$, e.g., $|\tilde{\Psi}\rangle = f(a_\xi^\dagger, b_{\sigma^-}^\dagger, \tilde{b}_{\sigma^-}^\dagger)|\Psi\rangle$ with f in eq. (3.28). This is what the semiclassical field theory is! It provides a unitary description of physics *unless* we consider processes involving the microscopic spacetime index k such as the Hawking emission process.

3.3 UV/IR correspondence and relation to the entanglement entropy

Before concluding this section, we would like to highlight a key assumption adopted in section 3.2 as well as its implications. The assumption may be called the *stretched horizon/second exterior (SH/SE) correspondence*:

The stretched horizon degrees of freedom representing excitations on a black hole vacuum state $|\psi_k(M)\rangle$ can be organized into the modes that are interpreted as being located outside the stretched horizon of the second (near) exterior region of the corresponding extended, or two-sided, black hole of mass M .

Under this correspondence, the “trans-Planckian” — or stringy — excitations in the region $R_S < r \leq r_s$ in the “original” one-sided picture are mapped into low energy field theory excitations outside the stretched horizon of the second exterior region in the corresponding

⁶The ξ modes correspond to the modes defined in the past null infinity using the Minkowski time if we treat the collapsing geometry faithfully (without invoking the eternal black hole approximation at late times) ignoring the trans-Planckian nature of the encounters of the modes with the collapsing matter.

two-sided picture. This implies that there are no (or few) excitations at a distance below the quantum gravitational lengthscale in the latter picture (at least around the bifurcation surface). We may therefore colloquially summarize this correspondence as: the full string theory excitations defined in the region outside the mathematical Schwarzschild horizon, $r > R_S$, correspond to field theory excitations defined in the extended, two-sided black hole spacetime.

While we have motivated the correspondence by the physical picture in semiclassical gravity, it would be desirable to derive or motivate it more directly in the fundamental theory of quantum gravity. In fact, we view this as an interesting opportunity. The description in section 3.2 requires specific properties for the dynamics of the stretched horizon (the UV degrees of freedom above the quantum gravitational scale), reminiscent of UV/IR relations seen in other settings in string theory (such as AdS/CFT and T duality). This therefore provides an explicit suggestion on what properties the fundamental theory of quantum gravity must have in order to reproduce the predictions of general relativity in the appropriate classical limit, in situations in which spacetime is far from simple Minkowski or anti-de Sitter space.

The picture described here also provides a simple interpretation of possible relations between the black hole and entanglement entropies contemplated long ago in ref. [58]. As we have discussed, the relevant spacetime region in a distant description of a black hole is the outside of the mathematical horizon $r > R_S$, namely the near side of the Schwarzschild horizon as viewed from the “origin of the reference frame” located outside the horizon. The number of independent quantum states in full quantum gravity describing this region — more precisely the near horizon part of it — is given by $n(M) = e^{\mathcal{A}/4l_P^2}$, where \mathcal{A} is the horizon area.⁷ How can we calculate this number without invoking the thermodynamic argument as was originally done by Bekenstein and Hawking?

One way to do so is the following. We *fictitiously* extend the spacetime beyond the horizon $r = R_S$ by considering another “mirror” region parameterized by $r' > R_S$ sewn to the original region at $r = r' = R_S$. (Note that the mirror region, $r' > R_S$, in this discussion is really fictitious, since we are describing the system from a distant viewpoint using the fundamental theory, so that the degrees of freedom in the original region, $r > R_S$, already contain the stretched horizon degrees of freedom, which are sufficient to construct the interior of the black hole.) We may then consider a state in which the degrees of freedom in the two regions, $r > R_S$ and $r' > R_S$, are (nearly) maximally entangled. The number of independent states available in one region can then be obtained as the exponential of the entanglement entropy between the two regions.

For black hole spacetime, we may consider the intersection of the two regions, $r = r' = R_S$, to be a bifurcation surface, and the mirror region to be the other side of that surface

⁷This statement may be generalized to a more concrete statement if we include the “boundary” of the ambient space as viewed from the origin of the reference frame p_0 : the number of independent quantum states representing the region enclosed by the black hole horizon and the boundary (as viewed from p_0) is given by the exponential of the sum of the horizon and boundary areas in units of $4l_P^2$. (In Minkowski space, the boundary may be located at infinite distances away and may have an infinite area.) For further discussions, see section 5 (and ref. [59]).

on an equal-time hypersurface passing through it. Since we are interested in the number of microstates in full quantum gravity, this construction must be done in the full theory of quantum gravity, implying that the mirror region, $r' > R_S$, also “mirrors” the effective second exterior region of the extended black hole, arising from the stretched horizon modes in the original region. Now, we expect that the local inertial vacuum state around $r \sim R_S$ in this “doubled spacetime” involves near maximal entanglement between the states in the two regions, $r > R_S$ and $r' > R_S$, when viewed from a static/distant observer because of a large relative acceleration between the two reference frames. We may therefore obtain the black hole entropy by calculating the entanglement entropy between the two regions for the local inertial vacuum state in the full theory of quantum gravity. This corresponds to the calculation performed in ref. [60] in string theory, which indeed found that the entanglement entropy is $\mathcal{A}/4l_P^2$ at the leading order in l_P^2/\mathcal{A} .

Do we really need string theory to calculate the entanglement entropy? In low energy quantum field theories, the result of the calculation will be divergent, which must be made finite by counterterms. The freedom in adjusting these counterterms represents a possible variation of UV theories that lead to consistent quantum field theories at low energies. At the leading order in l_P^2/\mathcal{A} , however, no such freedom is left after the Planck length, l_P , is renormalized to a finite value, and the calculation in ref. [61] indeed finds that the entanglement entropy after this renormalization is finite and takes an unambiguous value of $\mathcal{A}/4l_P^2$. (To be precise, ref. [61] showed this only for certain limited cases, but we may expect it applies more generally.) This implies that any consistent UV theory leads to the entanglement entropy between the two regions of $\mathcal{A}/4l_P^2$ in the local inertial vacuum state. As argued above, this measures the number of independent quantum states available in one of the regions in the full theory of quantum gravity. In the context of a black hole described in a distant reference frame, these states comprise all of the states in the near horizon region, since the mirror region $r' > R_S$ (again, not to be confused with the second exterior region in the extended black hole approximation) does not really exist.

4 Nature of microscopic dynamics

We now discuss the nature of microscopic dynamics associated with Hawking emission and measuring the black hole’s thermal atmosphere by a physical detector (the mining process). A complete treatment of this issue requires the fundamental theory of quantum gravity. Our focus here is how these dynamics manifest themselves in the semiclassical approximation, in particular what physical conclusions we may draw using semiclassical analyses.

We argue that a measurement of the black hole’s atmosphere by a static detector consists of two effects: the effect caused by acceleration of the detector (the Unruh effect) and that by nonzero curvature of spacetime. The two combined make the detector respond as if it is immersed in the thermal bath at a blueshifted local Hawking temperature. If the detector is located in a region away from the black hole, then the former acceleration induced effect becomes negligible, and its response is dominated by the latter curvature

induced effect. It is this latter effect that represents the Hawking radiation emitted to the asymptotic region.

We assert that the microscopic details associated with the information transfer are in general not visible in semiclassical gravity, although a part of the backreaction of the acceleration induced effect can be described in semiclassical field theory. In particular, this implies that at the level of semiclassical gravity, the curvature induced effect is described only as “physics of a vacuum”, which arises from an intrinsic quantum mechanical ambiguity of local vacuum energy density (or particle numbers) of order $1/(Ml_P^2)^4$ that cannot be resolved in semiclassical field theory. One must therefore be careful in applying intuition from semiclassical field theory in analyzing the microscopic processes underlying Hawking radiation.

4.1 Response of a static detector

Consider a physical detector located at a fixed Schwarzschild radial coordinate $r = r_d$. According to the semiclassical calculation, the detector responds as if it is immersed in the thermal bath of temperature $T(r_d)$ in eq. (3.2). What causes this phenomenon?

First of all, since the detector at a fixed coordinate point $r = r_d$ is accelerated with respect to local inertial frames, the effect in ref. [10] — the Unruh effect — makes the detector react accordingly. The magnitude of the proper acceleration of the detector is given by

$$a(r_d) = \frac{1}{\sqrt{1 - \frac{2Ml_P^2}{r_d}}} \frac{Ml_P^2}{r_d^2}. \quad (4.1)$$

If this were the only effect, then the detector would see a thermal bath of temperature

$$T_U(r_d) = \frac{a(r_d)}{2\pi} = \frac{Ml_P^2}{2\pi r_d^2 \sqrt{1 - \frac{2Ml_P^2}{r_d}}}, \quad (4.2)$$

which does not agree with $T(r_d)$ except when the detector is located at the horizon, $r_d \rightarrow 2Ml_P^2$. In particular, for $r_d \rightarrow \infty$ the effect from acceleration disappears, $T_U(r_d) \rightarrow 0$, so that it cannot be responsible for Hawking radiation measured at the asymptotic infinity. What is the remaining effect making up the difference between $T_U(r_d)$ and $T(r_d)$?

To look for that effect, we can make the following heuristic argument. Consider a detector placed somewhere in the zone. If there were the only acceleration effect, the local energy density at the detector location (as defined in the asymptotic region) would be

$$\rho_U(r_d) = \frac{c}{1 - \frac{2Ml_P^2}{r_d}} \left(\frac{Ml_P^2}{2\pi r_d^2} \right)^4, \quad (4.3)$$

where $c = \pi^2 g_*/30$ is a numerical coefficient with g_* being the effective number of relativistic degrees of freedom, and the quantity in the parentheses is the Unruh temperature in eq. (4.2) corrected by the redshift factor $\sqrt{1 - 2Ml_P^2/r_d}$. On the other hand, the semiclassical analysis tells us that this quantity must be

$$\rho(r_d) = \frac{c T_H^4}{1 - \frac{2Ml_P^2}{r_d}} = \frac{c}{1 - \frac{2Ml_P^2}{r_d}} \left(\frac{1}{8\pi Ml_P^2} \right)^4, \quad (4.4)$$

consistent with the blueshifted Hawking temperature. The difference is given by

$$\rho(r_d) - \rho_U(r_d) = c \left(\frac{1}{8\pi M l_P^2} \right)^4 f \left(\frac{r_d}{2M l_P^2} \right); \quad f(x) = \frac{1 - \frac{1}{x^8}}{1 - \frac{1}{x}}. \quad (4.5)$$

Since $f(x)$, defined for $x > 1$, is a monotonically decreasing function with $f(x \rightarrow 1) = 8$ and $f(x \gg 1) \rightarrow 1$, we find that the effect we seek is giving a contribution

$$\rho(r) - \rho_U(r) \approx O \left(\frac{1}{(8\pi M l_P^2)^4} \right), \quad (4.6)$$

throughout the entire zone region.

The local energy density of the amount in eq. (4.6) is precisely what we would expect to arise from the intrinsic quantum mechanical ambiguity of defining particle states in semiclassical gravity. Since the curvature lengthscale around the black hole is of order $M l_P^2$, the concept of particles is ambiguous beyond this lengthscale. This makes particle states with $\omega \lesssim 1/M l_P^2$ ambiguous, leading to an uncertainty of the local vacuum energy density of order $1/(M l_P^2)^4$. Here, ω is the frequency as defined in the asymptotic region. We assume that this curvature induced effect is responsible for the difference in eq. (4.6), i.e. the difference between $T_U(r_d)$ and $T(r_d)$. In particular, this effect is solely responsible for the response of a detector located in the far asymptotic region, i.e. the original Hawking radiation effect calculated using the in-out formalism [2]. Unlike the acceleration induced effect, this effect cannot be eliminated by going to a local inertial frame.

The separation of the origin of the detector response into the two effects as described above is useful in discussing what aspects of the underlying process are visible in semiclassical gravity. At the microscopic level, both these effects occur according to the full Hamiltonian evolution in quantum gravity, and all the microscopic details associated with them can be calculated, at least, in principle. The same, however, is not true in the semiclassical approximation. We now see to what extent the details of the underlying dynamics can be captured in semiclassical field theory for each of these two effects.

4.2 Effect from acceleration

Imagine that a physical detector is held near the horizon. The detector then responds to a highly blueshifted Hawking temperature. As can be seen from the fact that

$$\frac{\rho_U(r)}{\rho(r) - \rho_U(r)} = \frac{1}{\left(\frac{r}{2M l_P^2} \right)^4 - 1} \gg 1 \quad \text{for } r \simeq 2M l_P^2, \quad (4.7)$$

this response is caused mostly by the acceleration of the detector with respect to local inertial frames. The dynamics of this effect, including the backreaction, was analyzed in ref. [62] in semiclassical field theory, which we may import to study our black hole problem here.

Suppose the detector is coupled to a field φ . When described in a static reference frame, the backreaction of a detector response eliminates a φ particle from the thermal bath, interpreted as being absorbed by the detector: $|\Psi(M)\rangle \rightarrow b_\sigma |\Psi(M)\rangle$, where $|\Psi(M)\rangle$

represents the field theory black hole vacuum, and an appropriate superposition of b_σ is implied. In an infalling frame, however, the same process is described as an *emission* of a φ particle by the detector, i.e. the backreaction of the detector response is the creation of a particle, as can be seen from the fact that b_σ involves a superposition of a_ξ and a_ξ^\dagger , and $a_\xi|\Psi(M)\rangle = 0$. As expected, the effect of this backreaction is limited within the causal future of the detector response event. In this region, however, the backreaction indeed *does* make the system deviate from the vacuum state, which is visible in the semiclassical approximation. In fact, the effect can be quite significant if the detector is located very close to the horizon.

It is important to understand the physical origin of this backreaction. Suppose we describe the detector in an infalling reference frame. When it responds, i.e. gets excited, it emits a φ particle to the environment, so that the (appropriately defined) total energy of the system increases. Where does this energy come from? In order for the detector to measure a φ particle, it must be held at $r = r_d$ for a Schwarzschild time of order $\Delta t \approx M l_P^2$. This requires an external force — without it, the detector would fall into the horizon within a much shorter time. It is this force that is responsible for the energy needed for the detector response as well as the φ emission, i.e. the backreaction. Since the vacuum is physically disturbed by the external non-gravitational force, it is no surprise that there is a backreaction effect visible clearly in the semiclassical approximation.

What about the information transfer? It is known that one can accelerate the energy loss rate of a black hole by extracting its energy from the atmosphere using a physical apparatus: the mining process [63]. The most efficient, and essentially the only, way to do this is to thread the horizon with strings that have the maximal tension-to-linear-mass-density ratio allowed by the null energy condition, which can make the black hole lifetime as short as $O(M^2 l_P^3)$ [64]. Now, the fact that the energy is extracted by the apparatus means that the information must also be extracted, since otherwise the entropy of the black hole would be oversaturated beyond the Bekenstein-Hawking value. How can information be transferred in such a process?⁸

Suppose we mine a black hole with the model “detector” considered before, around eqs. (2.10) and (3.3). This detector may be viewed as a toy model for the strings described above. The evolution of the combined black hole and detector system then occurs in steps. First, due to the acceleration effect, the detector responds with some probability:

$$|\psi_k(M)\rangle|d_0\rangle \rightarrow \sum_{i=0}^{i_{\max}} \sum_{k'=1}^{n(M)} \zeta_{k'i}^k |\tilde{\psi}_{k'i}(M)\rangle|d_i\rangle, \quad (4.8)$$

where we have included the possibility that no response occurs in the relevant time interval, represented by $i = 0$ in the sum. The states $|\tilde{\psi}_{k'i}(M)\rangle$ arise as a result of the backreaction of the detector response, and are not vacuum states. In the static description, the evolution in eq. (4.8) occurs because the detector may absorb particles from the thermal bath; in the inertial description, it represents the (probabilistic) emission of particles from the detector, accompanied by the internal excitation.

⁸Our analysis in the previous version of this paper on this issue contained an error. We thank Joseph Polchinski for a comment that made us identify and correct that error.

It is important that the coefficients $\zeta_{k'i}^k$ in eq. (4.8) represents a unitary map from the k space to its image in the $(k'i)$ space, not in the k' space. More specifically, for a fixed i the rank of the matrix $(\zeta^i)_{kk'} \equiv \zeta_{k'i}^k$ in the $k-k'$ space is $n(M_i)$, which is smaller than $n(M)$, where M_i is given in eq. (2.10). This is because in the static description, which we adopt here, the excitations in the states $|\tilde{\psi}_{k'i}(M)\rangle$ have negative energy resulting from the elimination of particles from the thermal bath, so that they carry negative entropy. In other words, we are assuming that only a subspace of the entire space spanned by all the values of $k' = 1, \dots, n(M)$ is realized by a physical process when the state is accompanied by negative energy excitations i (which implies that the physical Hilbert space for the states having negative energy excitations is smaller than that of the naive Fock space suggested by the construction in section 3.2). This dynamical assumption allows us to avoid the argument for firewalls presented in refs. [11, 12]. (In the inertial, or infalling, description, there is no acceleration induced thermal bath, so that the corresponding part of the k index is not accessible locally by the detector. This is consistent because the description in spacetime in this case is available only for the timescale of the fall.⁹)

The excitations in $|\tilde{\psi}_{k'i}(M)\rangle$ are “emitted” from the detector. These excitations propagate within the causal future of the detector response event, and most of them will interact with the black hole degrees of freedom, in particular the stretched horizon. This results in the evolution of the states

$$|\tilde{\psi}_{k'i}(M)\rangle \rightarrow \sum_{k_i=1}^{n(M_i)} \eta_{k_i}^{k'i} |\psi_{k_i}(M_i)\rangle, \quad (4.9)$$

where we have made the simplifying assumption that all the excitations interact with the black hole degrees of freedom, and the resulting states are fully relaxed into (superpositions of) vacuum states. Substituting eq. (4.9) into eq. (4.8), we obtain the expression in eq. (2.10):

$$|\psi_k(M)\rangle |d_0\rangle \rightarrow \sum_i \sum_{k_i=1}^{n(M_i)} \alpha_{k_i i}^k |\psi_{k_i}(M_i)\rangle |d_i\rangle, \quad (4.10)$$

where $\alpha_{k_i i}^k = \sum_{k'=1}^{n(M)} \zeta_{k'i}^k \eta_{k_i}^{k'i}$. Since the distribution of the information about the vacuum state is expected to follow the thermal entropy calculated in semiclassical field theory, as discussed in eq. (2.6) and below, the rate of information recovery through a mining at $r = r_d$ will be determined by the thermal entropy at $r = r_d$, associated with the blueshifted local Hawking temperature. This implies that the mining process, in fact, is expected to achieve the required acceleration of the information transfer.

In the semiclassical approximation, the indices such as k and k' are not visible because the degrees of freedom represented by these indices cannot be resolved. In practice, this

⁹We may still suspect that the infalling picture provides a unitarily equivalent description for the distant one because of the existence of the “horizon” surrounding the origin, p_0 , of the infalling reference frame [59] (not to be confused with the horizon as viewed from the distant reference frame). As p_0 approaches the singularity, the “horizon” approaches p_0 , eventually making the whole quantum state a “singularity state” [65]; see section 5. This allows us to contemplate the possibility that the fate of the information mined by the detector is mapped to the intrinsically quantum gravitational dynamics of singularity states after p_0 hits the singularity.

implies that semiclassical field theory describes a process as if these indices were absent. For example, it describes the process in eq. (4.8) as

$$|\psi(M)\rangle|d_0\rangle \rightarrow \sum_{i=0}^{i_{\max}} \zeta_i |\tilde{\psi}_i(M)\rangle|d_i\rangle, \quad (4.11)$$

where $|\psi(M)\rangle$ is the field theory vacuum state, and $|\tilde{\psi}_i(M)\rangle$ represents the state having a negative energy excitation labeled by i on $|\psi(M)\rangle$. This corresponds to the description given in ref. [62]. The state $|\tilde{\psi}_i(M)\rangle$ may be considered to relax into the vacuum state $|\psi(M_i)\rangle$ after the excitation interacts with the stretched horizon, although this process cannot be described in semiclassical field theory. In any event, in the semiclassical approximation, the detector does not have any information about the original black hole microstate — the reduced density matrix representing the detector state after the response is simply $(\rho)_{ii'} = |\zeta_i|^2 \delta_{ii'}$, which does not depend on k .

It is instructive to consider what happens in the limit $M \rightarrow \infty$, in which the relevant spacetime in the static description reduces to a Rindler wedge, a portion of Minkowski space. In this case, the entropy density of eq. (2.1) diverges, $S = 4\pi M^2 l_P^2 \rightarrow \infty$, so that we have an infinitely large degeneracy of microscopic vacuum states in any finite energy interval.¹⁰ This implies that to describe any experiment performed in finite time, we must coarse-grain an infinite number of microstates labeled by the index k . This makes eq. (4.11) “exact” in the sense that any finite resolution (or finite time interval) forces us to describe the process as in eq. (4.11), in which the vacuum index k does not appear. In this way, the uniqueness of the Minkowski vacuum is recovered, which is crucial for our ability to describe physics without having (an infinite amount of) information about the specific vacuum we live in. (Another way to see the uniqueness of the Minkowski vacuum, in an inertial description, will be discussed in section 5.)

4.3 Effect from spacetime curvature

We now consider the other effect responsible for the detector response: the curvature induced effect. As discussed before, this effect is solely responsible for the spontaneous Hawking emission process, where the detection of emitted quanta is envisioned only in the asymptotic region. The situation in this case is different from that described above, in that the effect does not require an operation of a non-gravitational force. There is therefore no reason to expect that there must be any backreaction effect visible in the semiclassical approximation. In fact, we assert that the phenomenon of particle creation due to spacetime curvature must be viewed purely as an “activity of a vacuum”, whose details are invisible in semiclassical field theory.

Consider a process in which Hawking quanta are emitted from the black hole to the far exterior region. (We can imagine measuring these quanta by a physical detector at a

¹⁰The exact Minkowski vacuum space may be an artifact of mathematical idealization, and it is possible that in any physically relevant case, there is a boundary/horizon at a finite affine distance in any direction from the origin of the reference frame. Our argument below persists even in this case by replacing the infinities by appropriately large numbers.

faraway place, but we are not interested in that process here.) At the microscopic level, the elementary emission process occurring in a time interval of $t' - t \approx O(Ml_P^2)$ is given as in eq. (3.6), which, ignoring the difference between various E_i 's, reads

$$|\psi_k(M(t))\rangle|\phi_a\rangle \rightarrow \sum_i \sum_{k'=1}^{n(M(t'))} \beta_{k'i}^k |\psi_{k'}(M(t'))\rangle|\phi_{a+i}\rangle. \quad (4.12)$$

We assert that the states $|\psi_{k'}(M(t'))\rangle$ appearing here are vacuum states from the viewpoint of semiclassical field theory. In particular, when we act a field theory operator of the form in eqs. (3.29), (3.32) which has a support in a near horizon region, then all these states respond as the thermofield double state of the form in eq. (3.26).

Let us now discuss a physical picture behind the information extraction processes in eq. (4.12) and in eq. (4.8). The black hole vacuum states $|\psi_k(M)\rangle$ represent the set of microscopic quantum states which, in the semiclassical approximation, all look like the vacuum state in the background of a black hole of mass M in a fixed location. While invisible in semiclassical gravity, these $|\psi_k(M)\rangle$'s have nontrivial structures representing subtle quantum fluctuations of the spacetime beyond the resolution of the semiclassical approximation. The expressions in eqs. (4.8), (4.12) imply that the backreactions on the black hole relevant for the information transfer go into these subtle quantum modes associated with the spacetime. (Note that the excitation i of $|\tilde{\psi}_{k'i}(M)\rangle$ in eq. (4.8) is not directly relevant for the information transfer as can be seen in eq. (4.11).) These backreactions are not visible in semiclassical field theory — they are “too soft” to be resolved. In particular, the emission of a Hawking quantum must be viewed at the semiclassical level as occurring around the barrier region of the effective gravitational potential, with the associated information transfer occurring through the delocalization of the black hole information to the entire zone.

The discussion above says that one must be careful in applying intuition from semiclassical field theory to the black hole information release processes. In Hawking emission, the purifiers of the emitted Hawking quanta are what semiclassical gravity describes as vacuum states. In particular, they are not field theory modes associated with operators $b_{\sigma^+}^{(k)}$ and $b_{\sigma^+}^{(k)\dagger}$ as envisioned in refs. [8, 11, 12]. A similar statement also applies to the mining process; the purifiers of the modes mined from black hole degrees of freedom are black hole vacuum states. The microscopic details of the information transfer processes, in particular the flow of information within the black hole degrees of freedom, are not visible in semiclassical field theory; one can only see certain inclusive quantities that can be calculated as expectation values of field theory operators in the vacuum, e.g. the total energy flux over a large surface [66, 67]. If one wants to include the backreaction on spacetime at the semiclassical level, the best one can do seems to do it “by hand” such that the semiclassical Einstein equation, $R_{\mu\nu} - (1/2)g_{\mu\nu}R = 8\pi l_P^2 \langle T_{\mu\nu} \rangle$, is satisfied [68, 69], which employs the Vaidya metric near the horizon.

4.4 Horizon of an evaporating black hole

Consider describing the formation and evaporation of a black hole in a distant reference frame. At the microscopic level, the whole process is described as a unitary evolution:¹¹

$$|m_{\text{init}}\rangle \rightarrow \sum_{k=1}^{n(M(t))} c_k(t) |\psi_k(M(t))\rangle |r_k(t)\rangle \rightarrow |r_{\text{fin}}\rangle, \quad (4.13)$$

where $|m_{\text{init}}\rangle$, $|r_k(t)\rangle$, and $|r_{\text{fin}}\rangle$ represent the states for the initial collapsing matter, the subsystem complement to the black hole at time t (which includes Hawking radiation emitted earlier), and the final Hawking quanta after the black hole is completely evaporated. For generic initial states and microscopic emission dynamics, this evolution indeed satisfies the behavior outlined by Page [70] on general grounds.

We here highlight two important aspects of the evolution that have not been stated explicitly:

- (i) *Microscopic dynamics need not violate locality at large distances in the description based on a distant reference frame.* Since later Hawking emissions occur within the causal future of the region in which an earlier emission took place, there is no reason to expect that the information recovery process must violate locality, or causality, of the spacetime structure. Namely, while the detailed flow of information is not visible to semiclassical field theory, the speed at which physical information is transferred can still be bounded by the causal structure of the spacetime.
- (ii) *The semiclassical description of the evolution violates unitarity at all stages of the black hole formation and evaporation processes.* First, since the same semiclassical black hole (within the uncertainties needed to localize it in spacetime) can be formed by collapsing matter in different initial states, the description of the black hole formation is not unitary in semiclassical gravity. After the formation, the microscopic dynamics makes the black hole emit Hawking quanta through the unitary process in eq. (3.6) that occurs successively in each time interval of order $M(t)l_P^2$. Semiclassical gravity, however, describes this as a succession of the process in eq. (3.10); in particular, the state of the emitted Hawking quanta in each time interval is given by the incoherent thermal superposition, making the final Hawking radiation state a mixed thermal state. The semiclassical description of a process involving “microscopic degrees of freedom of spacetime” (the index k) is fundamentally non-unitary.

We now ask what a physical object will find if it falls into the horizon of an evaporating black hole. At the microscopic level, the state of a black hole is in general given by

$$\rho_{\text{BH}}(M(t)) = \sum_{k,l=1}^{n(M(t))} c_k(t) c_l^*(t) |\psi_k(M(t))\rangle \langle \psi_l(M(t))|, \quad (4.14)$$

¹¹As before, we ignore macroscopic dispersions of the black hole mass and location, which may be included explicitly if one wants.

obtained by integrating out the rest of the system in the middle expression in eq. (4.13). As discussed in section 4.3, the states $|\psi_k(M(t))\rangle$ are all vacuum states in the sense that their responses to the field theory operators are those of the vacuum state in semiclassical field theory. If a physical object is falling into the black hole, it is represented as excitations on the background: a_ξ^\dagger , $b_{\sigma-}^\dagger$, and $\tilde{b}_{\sigma-}^\dagger$ acting on the microstate. What happens to the object afterward?

Since a_ξ^\dagger consists of a linear combination of operators in the both exterior regions, $b_{\sigma+}$, $b_{\sigma+}^\dagger$ and $\tilde{b}_{\sigma+}$, $\tilde{b}_{\sigma+}^\dagger$, the excitations in general involve entanglement between those in the two regions. One might think this violates causality because the object that has just entered the zone already involves “excitations” in the second exterior region generated by $\tilde{b}_{\sigma+}$, $\tilde{b}_{\sigma+}^\dagger$. This apparent violation, however, is not physical — exciting a black hole vacuum state by mode operators in the first or second region, by itself, does not have an invariant meaning (as can be seen, e.g., from the fact that operating $b_{\sigma+}^\dagger$ and $\tilde{b}_{\sigma+}$ lead to the same state), so that the physical excitations are still confined in the first exterior region near the edge of the zone. The entangled nature of the excitations, however, becomes important when the object enters the horizon.

Because the algebra of the operators representing the object and the structure of the (vacuum) states on which they act are the same as in the semiclassical field theory in the regime of their validity indicated by eq. (2.3), the only condition needed for the object to smoothly pass the horizon is that the dynamics is also the same. Therefore, assuming that the microscopic dynamics describing the fallen object is well approximated by the Hamiltonian of the quantum field theory form in the relevant timescale for the fall:

$$H(M(t)) \approx H_{\text{QFT}}(a_\xi, b_{\sigma-}, \tilde{b}_{\sigma-}; M(t)) = \sum_k H_{\text{QFT}}(a_\xi^{(k)}, b_{\sigma-}^{(k)}, \tilde{b}_{\sigma-}^{(k)}; M(t)), \quad (4.15)$$

we can conclude that the object does not see anything special at the horizon; in particular, it does not see a violation of the equivalence principle there. Of course, eq. (4.15) is not directly derived from the microscopic theory of quantum gravity, but it is a well motivated assumption based on the success of general relativity. In fact, we do not find any inconsistency in postulating it; in particular, we do not find a necessity of introducing a drastically new physical effect such as the firewall discussed in refs. [8, 71–73]. If we mine the black hole by a physical detector, a part of its backreaction is visible to a semiclassical observer. This effect, however, is confined in the causal future of the mining event, and is caused by the (non-gravitational) force supporting the detector; this is not the firewall phenomenon.

5 Discussions

In this paper, we have discussed relations between the microscopic structure of quantum gravity and what is called the semiclassical approximation. In particular, we have discussed what semiclassical field theory really is, and how its salient features — including the existence of the interior of a black hole — might arise from the fundamental theory of quantum gravity.

This emergence of the semiclassical description by no means implies that semiclassical field theory is always valid in the regime in which it is conventionally thought to be valid. As is well known, the global spacetime picture of general relativity contradicts the assumption that the final state of Hawking evaporation carries the complete information about the initial state [74], the assumption also suggested by gauge/gravity duality [75]. To prevent the resulting inconsistency, the fundamental (presumably covariant) theory of quantum gravity must possess certain nonlocality beyond cutting off the ultraviolet region of semiclassical field theory. To explore further aspects of quantum gravity, our discussions must be put in this larger context.

In this section, we expand our discussions in more general contexts in quantum gravity. We first present relatively straightforward extensions of our analyses to de Sitter and Minkowski spaces. We then discuss how our picture applies in understanding complementarity, both in black hole and more general spacetimes. Finally, we discuss how all these ingredients might help to understand cosmology, especially the issue of the beginning/end of the eternally inflating multiverse.

The discussions in this section become more conjectural as we move forward. This is especially so in the discussions of the last two subjects — complementarity and the multiverse — which are mostly based on the picture one of the authors (Y.N.) has been promoting in the past few years. Our hope is that they set a ground for, or provide useful intuition for, further explorations of quantum gravity, which may eventually lead to a “complete understanding” of the world we live in.

Simple extensions — de Sitter and Minkowski spaces. Our analyses can be extended relatively straightforwardly to de Sitter space. Consider de Sitter space with the Hubble radius α . In cosmology, de Sitter space appears as a meta-stable state in the middle of the evolution of the universe, and in this sense it is similar to spacetimes with a black hole. Indeed, string theory suggests that there is no absolutely stable de Sitter vacuum state in full quantum gravity — it must decay, at least, before the recurrence time [76]. This implies that what we call de Sitter space cannot be an eigenstate of energy (at least in this context).

Suppose we want to determine the time at which the de Sitter space is created or decays within an uncertainty of order $\delta t \approx \alpha$, where t is taken as the proper time for a static observer (see later for more discussions on this point). This implies the existence of uncertainty in the corresponding energy of order $\delta E \approx 1/\alpha$. Estimating this energy to be of order the vacuum energy ρ_Λ integrated over a Hubble volume, $E \approx O(\rho_\Lambda \alpha^3) \approx O(\alpha/l_P^2)$, this uncertainty is translated into

$$\delta\alpha \approx O\left(\frac{l_P^2}{\alpha}\right). \quad (5.1)$$

As in the case for black holes, we may consider the Gibbons-Hawking entropy [77]

$$S_{\text{dS}} = \frac{\mathcal{A}_{\text{dS}}}{4l_P^2} = \frac{\pi\alpha^2}{l_P^2}, \quad (5.2)$$

to be the entropy density for de Sitter vacuum states; namely, the number of de Sitter vacuum states with the Hubble radius between α and $\alpha + \delta\alpha$ is given by $\approx e^{S_{\text{dS}}}\delta\alpha/\alpha$. Here, $\mathcal{A}_{\text{dS}} = 4\pi\alpha^2$ is the area of the de Sitter horizon, and the expression in eq. (5.2) is understood to be valid at the leading order in expansion in powers of l_P/α . It is easy to see that the entropy density of eq. (5.2), in fact, does not depend on the precise choice of $\delta\alpha$ as long as $\delta\alpha/\alpha \gg e^{-S_{\text{dS}}}$.

Using static coordinates, appropriate for a static observer, the de Sitter horizon is located at the radius $r = \alpha$, and the stretched horizon is at $r_s = \alpha - O(l_P^2/\alpha)$, where we have identified the string (cutoff) scale, l_s , with the Planck scale, l_P . When a physical detector is located at $r = r_d$, it sees the blue-shifted local Gibbons-Hawking temperature

$$T_{\text{dS}}(r_d) = \frac{1}{2\pi\alpha} \frac{1}{\sqrt{1 - (r_d/\alpha)^2}}. \quad (5.3)$$

The interpretation of this phenomenon goes as in the black hole case. If the detector is held at a constant radius r_d which is close to the horizon, then most of the detector response arises because of the acceleration of the detector with respect to local inertial frames there. (This effect by itself would give a response with temperature $T_U(r_d) = r_d T_{\text{dS}}(r_d)/\alpha$.) The energy responsible for the response of the detector as well as the backreaction disturbing the vacuum comes from the external force needed to hold the detector at a constant $r = r_d$ for time interval of order $\Delta t \approx \alpha$, as measured at $r = 0$. This acceleration induced effect, however, does not fully explain eq. (5.3); in particular, it cannot make a detector located at $r = 0$ click as implied by eq. (5.3). As in the case of black holes, the remaining effect comes from spacetime curvature, which leads to the uncertainty of the local vacuum energy density of order $(1/2\pi\alpha)^4$ throughout the Hubble volume. It is this curvature induced effect that is responsible for the response of inertial detectors, including the detector located at $r = 0$.

If we denote the de Sitter vacuum states with the Hubble radius between α and $\alpha + \delta\alpha$ by $|\Psi_k(\alpha)\rangle$, the “vacuum index” k runs over

$$k = 1, \dots, e^{S_{\text{dS}}} = e^{\frac{\pi\alpha^2}{l_P^2}}. \quad (5.4)$$

As in the black hole case, most of this information will be in the region close to the horizon $r = \alpha$, although some of the information must spread over the entire Hubble volume. The expected distribution follows that of the thermal entropy calculated in semiclassical field theory using the local temperature in eq. (5.3). The existence of excitations in the interior or near exterior region is expected to provide only a small perturbation in terms of the entropy counting.

Given the absence of evidence otherwise, we could expect that the stretched horizon of de Sitter space behaves similarly to that of a black hole. When viewed from a static reference frame, a basis vector of the Hilbert space $\mathcal{H}_{\tilde{\Psi}_k}$, spanned by $|\Psi_k(\alpha)\rangle$ and its excited states for fixed k , will be specified by the state of the stretched horizon, \tilde{i} , as well as the number of excitations n_σ in each field theory mode σ inside the stretched horizon:

$$|\tilde{\Psi}_{k;\tilde{i};\{n_\sigma\}}(\alpha)\rangle = |\tilde{i}; k\rangle |\{n_\sigma\}; k\rangle. \quad (5.5)$$

Here, the meaning of the index k must be understood similarly to the black hole case in eq. (3.12); see the discussion in the paragraph containing eqs. (3.13)–(3.15). While it is not fully clear how the region outside the horizon is encoded in general in a given state, we may expect in analogy with the black hole case that some of the stretched horizon degrees of freedom are organized such that they can be viewed as the mirror modes for σ . The vacuum states will then take the form

$$|\Psi_k(\alpha)\rangle \approx \frac{1}{\sqrt{Z}} \sum_{\{n_\sigma\}} e^{-\frac{E_{\{n_\sigma\}}}{2T_{dS}}} |\{\tilde{n}_\sigma = n_\sigma\}; k\rangle |\{n_\sigma\}; k\rangle; \quad Z = \sum_{\{n_\sigma\}} e^{-\frac{E_{\{n_\sigma\}}}{T_{dS}}}, \quad (5.6)$$

in the near horizon region. Here, $T_{dS} = 1/2\pi\alpha$, and $E_{\{n_\sigma\}}$ is the energy of the state $|\{n_\sigma\}; k\rangle$ as defined at $r = 0$. This structure will ensure that the de Sitter horizon is smooth.

When an object hits the horizon in a static description, it can be thought of as going to space outside the horizon. The information about the object that goes outside will be stored in the state, consisting of the interior and the stretched horizon degrees of freedom (although there will be some maximum capacity for this information content). We expect that such information can be recovered later; otherwise, there is no Poincaré recurrence. This information recovery may not be necessarily in the form of Hawking radiation if the system evolves, for example, into Minkowski space or another de Sitter space with a smaller vacuum energy. Indeed, this is believed to have happened to density fluctuations generated in the early inflationary phase in our own universe [78–81].

The situation in Minkowski space is obtained by taking the limit $\alpha \rightarrow \infty$. In this limit, the “horizon” area becomes infinity, $\mathcal{A}_{\text{Min}} \rightarrow \infty$, so that the number of vacuum states also becomes infinity, $e^{S_{\text{Min}}} \rightarrow \infty$. Since the “horizon” is located at an infinite distance, however, a local observer cannot probe which vacuum he/she is in by directly interacting with the “horizon”. Also, while the information about the vacuum is to some extent delocalized, the effect the observer can probe locally away from the “horizon” in any finite time is expected to be suppressed exponentially in $\mathcal{A}_{\text{Min}}/l_P^2$ (and suppressed, at least, by powers of $\mathcal{A}_{\text{Min}}/l_P^2$), and hence negligible. The uniqueness of the Minkowski vacuum is thus recovered for the purpose of describing physics in any finite spacetime region. Note that this provides a way of arriving at the same conclusion in an inertial reference frame as the one reached at the end of section 4.2 using an accelerated reference frame.

Complementarity. A naive formulation of local quantum field theory on spacetime in general relativity allows us to use arbitrary spacelike (or null) hypersurfaces as the quantization surfaces, as long as spacetime curvature is sufficiently small everywhere on it. In black hole physics, this allows for a hypersurface that passes through both an object fallen inside the horizon as well as late Hawking quanta, contradicting the idea that final-state Hawking radiation contains the full information about the system that collapsed and fell into the black hole (since it leads to a cloning of quantum information) [74]. We expect that this problem is avoided in the fundamental theory of quantum gravity due to some nonlocality at large distances that becomes prominent when certain quantization surfaces are chosen.

The idea of black hole complementarity [5–7] is that we may quantize the system such that this nonlocality is “minimized”, and that there are multiple equivalent ways to do so. For example, by restricting the quantization surfaces appropriately to foliate only the exterior region, we may obtain a unitary description of physics in which the long-range nonlocality effect described above is absent. In addition, there are other “infalling” ways to quantize the system in which portions of the interior spacetime are included in the descriptions. Our discussions in this paper support this idea. In particular, we do not find any fundamental inconsistency between the existence of infalling descriptions and unitarity of the Hawking evaporation process as viewed from a distant reference frame.

While the general and precise formulation of complementarity is not yet known, a simple way to implement it seems to describe a system as viewed from a freely falling (local Lorentz) reference frame [59, 65, 82]. Specifically, we may consider a fixed reference point p_0 and a null or spacelike hypersurface associated with it, which in general is bounded by certain “horizons” signaling the breakdown of the local spacetime picture. It is important that the *procedure* to determine the hypersurface here is given independently of background spacetime; for example, it can be generated by past-directed light rays emanating from p_0 in all angular directions, with each light ray terminated when a certain condition on its expansion or other quantities is met. A quantum state that allows for a semiclassical spacetime interpretation is then given by specifying the configuration of physical degrees of freedom on the hypersurface using a coordinate system related to a local Lorentz frame elected at p_0 . In the fundamental theory of quantum gravity, restricting quantum states to this kind corresponds to partially fixing the gauge associated with spacetime transformations; after fixing the coordinates on the hypersurface, the remaining gauge freedom corresponds to the choice of the reference frame, whose generators we denote by H , P_i , $J_{[ij]}$, and K_i , borrowing the standard notation for the Poincaré algebra.

Let us denote the Hilbert space spanned by all the states obtained as described above by $\mathcal{H}_{\text{spacetime}}$. The relevant, partially gauge-fixed Hilbert space \mathcal{H} is then given by the direct sum of $\mathcal{H}_{\text{spacetime}}$ and another Hilbert space $\mathcal{H}_{\text{sing}}$ containing quantum states that do not allow for a spacetime interpretation (which become relevant when p_0 hits a spacetime singularity):¹²

$$\mathcal{H} = \mathcal{H}_{\text{spacetime}} \oplus \mathcal{H}_{\text{sing}}. \quad (5.7)$$

Describing a system from a fixed reference frame corresponds to further fixing the gauge for the transformations associated with P_i , $J_{[ij]}$, and K_i by taking an appropriate section of the corresponding gauge orbit in \mathcal{H} . A quantum state representing the system at a fixed time is then given by an element in \mathcal{H} , which may involve a superposition of macroscopically different spacetimes or even a superposition of spacetime and singularity states. Its evolution is given by the (effective) Hamiltonian H_{QG} , generating a translation of the proper time τ measured at p_0 . (This transformation corresponds to the remaining gauge freedom represented by generator H . The emergence of the effective time evolution picture from a more fundamental description in quantum gravity will be discussed below.)

¹²Here, we have changed the notation from that in ref. [59]. The Hilbert spaces denoted by \mathcal{H} and $\mathcal{H}_{\text{spacetime}}$ here correspond to \mathcal{H}_{QG} and \mathcal{H} in ref. [59], respectively.

In this framework, a complementarity transformation is simply choosing a different local Lorentz frame at some time τ , which corresponds to choosing a different section in the orbit of the gauge transformations generated by P_i and K_i (and $J_{[ij]}$, which however yields only a trivial rotation of the frame). This framework was applied in particular to black hole physics in ref. [49], in which it was shown how the interior spacetime may appear effectively by combining various descriptions in different reference frames.

We may take a basis in $\mathcal{H}_{\text{spacetime}}$ in which a basis element represents a hypersurface that contains p_0 and is surrounded by the boundary (horizon) that indicates the breakdown of the semiclassical spacetime picture. For example, if p_0 is located outside the Schwarzschild horizon of a black hole in de Sitter space, then p_0 will “see” the black hole horizon in some directions and the de Sitter horizon in the others; specifically, if the hypersurface is generated by past-directed light rays emanating from p_0 , then each light ray hits either the (stretched) black hole or de Sitter horizon, depending on the angular direction in which it is emitted. A general criterion for determining the location of the boundary is not fully understood, but a possible procedure was described in ref. [59] which applies to certain simple cases. Now, consider a set of quantum states that share the “same boundary” $\partial\mathcal{M}$ in some appropriate sense (although the precise general definition of this statement is also not yet available). We may then expect, based on the conjectured entropy bound in refs. [53, 54], that the dimension of the Hilbert space $\mathcal{H}_{\partial\mathcal{M}}$ spanned by all these states is given by

$$\ln \dim \mathcal{H}_{\partial\mathcal{M}} \approx \frac{\mathcal{A}_{\partial\mathcal{M}}}{4l_P^2}, \quad (5.8)$$

where $\mathcal{A}_{\partial\mathcal{M}}$ is the total area of the boundary; in the above example, $\mathcal{A}_{\partial\mathcal{M}} = \mathcal{A}_{\text{BH}} + \mathcal{A}_{\text{dS}}$ where \mathcal{A}_{BH} and \mathcal{A}_{dS} are the areas of the black hole and de Sitter portions of the boundary. In our present context, $\dim \mathcal{H}_{\partial\mathcal{M}}$ represents the number of independent microstates that must be coarse-grained to obtain the semiclassical description on a hypersurface bounded by $\partial\mathcal{M}$.

The static quantum multiverse. By describing a system from a fixed reference frame, we are fixing all the relevant gauge redundancies except for that associated with H : translation of the proper time τ at p_0 . In the most fundamental treatment of quantum gravity, this residual gauge redundancy must also be fixed, which we may do by imposing the constraint on an element $|\Phi\rangle$ of \mathcal{H} in eq. (5.7) (equivalent to taking an appropriate section in the corresponding gauge orbit):

$$H|\Phi\rangle = 0. \quad (5.9)$$

The effective time evolution picture may then arise as correlations between physical configurations of subsystems [83]. Specifically, in $|\Phi\rangle$ we can identify a (small) subsystem as the “clock” degrees of freedom, and reinterpret the entanglement of these degrees of freedom — represented e.g. by a set of states $|i\rangle$ — with the rest of the degrees of freedom — represented e.g. by a set of states $|\Psi_i\rangle$ — as the time evolution of a state $|\Psi_i\rangle$ with i playing the role of time.

We emphasize that $|\Phi\rangle$ in eq. (5.9) is the quantum state for the *entire* system, including all the degrees of freedom existing in the world. If we omit any, even small, subsystem from

our description, then the resulting state need not satisfy eq. (5.9) (hence making the time evolution picture possible). In ref. [84], it was conjectured that the condition of eq. (5.9) together with the normalizability of the state in \mathcal{H} , $\langle \Phi | \Phi \rangle < \infty$, is enough to select the *state of the multiverse*.¹³ In this picture, practically all the systems we observe, including the whole universe of our own, are tiny portions of the branches of $|\Phi\rangle$, and their “evolutions” are well described by

$$|\Psi(\tau)\rangle \rightarrow |\Psi(\tau + \Delta\tau)\rangle = e^{-iH\Delta\tau}|\Psi(\tau)\rangle, \quad (5.10)$$

using the proper time τ at p_0 . This is the time evolution of a state we have been discussing so far — it is a parametric representation of correlations between configurations of subsystems, analogous to a parametric representation of a curve on a plane, $(x(\tau), y(\tau))$. Note that this does not contradict the fact that the state of the entire multiverse is static: $e^{-iH\Delta\tau}|\Phi\rangle = |\Phi\rangle$.

If the state of the multiverse is indeed selected as described above, then we may, in principle, predict everything we are allowed to predict once the rule of extracting it is given (assuming that the explicit form of relevant operators, such as H , and their algebra is known). This rule must reduce to the standard Born rule in appropriate circumstances, but it will involve ingredients beyond that because the state of the multiverse $|\Phi\rangle$ does not have any environment to interact with, whose existence is implicitly assumed by the standard Born rule (e.g. in determining the basis for projections). In fact, the normalizability of $|\Phi\rangle$ would mean that the number of components in $|\Phi\rangle$, when expanded in a basis in which locality is manifest, is effectively finite, and the rule would have to be formulated in a way that it works internally in the finite-dimensional Hilbert space spanned by these components. Since we cannot use an external environment to select a measurement basis, such a rule will have to involve the structure of operators, rather than just a state. In our view, this may very well comprise one of the most important challenges to obtain a complete theory of quantum gravity, about which the current formulation of string theory has little to say.

Acknowledgments

We thank Raphael Bousso, Ben Freivogel, Daniel Harlow, Juan Maldacena, Joseph Polchinski, Jaime Varela, Herman Verlinde, and I-Sheng Yang for various conversations during our exploration of this subject. This work was supported in part by the Director, Office of Science, Office of High Energy and Nuclear Physics, of the U.S. Department of Energy under Contract DE-AC02-05CH11231, and in part by the National Science Foundation under grant PHY-1214644.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](#)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

¹³If the state is not uniquely selected, but multiple discrete states $|\Phi_n\rangle$ ($n = 1, \dots, N$) are allowed, then we may expect that the state of the multiverse is mixed: $\sum_{n=1}^N |\Phi_n\rangle \langle \Phi_n|/N$. Our argument below still applies in this case with appropriate modifications.

References

- [1] J.D. Bekenstein, *Black holes and entropy*, *Phys. Rev. D* **7** (1973) 2333 [[INSPIRE](#)].
- [2] S.W. Hawking, *Particle creation by black holes*, *Commun. Math. Phys.* **43** (1975) 199 [*Erratum ibid.* **46** (1976) 206] [[INSPIRE](#)].
- [3] G. 't Hooft, *The black hole interpretation of string theory*, *Nucl. Phys. B* **335** (1990) 138 [[INSPIRE](#)].
- [4] C.R. Stephens, G. 't Hooft and B.F. Whiting, *Black hole evaporation without information loss*, *Class. Quant. Grav.* **11** (1994) 621 [[gr-qc/9310006](#)] [[INSPIRE](#)].
- [5] L. Susskind, L. Thorlacius and J. Uglum, *The stretched horizon and black hole complementarity*, *Phys. Rev. D* **48** (1993) 3743 [[hep-th/9306069](#)] [[INSPIRE](#)].
- [6] L. Susskind and L. Thorlacius, *Gedanken experiments involving black holes*, *Phys. Rev. D* **49** (1994) 966 [[hep-th/9308100](#)] [[INSPIRE](#)].
- [7] P. Hayden and J. Preskill, *Black holes as mirrors: quantum information in random subsystems*, *JHEP* **09** (2007) 120 [[arXiv:0708.4025](#)] [[INSPIRE](#)].
- [8] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black holes: complementarity or firewalls?*, *JHEP* **02** (2013) 062 [[arXiv:1207.3123](#)] [[INSPIRE](#)].
- [9] S.W. Hawking, *Breakdown of predictability in gravitational collapse*, *Phys. Rev. D* **14** (1976) 2460 [[INSPIRE](#)].
- [10] W.G. Unruh, *Notes on black hole evaporation*, *Phys. Rev. D* **14** (1976) 870 [[INSPIRE](#)].
- [11] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford and J. Sully, *An apologia for firewalls*, *JHEP* **09** (2013) 018 [[arXiv:1304.6483](#)] [[INSPIRE](#)].
- [12] D. Marolf and J. Polchinski, *Gauge/gravity duality and the black hole interior*, *Phys. Rev. Lett.* **111** (2013) 171301 [[arXiv:1307.4706](#)] [[INSPIRE](#)].
- [13] Y. Nomura and J. Varela, *A note on (no) firewalls: the entropy argument*, *JHEP* **07** (2013) 124 [[arXiv:1211.7033](#)] [[INSPIRE](#)].
- [14] Y. Nomura, J. Varela and S.J. Weinberg, *Black holes or firewalls: a theory of horizons*, *Phys. Rev. D* **88** (2013) 084052 [[arXiv:1308.4121](#)] [[INSPIRE](#)].
- [15] Y. Nomura and S.J. Weinberg, *The entropy of a vacuum: what does the covariant entropy count?*, [arXiv:1310.7564](#) [[INSPIRE](#)].
- [16] S.B. Giddings, *Black holes, quantum information and unitary evolution*, *Phys. Rev. D* **85** (2012) 124063 [[arXiv:1201.1037](#)] [[INSPIRE](#)].
- [17] S.B. Giddings, *Nonviolent nonlocality*, *Phys. Rev. D* **88** (2013) 064023 [[arXiv:1211.7070](#)] [[INSPIRE](#)].
- [18] K. Papadodimas and S. Raju, *An infalling observer in AdS/CFT*, *JHEP* **10** (2013) 212 [[arXiv:1211.6767](#)] [[INSPIRE](#)].
- [19] K. Papadodimas and S. Raju, *Black hole interior in the holographic correspondence and the information paradox*, *Phys. Rev. Lett.* **112** (2014) 051301 [[arXiv:1310.6334](#)] [[INSPIRE](#)].
- [20] K. Papadodimas and S. Raju, *State-dependent bulk-boundary maps and black hole complementarity*, *Phys. Rev. D* **89** (2014) 086010 [[arXiv:1310.6335](#)] [[INSPIRE](#)].

- [21] E. Verlinde and H. Verlinde, *Black hole entanglement and quantum error correction*, *JHEP* **10** (2013) 107 [[arXiv:1211.6913](#)] [[INSPIRE](#)].
- [22] E. Verlinde and H. Verlinde, *Passing through the firewall*, [arXiv:1306.0515](#) [[INSPIRE](#)].
- [23] E. Verlinde and H. Verlinde, *Behind the horizon in AdS/CFT*, [arXiv:1311.1137](#) [[INSPIRE](#)].
- [24] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortschr. Phys.* **61** (2013) 781 [[arXiv:1306.0533](#)] [[INSPIRE](#)].
- [25] R. Bousso, *Complementarity is not enough*, *Phys. Rev. D* **87** (2013) 124023 [[arXiv:1207.5192](#)] [[INSPIRE](#)].
- [26] Y. Nomura, J. Varela and S.J. Weinberg, *Complementarity endures: no firewall for an infalling observer*, *JHEP* **03** (2013) 059 [[arXiv:1207.6626](#)] [[INSPIRE](#)].
- [27] S.D. Mathur and D. Turton, *Comments on black holes I: the possibility of complementarity*, *JHEP* **01** (2014) 034 [[arXiv:1208.2005](#)] [[INSPIRE](#)].
- [28] S.D. Mathur and D. Turton, *The flaw in the firewall argument*, *Nucl. Phys. B* **884** (2014) 566 [[arXiv:1306.5488](#)] [[INSPIRE](#)].
- [29] B.D. Chowdhury and A. Puhm, *Is Alice burning or fuzzing?*, *Phys. Rev. D* **88** (2013) 063509 [[arXiv:1208.2026](#)] [[INSPIRE](#)].
- [30] T. Banks and W. Fischler, *Holographic space-time does not predict firewalls*, [arXiv:1208.4757](#) [[INSPIRE](#)].
- [31] T. Banks and W. Fischler, *No firewalls in holographic space-time or matrix theory*, [arXiv:1305.3923](#) [[INSPIRE](#)].
- [32] R. Brustein, *Origin of the blackhole information paradox*, *Fortschr. Phys.* **62** (2014) 255 [[arXiv:1209.2686](#)] [[INSPIRE](#)].
- [33] L. Susskind, *The transfer of entanglement: the case for firewalls*, [arXiv:1210.2098](#) [[INSPIRE](#)].
- [34] L. Susskind, *Butterflies on the stretched horizon*, [arXiv:1311.7379](#) [[INSPIRE](#)].
- [35] L. Susskind, *Computational complexity and black hole horizons*, [arXiv:1402.5674](#) [[INSPIRE](#)].
- [36] S.G. Avery, B.D. Chowdhury and A. Puhm, *Unitarity and fuzzball complementarity: “Alice fuzzes but may not even know it!”*, *JHEP* **09** (2013) 012 [[arXiv:1210.6996](#)] [[INSPIRE](#)].
- [37] D. Harlow and P. Hayden, *Quantum computation vs. firewalls*, *JHEP* **06** (2013) 085 [[arXiv:1301.4504](#)] [[INSPIRE](#)].
- [38] R. Brustein and A.J.M. Medved, *Restoring predictability in semiclassical gravitational collapse*, *JHEP* **09** (2013) 015 [[arXiv:1305.3139](#)] [[INSPIRE](#)].
- [39] R. Brustein and A.J.M. Medved, *Phases of information release during black hole evaporation*, *JHEP* **02** (2014) 116 [[arXiv:1310.5861](#)] [[INSPIRE](#)].
- [40] R. Brustein and A.J.M. Medved, *Horizons of semiclassical black holes are cold*, [arXiv:1312.0880](#) [[INSPIRE](#)].
- [41] D.A. Lowe and L. Thorlacius, *Pure states and black hole complementarity*, *Phys. Rev. D* **88** (2013) 044012 [[arXiv:1305.7459](#)] [[INSPIRE](#)].
- [42] D.A. Lowe and L. Thorlacius, *Black hole complementarity: the inside view*, *Phys. Lett. B* **737** (2014) 320 [[arXiv:1402.4545](#)] [[INSPIRE](#)].

- [43] D.N. Page, *Excluding black hole firewalls with extreme cosmic censorship*, *JCAP* **06** (2014) 051 [[arXiv:1306.0562](#)] [[INSPIRE](#)].
- [44] M. Van Raamsdonk, *Evaporating firewalls*, [arXiv:1307.1796](#) [[INSPIRE](#)].
- [45] S. Lloyd and J. Preskill, *Unitarity of black hole evaporation in final-state projection models*, *JHEP* **08** (2014) 126 [[arXiv:1308.4209](#)] [[INSPIRE](#)].
- [46] S.D.H. Hsu, *Factorization of unitarity and black hole firewalls*, [arXiv:1308.5686](#) [[INSPIRE](#)].
- [47] I. Ilgin and I.-S. Yang, *Causal patch complementarity: the inside story for old black holes*, *Phys. Rev. D* **89** (2014) 044007 [[arXiv:1311.1219](#)] [[INSPIRE](#)].
- [48] B. Freivogel, *Energy and information near black hole horizons*, *JCAP* **07** (2014) 041 [[arXiv:1401.5340](#)] [[INSPIRE](#)].
- [49] Y. Nomura, J. Varela and S.J. Weinberg, *Black holes, information and Hilbert space for quantum gravity*, *Phys. Rev. D* **87** (2013) 084050 [[arXiv:1210.6348](#)] [[INSPIRE](#)].
- [50] D.N. Page, *Is black hole evaporation predictable?*, *Phys. Rev. Lett.* **44** (1980) 301 [[INSPIRE](#)].
- [51] G. 't Hooft, *Dimensional reduction in quantum gravity*, [gr-qc/9310026](#) [[INSPIRE](#)].
- [52] L. Susskind, *The world as a hologram*, *J. Math. Phys.* **36** (1995) 6377 [[hep-th/9409089](#)] [[INSPIRE](#)].
- [53] R. Bousso, *A covariant entropy conjecture*, *JHEP* **07** (1999) 004 [[hep-th/9905177](#)] [[INSPIRE](#)].
- [54] R. Bousso, *The holographic principle*, *Rev. Mod. Phys.* **74** (2002) 825 [[hep-th/0203101](#)] [[INSPIRE](#)].
- [55] Y. Nomura, *A theory of horizons*, talk given at KITP Rapid Response Workshop. *Black Holes: Complementarity, Fuzz, or Fire?*, KITP, Santa Barbara U.S.A., 19–30 Aug 2013.
- [56] D.N. Page, *Particle emission rates from a black hole: massless particles from an uncharged, nonrotating hole*, *Phys. Rev. D* **13** (1976) 198 [[INSPIRE](#)].
- [57] W. Israel, *Thermo field dynamics of black holes*, *Phys. Lett. A* **57** (1976) 107 [[INSPIRE](#)].
- [58] R.D. Sorkin, *Diffusion, differential geometry, and black holes*, in Proceedings of the 10th International Conference on General Relativity and Gravitation, Padova Italy, 4–9 Jul 1983, B. Bertotti, F. de Felice and A. Pascolini eds., volume II, pp. 737–739.
- [59] Y. Nomura, J. Varela and S.J. Weinberg, *Low energy description of quantum gravity and complementarity*, *Phys. Lett. B* **733** (2014) 126 [[arXiv:1304.0448](#)] [[INSPIRE](#)].
- [60] L. Susskind and J. Uglum, *Black hole entropy in canonical quantum gravity and superstring theory*, *Phys. Rev. D* **50** (1994) 2700 [[hep-th/9401070](#)] [[INSPIRE](#)].
- [61] J.H. Cooperman and M.A. Luty, *Renormalization of entanglement entropy and the gravitational effective action*, [arXiv:1302.1878](#) [[INSPIRE](#)].
- [62] W.G. Unruh and R.M. Wald, *What happens when an accelerating observer detects a Rindler particle*, *Phys. Rev. D* **29** (1984) 1047 [[INSPIRE](#)].
- [63] W.G. Unruh and R.M. Wald, *Acceleration radiation and generalized second law of thermodynamics*, *Phys. Rev. D* **25** (1982) 942 [[INSPIRE](#)].
- [64] A.R. Brown, *Tensile strength and the mining of black holes*, *Phys. Rev. Lett.* **111** (2013) 211301 [[arXiv:1207.3342](#)] [[INSPIRE](#)].

- [65] Y. Nomura, *Quantum mechanics, spacetime locality and gravity*, *Found. Phys.* **43** (2013) 978 [[arXiv:1110.4630](#)] [[INSPIRE](#)].
- [66] S.M. Christensen and S.A. Fulling, *Trace anomalies and the Hawking effect*, *Phys. Rev. D* **15** (1977) 2088 [[INSPIRE](#)].
- [67] P. Candelas, *Vacuum polarization in Schwarzschild space-time*, *Phys. Rev. D* **21** (1980) 2185 [[INSPIRE](#)].
- [68] J.M. Bardeen, *Black holes do evaporate thermally*, *Phys. Rev. Lett.* **46** (1981) 382 [[INSPIRE](#)].
- [69] R. Balbinot, *Hawking radiation and the back reaction — a first approach*, *Class. Quant. Grav.* **1** (1984) 573.
- [70] D.N. Page, *Information in black hole radiation*, *Phys. Rev. Lett.* **71** (1993) 3743 [[hep-th/9306083](#)] [[INSPIRE](#)].
- [71] S.L. Braunstein, S. Pirandola and K. Życzkowski, *Better late than never: information retrieval from black holes*, *Phys. Rev. Lett.* **110** (2013) 101301 [[arXiv:0907.1190](#)] [[INSPIRE](#)].
- [72] S.D. Mathur, *The information paradox: a pedagogical introduction*, *Class. Quant. Grav.* **26** (2009) 224001 [[arXiv:0909.1038](#)] [[INSPIRE](#)].
- [73] S.B. Giddings, *Nonlocality versus complementarity: a conservative approach to the information problem*, *Class. Quant. Grav.* **28** (2011) 025002 [[arXiv:0911.3395](#)] [[INSPIRE](#)].
- [74] J. Preskill, *Do black holes destroy information?*, in *Blackholes, membranes, wormholes and superstrings*, S. Kalara and D.V. Nanopoulos eds., World Scientific, Singapore (1993), pg. 22 [[hep-th/9209058](#)] [[INSPIRE](#)].
- [75] J.M. Maldacena, *The large- N limit of superconformal field theories and supergravity*, *Adv. Theor. Math. Phys.* **2** (1998) 231 [*Int. J. Theor. Phys.* **38** (1999) 1113] [[hep-th/9711200](#)] [[INSPIRE](#)].
- [76] S. Kachru, R. Kallosh, A.D. Linde and S.P. Trivedi, *de Sitter vacua in string theory*, *Phys. Rev. D* **68** (2003) 046005 [[hep-th/0301240](#)] [[INSPIRE](#)].
- [77] G.W. Gibbons and S.W. Hawking, *Cosmological event horizons, thermodynamics and particle creation*, *Phys. Rev. D* **15** (1977) 2738 [[INSPIRE](#)].
- [78] S.W. Hawking, *The development of irregularities in a single bubble inflationary universe*, *Phys. Lett. B* **115** (1982) 295 [[INSPIRE](#)].
- [79] A.A. Starobinsky, *Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations*, *Phys. Lett. B* **117** (1982) 175 [[INSPIRE](#)].
- [80] A.H. Guth and S.Y. Pi, *Fluctuations in the new inflationary universe*, *Phys. Rev. Lett.* **49** (1982) 1110 [[INSPIRE](#)].
- [81] V.F. Mukhanov and G.V. Chibisov, *Quantum fluctuation and nonsingular universe* (in Russian), *JETP Lett.* **33** (1981) 532 [*Pisma Zh. Eksp. Teor. Fiz.* **33** (1981) 549] [[INSPIRE](#)].
- [82] Y. Nomura, *Physical theories, eternal inflation and quantum universe*, *JHEP* **11** (2011) 063 [[arXiv:1104.2324](#)] [[INSPIRE](#)].
- [83] B.S. DeWitt, *Quantum theory of gravity. I. The canonical theory*, *Phys. Rev.* **160** (1967) 1113 [[INSPIRE](#)].
- [84] Y. Nomura, *The static quantum multiverse*, *Phys. Rev. D* **86** (2012) 083505 [[arXiv:1205.5550](#)] [[INSPIRE](#)].