

**Quantization in black hole backgrounds**

Steven B. Giddings\*

*Department of Physics University of California Santa Barbara, California 93106-9530, USA*

(Received 5 April 2007; published 24 September 2007)

Quantum field theory in a semiclassical background can be derived as an approximation to quantum gravity from a weak-coupling expansion in the inverse Planck mass. Such an expansion is studied for evolution on “nice slices” in the spacetime describing a black hole of mass  $M$ . Arguments for a breakdown of this expansion are presented, due to significant gravitational coupling between fluctuations, which is consistent with the statement that existing calculations of information loss in black holes are not reliable. For a given fluctuation, the coupling to subsequent fluctuations becomes of order unity by a time of order  $M^3$ . Lack of a systematic derivation of the weakly coupled/semiclassical approximation would indicate a role for the nonperturbative dynamics of gravity, and possibly for the proposal that such dynamics has an essentially nonlocal quality.

DOI: [10.1103/PhysRevD.76.064027](https://doi.org/10.1103/PhysRevD.76.064027)

PACS numbers: 04.70.Dy, 04.60.-m

**I. INTRODUCTION**

Hawking’s discovery of black hole radiance [1] has produced a paradox that may be as important to finding a quantum description of gravity as the paradox of the classical instability of matter was in the foundation of quantum mechanics. There is no commonly accepted explanation for what is wrong with Hawking’s original argument that black holes destroy information [2]. This is despite widespread belief that black holes respect unitary quantum evolution, which is now shared by the originator of the paradox himself [3].

From the beginning it has seemed that the challenge of Hawking radiation requires a major conceptual advance. Any proposal regarding the fate of information apparently conflicts with a cherished principle of physics.<sup>1</sup> Hawking’s original proposal of information loss violates unitary quantum evolution, and closer studies of the resulting dynamics [6,7] argue it leads to drastic violations of energy conservation. Alternatively, unitary evolution might be preserved if information escapes after Hawking’s calculation breaks down in the Planck phase of evaporation. But basic quantum restrictions on how rapidly information can be transmitted with a given energy then apparently imply a long-lived remnant. Remnants seem equally disastrous; to encode the information from an arbitrarily large black hole, they are expected to come in infinite species, and correspondingly they would be infinitely produced in any process with sufficient total energy. These outcomes suggest that information must escape early in black hole evaporation, but Hawking’s argument to the contrary apparently requires only very basic assumptions such as locality.

Attempts at a resolution of this paradox have increasingly focussed on some breakdown of locality [8–10], and this possibility is also strongly suggested by aspects of string theory such as the anti-de Sitter/conformal field

theory (AdS/CFT) correspondence [11]. However, what has not been clear in much of this discussion is the mechanism for a breakdown of locality that obviates Hawking’s argument.

The apparently fundamental nature of the paradox suggests comparison with the paradoxes that arose in the transition from classical to quantum physics, and the instability of the classical atomic model seems particularly apt. In a gedanken history, were it not for the fact that the stability of matter is so critical, classical physicists might have reasoned that the problem lay in the singularity in classical evolution, when the electron reaches the charge center, and might have tried various methods to regulate this singularity. History shows us this is not the correct approach; indeed, the problem is that classical physics simply ceases to be a valid description long before a classical electron reaches the nucleus. Classical physics does not exhibit a *breakdown*, but rather quantum physics *supplants* it at scales set by the Bohr radius.

This suggests a possible way that the black hole paradox could be resolved: a local quantum field theory description simply ceases to be a good description in the vicinity of the Schwarzschild radius, long before it breaks down near the singularity. Like in the case of the classical atom, we may not see the failure in the present physical framework. What is of course missing is the more complete framework within which such a statement could be properly understood.

Another different alternative is that local quantum field theory actually does signal its limitations through some explicit failure, following a slightly different path than that from classical to quantum. It is worth pursuing this possibility as far as possible, in an effort to find any relevant clues to the problem.

This paper will investigate the question of the validity of Hawking’s argument for loss of information [1,2]. With more recent refinements, the argument goes as follows. First, one can draw a family of smooth spacelike slices through the interior of a black hole of large mass  $M$ ,

\*Electronic address: [giddings@physics.ucsb.edu](mailto:giddings@physics.ucsb.edu)<sup>1</sup>For more in depth reviews, see [4,5].

avoiding the region of strong curvature, and asymptoting to usual time slices. Away from the singularity, the geometry does not fluctuate, so the evolution on these slices can be described in terms of local quantum field theory on a semiclassical background. This evolution produces states on late-time slices that contain significant quantum entanglement between interior and exterior. Tracing over internal states gives a density matrix that describes physics outside, and the entanglement implies that this density matrix has large entropy, of order  $M^2$ . This parametrizes missing information assumed to vanish with the black hole.

We begin by describing in general the steps that can be used to justify the approximation of quantum field theory in a semiclassical background; this arises from a weak-coupling expansion in the inverse Planck mass  $M_p$ . This approximation is described both from the functional and canonical perspectives, and then a more complete argument for information loss is given. Section III then carefully examines the validity of this expansion in the specific situation of evolution on the slices described above, called “nice slices” in [12]. Such a careful treatment reveals an apparent flaw in this weak-coupling expansion, arising from strong coupling between fluctuations at early and late times. This effect is uniquely gravitational. Section IV then examines this effect more carefully, discussing the diagrammatic expansion, the connection to large relative boosts in the black hole geometry, possible lessons for locality, possible extensions to inflationary cosmology, and other issues. Concluding remarks appear in Sec. V.

This paper is thus consistent with the second possibility above, that the semiclassical approximation breaks down due to strong gravitational effects.<sup>2</sup> This has been previously advocated in [16–21].<sup>3</sup> If indeed there is no reliable calculation of the information destroyed by a black hole, there is no information paradox. This paper improves on the previous arguments, accounting for the effects of fluctuations on the nice-slice states used to calculate the density matrix and thus its entropy. Reference [12] attempted to produce a related conclusion due to effects of long strings, but failed to find an unambiguous effect [22]; moreover, [20] argues that in high-energy collisions (which are closely related to black hole evolution) there is no apparent evidence for the relevant such long-string effects, but rather strong gravitational dynamics plays a central role.

## II. PERTURBATIVE GRAVITY AND THE SEMICLASSICAL DESCRIPTION

A good starting point is a description of the quantization of gravity, reviewing its perturbative quantization either in

path integral or canonical terms. Let us particularly focus on the origin of the semiclassical approximation. This has been previously described in [23–25] as derived in a WKB approach, but we will find that in a slightly different approach it can be derived in a weak-coupling expansion about a semiclassical spacetime. We will find that this approximation indeed yields dynamics described as quantum field theory in a curved background, and thus demonstrate how Hawking’s argument for information loss arises.

### A. Path-integral quantization

A concrete approach to quantization is the path integral. Consider a theory with gravity coupled to matter, which will for simplicity be taken to be a single massless scalar field, with Lagrangian

$$S[g, \phi] = -\frac{1}{2} \int d^4x \sqrt{-g} (\nabla \phi)^2. \quad (2.1)$$

Also for simplicity we work in four dimensions, although the generalization to higher dimensions is straightforward. The gravitational action is taken to be

$$S[g] = \frac{M_p^2}{2} \int d^4x \sqrt{-g} R + M_p^2 \oint \sqrt{|^3g|} K, \quad (2.2)$$

which includes the well-known extrinsic-curvature surface term, and where  $M_p$  is the Planck mass, defined as  $M_p^{-2} = 8\pi G$ .

In order to have a well-defined time, we will work with such amplitudes defined over geometries that are asymptotically flat. Consider, in particular, a geometry described in terms of a family of spatial slices, with coordinates  $x^i$ , labeled by a time variable  $t$ . The Arnowitt-Deser-Misner (ADM) decomposition of the metric is

$$ds^2 = -(N^2 - N_i N^i) dt^2 + 2N_i dt dx^i + g_{ij} dx^i dx^j, \quad (2.3)$$

where  $N$  is the lapse,  $N^i$  is the shift, and spatial (latin) indices are raised and lowered with the spatial metric  $g_{ij}$  which describes the intrinsic geometry of the slices. For asymptotically flat geometries, the asymptotic time is

$$T = \int dt N(\infty). \quad (2.4)$$

We are interested in amplitudes for transitions between slices with definite asymptotic time difference. These can be defined as

$$\begin{aligned} \mathcal{A}[\Psi_i, \Psi_f, T] &= \int_{\Psi_i}^{\Psi_f} \mathcal{D}g \mathcal{D}\phi e^{iS[g] + iS[g, \phi]} \\ &\times \delta\left(T - \int dt N(\infty)\right). \end{aligned} \quad (2.5)$$

In this expression, the state is expressed in terms of data on the metric and field and their momenta. For example, in the field representation, one might specify the metric  $g_{ij}(x)$  and the field  $\phi(x)$  on the initial and final slices.

<sup>2</sup>For other potential sources of breakdown, see [13–15].

<sup>3</sup>This issue has also been explored in unpublished work by N. Arkani-Hamed.

A standard approach is to expand the metric as a fluctuating field about some background metric,

$$g_{\mu\nu} = g_{0\mu\nu} + M_p^{-1} h_{\mu\nu}. \quad (2.6)$$

For example, in such an expansion the ground-state wave function has been derived in the quadratic approximation in [26]. To do so, one rewrites (2.5) in terms of an integral over the perturbations  $h_{\mu\nu}$ , and gauge fixes. The expansion of the action (2.2) in powers of  $M_p^{-1}$  is of the form

$$\begin{aligned} S[g_0 + h] &= S[g_0] + \int d^4x \sqrt{-g_0} \left[ M_p^{-1} h \delta_g S[g_0] \right. \\ &\quad \left. - \frac{1}{2} (Dh)^2 + M_p^{-1} h (\nabla h)^2 + \dots \right] \\ &= \sum_n S_n^g. \end{aligned} \quad (2.7)$$

Here  $(Dh)^2$  is the kinetic term for  $h$  (integration of  $(Dh)^2$  by parts gives the Lichnerowicz Laplacian), and this and higher terms can be straightforwardly worked out in detail. One can likewise expand  $S[g, \phi]$  in an expansion in  $h$  and  $\phi$ . For simplicity we assume that the classical background for  $\phi$  vanishes (although this could easily be generalized), and as a result

$$\begin{aligned} S[g_0 + h, \phi] &= \int d^4x \sqrt{-g_0} \left\{ -\frac{1}{2} (\nabla^0 \phi)^2 - \frac{1}{2} M_p^{-1} h^{\mu\nu} T_{\mu\nu}^\phi \right. \\ &\quad \left. + M_p^{-2} \mathcal{O}[h^2(\partial\phi)^2] \right\} = \sum_n S_n^\phi, \end{aligned} \quad (2.8)$$

where the leading term is  $n = 2$ , to match (2.7).

A well-defined amplitude requires gauge fixing the diffeomorphism symmetry

$$\begin{aligned} g_{\mu\nu} &\rightarrow g_{\mu\nu} + M_p^{-1} \mathcal{L}_\xi g_{\mu\nu} \\ &= g_{\mu\nu} + M_p^{-1} \mathcal{L}_\xi g_{0\mu\nu} + M_p^{-2} \mathcal{L}_\xi h_{\mu\nu}, \end{aligned} \quad (2.9)$$

whose linearized version is thus

$$\delta h_{\mu\nu} = 2\nabla_{(\mu}^0 \xi_{\nu)}. \quad (2.10)$$

This is accomplished by applying a suitable gauge condition; one useful example at the linear level and in vacuo is Hilbert-DeDonder gauge, defined via

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} g_{0\mu\nu} h \quad (2.11)$$

as

$$\nabla_0^\mu \bar{h}_{\mu\nu} = 0. \quad (2.12)$$

Such a gauge has a residual set of symmetries, corresponding to diffeomorphisms satisfying

$$\nabla_0^2 \xi_\mu = 0. \quad (2.13)$$

These are fixed by also specifying conditions on  $h$  on the initial/final surfaces, resulting in the expected  $10 - 4 -$

$4 = 2$  degrees of freedom; these slice conditions can alternately be thought of as a choice of the initial/final slices.

Denote the general such combined gauge conditions by

$$C^I(h) = 0; \quad (2.14)$$

these are enforced by inserting the expression

$$\Delta[C^I(h)] \det(\delta C^I) \quad (2.15)$$

into the path integral, where  $\Delta$  is a functional delta function, and  $\det(\delta C^I)$  is the corresponding Fadeev-Popov determinant.

The resulting amplitude takes the form

$$\begin{aligned} \mathcal{A}[\Psi_i, \Psi_f, ; T] &= e^{iS[g_0]} \int_{\Psi_i}^{\Psi_f} \mathcal{D}_{g_0+h} h \mathcal{D}_{g_0+h} \phi \Delta[C^I(h)] \\ &\quad \times \det(\delta C^I) e^{iM_p^{-1} \int d^4x \sqrt{-g_0} h \delta_g S[g_0]} \\ &\quad \times e^{i(S_2 + S_3 + \dots)}. \end{aligned} \quad (2.16)$$

Here, as is explicitly indicated, the integration measures depend on the perturbation, as does the gauge-fixing measure, and we define  $S_n = S_n^g + S_n^\phi$ . The initial and final states are specified by giving wave functionals of the perturbations  ${}^3h$  of the three-geometry and fields on the initial/final slices. In the case with  $g_{0\mu\nu} = \eta_{\mu\nu}$ , Ref. [26] computes an explicit ground-state wave function for the linearized theory from such an expression.

To simplify the perturbation expansion about a more general metric  $g_0$ , one should choose it to eliminate the linear terms in  $h$ . There is both the explicit term and the terms from the measures. A useful prescription starts from the quantum stress tensor in the initial state,

$$\langle T_{\mu\nu} \rangle_i = \langle \Psi_i | T_{\mu\nu} | \Psi_i \rangle = 2iM_p \frac{\delta}{\delta h^{\mu\nu}} \ln[Z_i], \quad (2.17)$$

where  $Z_i$  is the functional integral (2.16) without the classical and linear terms, and with  $T = \infty$  and  $\Psi_f = \Psi_i$ . Specifically, define  $g_0$  as a solution to

$$2 \frac{\delta S}{\delta g^{\mu\nu}}[g_0] = \langle T_{\mu\nu} \rangle_i. \quad (2.18)$$

In a black hole spacetime, such an equation precisely determines the semiclassical metric sourced by the average Hawking flux; such equations were studied explicitly in the two-dimensional context in [27]. For such a  $g_0$ , the linear term is eliminated, with the understanding [19] that at the same time we must replace  $T^\phi$  in (2.8) by its ‘‘normal-ordered’’ version,

$$\Delta T^\phi = T^\phi - \langle T^\phi \rangle_i, \quad (2.19)$$

and likewise for the corresponding quadratic expression in  $\nabla h$  in (2.7) that corresponds to a gravitational stress pseudotensor. The final result is of the form

$$\begin{aligned} \mathcal{A}[\Psi_i, \Psi_f; T] &= e^{iS[g_0, T]} \int_{\Psi_i}^{\Psi_f} \mathcal{D}h \Delta[C^I(h)] \\ &\times \det(\delta C^I) \mathcal{D}\phi e^{i(S_2 + S_3 + \dots)} \end{aligned} \quad (2.20)$$

with the replacement (2.19) implicit.

The semiclassical limit now appears as the weak-coupling limit  $M_p \rightarrow \infty$ , or more specifically  $p_i \cdot p_j \ll M_p^2$  where  $p_i \cdot p_j$  is a typical momentum invariant. In this case, one drops all terms beyond  $S_2$  in (2.20), as in [26]. The corresponding expression computes quantum amplitudes for the field  $\phi$  and the spin-two perturbation  $h_{\mu\nu}$  in the fixed background geometry  $g_0$ —this limiting procedure precisely produces the amplitudes of quantum field theory in a curved background. For a black hole, taking this limit holding  $M/M_p^2$  fixed preserves a nontrivial geometry.

Of course, when one attempts to evaluate (2.20) beyond quadratic order, one finds the usual UV divergences of quantum gravity. These are commonly believed to be *short-distance* issues. One could imagine regulating these by a cutoff, and working with an effective description of gravity below the cutoff. For example, in the context of a higher-dimensional theory, one could even think of string theory as a mechanism to provide a UV regulator. We will thus regard (2.20) as a useful effective description of long-distance gravity. A focus of this paper will be, assuming that such short-distance problems can be cured, are there any unanticipated issues in the *macroscopic* dynamics of gravity, that can be understood from this expression?

## B. Hamiltonian quantization

The Hamiltonian description of quantization gives a complementary perspective which will be particularly useful for considering evolution on time slices. This is of course directly connected to the path-integral description. For example, beginning with (2.5), differentiate with respect to  $T$  to determine the form of the evolution equation. The quantity  $T$  only enters through the asymptotic separation of the time slices at infinity. To be more specific, let us imagine that the amplitudes are defined in a region inside some very large radius  $R$ , in which case the slice separation and surface term in (2.2) are determined at  $R$ . The time derivative of (2.5) with respect to  $T$  then gives

$$\begin{aligned} i \frac{\partial}{\partial T} \mathcal{A}[\Psi_i, \Psi_f, T] &= \int_{\Psi_i}^{\Psi_f} \mathcal{D}g \mathcal{D}\phi e^{iS[g] + iS[g, \phi]} \\ &\times \delta \left( T - \int dt N(R) \right) \hat{H}, \end{aligned} \quad (2.21)$$

where

$$\hat{H} = -M_p^2 \int_R d^2\Omega \sqrt{2g} K \quad (2.22)$$

is, in the limit  $R \rightarrow \infty$ , the Bondi energy.

Suppose that the asymptotic form of the background metric is  $g_0 = \eta + M_p^{-1} h_0$ , and thus the full metric (2.6)

behaves asymptotically as

$$g \rightarrow \eta + M_p^{-1} (h_0 + h). \quad (2.23)$$

Then the energy (2.22) becomes

$$\hat{H} = M_0 - M_p^2 \int_R d^2\Omega (\sqrt{2g} K - \sqrt{2g_0} K_0) = M_0 + H, \quad (2.24)$$

where  $M_0$  is the mass of the background solution<sup>4</sup>  $g_0$ . The latter term can be derived as the spatial integral of a total derivative, which is the term in the projected constraint

$$N\mathcal{H} = -M_p^2 N G_{\perp\perp} + N T_{\perp\perp}^{\phi} \quad (2.25)$$

that is linear in  $h$ . From the constraint equation  $\mathcal{H} = 0$ , one then finds an expression for the Hamiltonian  $H$  in terms of an integral over the slices:

$$H = \int d^3x \sqrt{3g_0} N_0 (T_{\perp\perp}^{\phi} + t_{\perp\perp}^{\text{grav}}), \quad (2.26)$$

where  $T^{\phi}$  is the stress tensor for matter, and  $t^{\text{grav}}$ , which arises from the quadratic and higher terms in  $h$  in the constraint (2.25), is a gravitational energy-momentum pseudotensor. Thus, to summarize so far, the derivation of an analog of the Schrödinger equation for  $T$  evolution has been outlined, with Hamiltonian given by (2.24) and (2.26).

The Hamiltonian (2.26) generates the expected evolution of the fields on the time slices labeled by  $T$ . To see this, first consider the scalar action, written in terms of the metric decomposition (2.3),

$$\begin{aligned} S[g, \phi] &= \frac{1}{2} \int d^4x N \sqrt{3g} \left[ \frac{1}{N^2} (\partial_0 \phi - N^i \partial_i \phi)^2 \right. \\ &\quad \left. - g^{ij} \partial_i \phi \partial_j \phi \right]. \end{aligned} \quad (2.27)$$

The canonical momentum is

$$\Pi_{\phi} = N \frac{\delta S}{\delta \dot{\phi}} = \frac{1}{N} (\partial_0 \phi - N^i \partial_i \phi), \quad (2.28)$$

and canonical commutation relations are

$$[\Pi_{\phi}(x), \phi(y)] = -i \frac{1}{\sqrt{3g}} \delta^3(x - y). \quad (2.29)$$

One can likewise find the canonical momentum conjugate to the metric, which is given in terms of the extrinsic curvature  $K$  of the time slices by

$$\Pi_{ij} = \frac{M_p^2}{2} (K_{ij} - g_{ij} K). \quad (2.30)$$

The constraint equation takes the form

<sup>4</sup>Note that in the case of the semiclassical metric satisfying (2.18), this mass will actually be  $T$  dependent due to the average Hawking flux.

$$0 = N\mathcal{H} = 2M_p^{-2}N\left[\Pi_{ij}\Pi^{ij} - \frac{1}{2}(\Pi_i^i)^2\right] - \frac{M_p^2}{2}N(^3R) + \frac{N}{2}(\Pi_\phi^2 + g^{ij}\partial_i\phi\partial_j\phi). \quad (2.31)$$

Add to this the spatial diffeomorphism constraint,  $0 = N^i\mathcal{H}_i$ , where

$$\mathcal{H}_i = -2D^j\Pi_{ij} + \Pi_\phi\partial_i\phi. \quad (2.32)$$

The matter part is thus

$$N\mathcal{H}_\phi = \frac{N}{2}\left(\Pi_\phi^2 + g^{ij}\partial_i\phi\partial_j\phi + 2\frac{N^i}{N}\Pi_\phi\partial_i\phi\right). \quad (2.33)$$

This precisely generates, via the commutators (2.29), the evolution equation (2.28), and likewise the equation for  $\partial_0\Pi_\phi$ . The gravitational piece  $\mathcal{H}_h$  of  $\bar{\mathcal{H}} = N\mathcal{H} + N^i\mathcal{H}_i$  similarly generates evolution of the metric perturbation  $h$ .

Corresponding to the expansions of the action in powers of  $h$ , (2.7) and (2.8), there is an expansion of  $\bar{\mathcal{H}}$  in powers of  $h$ ,

$$\bar{\mathcal{H}} = \sum_n \mathcal{H}_n. \quad (2.34)$$

Keeping the quadratic term in this expansion corresponds to the approximation of free quantum fields  $\phi$  and  $h$  in the background  $g_0$ , as described at the end of the preceding subsection. Terms with  $n \geq 3$  describe interactions between these fluctuations. For validity of the weak-coupling approximation, these terms should be small.

### C. Hawking's argument for information loss

We are now prepared to review the argument for information loss [1,2]. Let  $g_0$  describe the metric of a Schwarzschild black hole of initial large mass,  $M_0 \gg M_p$ ; we could imagine that this black hole is formed in the distant past by a collapsing distribution of  $\phi$  matter, but the details of this will not concern us. The evolution equation (2.18) actually tells us that the geometry evolves as it radiates Hawking quanta, and, in particular, that its mass evolves as

$$\frac{dM}{dT} = -\gamma \frac{M_p^4}{M^2} \quad (2.35)$$

with  $\gamma$  a near-constant factor. Since the black hole is large, curvatures are weak away from the singularity and the last instant of evaporation. This was argued to justify use of the semiclassical approximation, which we see corresponds to keeping only  $S_2$  and  $\mathcal{H}_2$  in our preceding description.<sup>5</sup>

Thus the state of the matter field (as well as the fluctuating metric  $h$ ) can be derived from the viewpoint of quan-

<sup>5</sup>One could clearly generalize this to describe interacting matter weakly coupled to gravity.

tum field theory in the background metric  $g_0$ . Specifically, choose a set of slices that asymptote to constant Schwarzschild time slices far from the black hole, but that smoothly cross the horizon into the black hole and stay away from the singularity. Explicit descriptions of such nice slices were given by [28,12,22,19], and will appear in the next section. Standard quantum field theory methods show that an initial state such as the vacuum evolves to the form

$$|\Psi\rangle = \sum_{\alpha,\hat{\alpha}} \Psi_{\hat{\alpha}\alpha} |\hat{\alpha}\rangle |\alpha\rangle, \quad (2.36)$$

where the states  $|\hat{\alpha}\rangle$ ,  $|\alpha\rangle$  form bases for perturbations inside and outside the black hole, respectively. Outside observations are described by a density matrix,

$$\rho = \text{Tr}_{\hat{\alpha}} |\Psi\rangle\langle\Psi|, \quad (2.37)$$

and one easily argues that this density matrix has entropy

$$S = -\text{Tr}(\rho \ln \rho) \quad (2.38)$$

which is of order the Bekenstein-Hawking value,

$$S \sim S_{\text{BH}} = M_0^2/2M_p^2. \quad (2.39)$$

Moreover, locality/causality of quantum field theory in the background imply that the missing information described by the entropy  $S$  cannot escape, at least until evolution breaks down when the black hole approaches the Planck size. The simplest possibility is that the black hole completely and rapidly evaporates once it reaches this regime. But in this case, basic considerations of energy and entropy imply that the missing information cannot immediately escape the black hole. Indeed, as argued by Page [29], if information is to escape, it must begin to do so by the time the black hole has appreciably evaporated, say to size  $M_0/2$ , and thus on time scales  $T_{\text{evap}} \sim M_0^3$ . As a consequence, information was argued to be destroyed, violating unitary quantum evolution [2]. Another alternative is that a black hole leaves behind a long-lived remnant containing the information, but such scenarios appear ruled out by their prediction of infinite remnant species and thus divergent production rates.

This scenario produces a paradox, since an effective description of such information loss was argued [6,7] to lead to violent conflict with energy conservation, corresponding to effective vacuum temperatures of order  $M_p$ .

### III. QUANTIZATION ON NICE SLICES

To improve our understanding of the information loss argument, let us examine it in the context of our description of quantum gravity and the emergence of the weak-coupling/semiclassical approximation.

### A. Nice slices and semiclassical states

Begin with a more precise characterization of nice slices. As stated in the previous section, these should asymptote to slices of constant Schwarzschild time, but also enter the horizon while avoiding the singularity. Such a choice of slicing corresponds to a particular choice of gauge. There are many ways to choose such slices, but they all share certain characteristics. In particular, remaining spatial while avoiding the strong-curvature regime requires that inside the black hole they be closely spaced together in the large  $T$  limit; in other words, the lapse  $N$  becomes extremely small, roughly of order  $1/T$ . There are various ways that one could choose the bound on how close such slices get to the singularity; in the semiclassical approximation one could specify an explicit constraint on the minimum radius  $r_c$ , or alternatively on the maximal tetrad components of the curvature, maximal extrinsic curvature of the slice, etc. When reformulated in the quantum framework, all of these correspond to different choices of gauge, specifically the surface conditions described below Eq. (2.13).

Let us give an explicit example. One family of nice slices that is particularly convenient for our purposes can be constructed by taking a constant Schwarzschild time slice down to some outside radius  $r_o \sim 10M_0$ . At this point, we attach an arc, with moderate extrinsic curvature, that smoothly joins the  $r > r_o$  piece of the slice to the curve  $r = r_c$  inside the horizon. Once these curves meet, the rest of the nice slice, at Schwarzschild  $t$  coordinate before the joining point, is simply taken to be the surface  $r = r_c$ , where, say,  $r_c \sim M_0/10$ . The family of slices is generated by acting on this slice by a Schwarzschild time translation<sup>6</sup> (see Fig. 1). In particular, note that in this gauge the lapse  $N$  vanishes where the slices degenerate at  $r = r_c$ . Evolution is frozen here. This makes sense from the perspective of the scalar Lagrangian, (2.27): fluctuations of  $\partial_0\phi$  around  $N^i\partial_i\phi$  are suppressed—the only possible evolution is due to a nontrivial shift.

An interesting feature of such nice slices is that they continuously grow by stretching.<sup>7</sup> Specifically, consider a point  $P$  on the  $r_c$  section of the  $T$  nice slice, and compute the distance in the intrinsic slice geometry to the point at radius  $r_o$ . Then consider a later slice, with time  $T + \Delta T$ . The distance from the point  $P$  to  $r_o$  along this slice is larger by an amount

$$\Delta s = \Delta T \sqrt{\frac{2M}{r_c} - 1}. \quad (3.1)$$

<sup>6</sup>More carefully, the geometry  $g_0$  does not have an exact Killing vector; so one duplicates this construction for later Schwarzschild times, requiring a small deformation in the arc piece for large times. This will not be critical to our arguments.

<sup>7</sup>For one description of this phenomenon and an attempt to connect it to the breakdown of the semiclassical approximation, see [30].

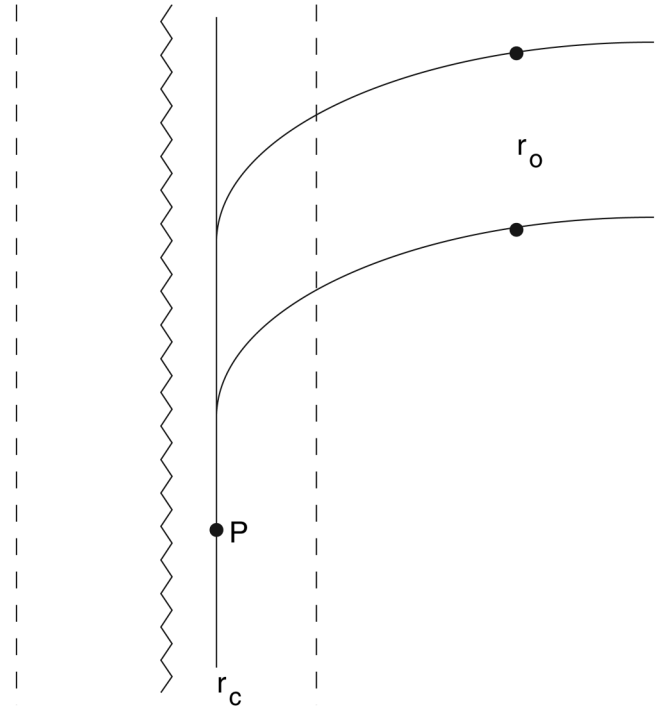


FIG. 1. A sketch of two nice slices in the Eddington-Finkelstein representation of the black hole geometry. The dotted line is the horizon. The stretching effect is easily seen, as the distance from point  $P$  to radius  $r_o$  on subsequent slices increases linearly with  $T$ .

This arises because an extra piece has been added to the slice at  $r = r_c$ ; the rest of the slice is the same. While the precise formula (3.1) is particular to our choice of slices, the generic phenomena holds for other choices, perhaps with more complicated formulas for the stretch.

Next consider the state on the nice slice that is predicted by Hawking's arguments [1] and subsequent refinements. In the semiclassical approximation, quanta are produced in pairs near the horizon. These have a thermal distribution of asymptotic energies and wavelengths at the Hawking temperature

$$T_H = \frac{M_p^2}{M}. \quad (3.2)$$

One can work in a basis where the asymptotic quanta are wave packets, which when described in terms of asymptotic retarded time  $u$  have widths  $\Delta u \approx M$ . A corresponding basis can be chosen for the paired quanta inside the black hole. Specifically, if we consider the initial state of the  $\phi$  field to be the (in) vacuum, we have

$$|0\rangle = \sum_{\{n_k\}} e^{-E(\{n_k\})/2T_H} |\{\hat{n}_k\}, \{n_k\}\rangle, \quad (3.3)$$

where  $\{\hat{n}_k\}$  and  $\{n_k\}$  label the states by occupation numbers for the inside and outside wave packet bases. (For more details, see [19].) One can think of a "typical" state as a

state where one Hawking quantum of wavelength  $M$  is emitted from the black hole for each (asymptotic) time step  $\delta T = M$ ; this indeed reproduces the evolution law (2.35). Correspondingly, the inside partner quanta (with paired correlations described by the state (3.3)) evolve towards  $r = 0$  and in the nice-slice description freeze at  $r = r_c$ . For asymptotic time step  $\delta T$ , the corresponding proper-distance separation between the ‘‘impact points’’ of the centers of the inside wave packets on the nice slice is

$$\delta s = \delta T \sqrt{\frac{2M}{r_c} - 1}, \quad (3.4)$$

and this formula also gives their characteristic proper-distance width. So, in summary, the Hawking state consists of a sum over states with paired inside and outside quanta, where the inside and outside quanta are spaced by typical proper distances along the slice  $\delta s$  and  $\delta T$ , respectively.

### B. Perturbations in nice-slice gauge

As explained in both Secs. II A and II B, the approximation of quantum field theory in a semiclassical background can be derived by taking the leading order terms in an expansion in  $M_p^{-1}h$ . Specifically, one drops the terms  $S_3$  and higher in (2.7) and (2.8), or  $\mathcal{H}_3$  and higher in (2.34), that describe couplings of the fluctuations in  $\phi$  and  $h$  to  $h$ . Moreover, as argued in Sec. II C, this is precisely the approximation in which a refined version of Hawking’s argument for information loss is made.

In examining the validity of the argument for information loss, one would like to assess the validity of such a weak-coupling expansion. In particular, this means examining the relative size of the subleading terms in the expansion (2.34). For a measure of this, phrase the question in the context of usual perturbation theory. If we have a quantum-mechanical system with Hamiltonian  $H = H_0 + H'$ , where  $H'$  is considered a perturbation of  $H_0$ , the effect of the perturbation can be easily described in the interaction picture. The time-dependent perturbed state takes the form

$$|\psi(t)\rangle = T \exp\left\{-i \int^t dt H'(t)\right\} |\psi\rangle_0, \quad (3.5)$$

where  $|\psi\rangle_0$  is the unperturbed state. Perturbation theory is valid as long as matrix elements of  $\int dt H'(t)$  are small. In contrast, suppose that

$$\langle \psi(t) | \psi \rangle_0 = 0 \quad (3.6)$$

for some  $t$ . This would be an indicator that perturbation theory has broken down, since in this case  $\langle \delta\psi | \psi \rangle_0 = -1$ : the perturbation has effect of order unity.

Specifically, in the present context, think of the terms  $S_2$  and  $\mathcal{H}_2$  as describing the unperturbed evolution, and the terms  $S_n$  and  $\mathcal{H}_n$  for  $n \geq 3$  as describing a perturbation of the Hamiltonian. One would like to assess whether these

higher-order terms indeed have small effect on the evolution, for example, as measured by their effect on the state as described in the preceding paragraph.

Thus we will consider the difference in evolution and, in particular, in final nice-slice states with and without the  $n \geq 3$  terms. In order to most cleanly exhibit the effect, consider an initial state that is a small perturbation of the ‘‘vacuum’’ black hole state. Specifically, one could consider a perturbation that corresponds to throwing one quantum into the black hole, at some early time, designated  $T = 0$ , changing its mass by  $m$ . A minimal such change is

$$m \sim M_p^2/M, \quad (3.7)$$

corresponding to the energy of one Hawking quantum. Or, if we think of the black hole evaporation as a statistical process, there will be fluctuations. The average energy flux is given by (2.18), but we might consider a state where there is a (typical) fluctuation such that one expected Hawking quantum fails to be emitted. This can be made explicit by, for example, projecting the state (3.3) onto states with a ‘‘missing’’ quantum in a particular mode  $k$  corresponding to an outgoing Hawking particle at time  $T = 0$ . This will change the mass of the hole by  $m$ , also given by (3.7). Such a fluctuation is representative of a typical fluctuation of the full quantum dynamics. Our goal is to compute the difference in the resulting state at some later time  $T$ , in either of these cases, that arises from including or excluding the  $n \geq 3$  terms in the Hamiltonian.

Of course, the approximation where  $n \geq 3$  terms are dropped is precisely the approximation where the back-reaction of the fluctuation on the metric is neglected; in this case the state  $|\Psi(T)\rangle$  is simply given by the original Hawking calculation, as described above.

One can estimate the effects of the perturbation terms by considering a perturbation that is purely  $s$  wave, as is in fact typical of the Hawking quanta. To compute the resulting state, one can shift the background  $g_0$  to include the effect of the perturbation. For example for an infalling quantum of mass  $m$ , this would be a new metric that is approximately described as an ingoing Vaidya solution [31], together with the effects of Hawking radiation described as in (2.18). Or, for a perturbation such that one Hawking quantum fails to escape, this would be a solution that looks like outgoing Vaidya, with the source stress tensor (2.19) describing a positive energy particle falling into the singularity, and the outside fluctuation represented by a negative energy particle travelling out to infinity, raising the mass of the hole by  $m$  at  $T = 0$ , with subsequent evolution again given by (2.18).

For simplicity, in evaluating the effect of the perturbation  $m$  we will ignore the Hawking flux represented by the stress tensor in (2.18), although analogous statements can be argued to hold there as well. Thus, we treat the perturbed geometry as simply Schwarzschild with mass  $M_0 + m$ .

How is the nice-slice state changed? This of course depends on the choice of slicing, i.e. gauge. In other words, clearly one can take a slicing of the perturbed spacetime such that there *appears* to be a big effect. To avoid this, one tries to specify gauge conditions that are “as nice as possible,” in the sense that they lead to *minimal* change in the resulting state. Outside the black hole, one can in this spirit take the new nice slices to still be described as Schwarzschild time slices far from the black hole, and take some small modification of the arc crossing the horizon. Inside the black hole, there are a number of options. The simplest is to keep the nice slices at the radius  $r_c$  determined by the original geometry; in this case the angular part of the intrinsic geometry of the slices is fixed. Alternately, one might choose a gauge such that  $\nabla^i \bar{h}_{ij} = 0$  on the slices, or  $\delta K_{ij} = 0$ , or even such that components of the Riemann tensor are fixed on the slices. Any of these conditions, and a multitude of others, will lead to a fractional change of  $M/r_c$  of order  $m/M \gtrsim M_p^2/M^2$ . In particular, this means a small change of the relation (3.1) for stretch versus asymptotic time,

$$\Delta s = \Delta T \sqrt{\frac{2M}{r_c} - 1} [1 + \mathcal{O}(m/M)]. \quad (3.8)$$

Likewise the description of the evaporation changes by  $\mathcal{O}(m/M)$  in the Hawking temperature, and thus energy of the typical outgoing Hawking quanta on the nice slices, as well as in the spacing  $\delta s$  of the wave packet crests hitting the nice slice, and in their widths.

We might compare the resulting perturbed nice-slice state  $|\Psi'(T)\rangle$  to the state where we ignore the effects of  $\mathcal{H}_3$ , etc., by taking their inner products. Considering, for example, the  $r = r_c$  part of the slice, they both have corresponding wave packets that hit the slices at proper distance intervals separated by  $\delta s$  or  $\delta s'$ , which are nearly identical. Thus the contribution to the inner product from corresponding wave packets is essentially unity, to  $\mathcal{O}(m^2/M^2)$ . Quantum by quantum, the states are nearly identical.

However, if we take into account the contributions of quanta corresponding to Hawking emission at very late times, this changes. For a first estimate, note that the cumulative shift in the center of the  $n$ th quantum in the perturbed state, relative to the unperturbed state, is of order

$$s'_n - s_n \sim n \frac{m}{M_p^2} \sqrt{\frac{2M}{r_c} - 1}. \quad (3.9)$$

A crude estimate is that when this becomes of order the size of the wave packets, (3.4), then the overlaps of the wave packets, and thus the states, nearly vanish, as illustrated in Fig. 2. This occurs for a time  $T$  given by

$$T \sim \frac{nM}{M_p^2} \sim \frac{M^2}{mM_p^2}, \quad (3.10)$$

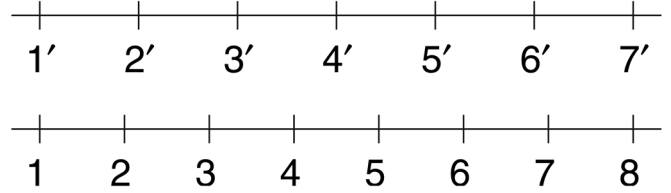


FIG. 2. An illustration of the comparison of the state on the internal part of a nice slice with and without the effects of the perturbations  $\mathcal{H}_n$ ,  $n \geq 3$ . Vertical marks represent the centers of corresponding wave packets. Initially the mismatch is small, but it increases along the slice. A similar mismatch develops on part of the slice outside the black hole.

or, in a  $D$ -dimensional generalization,

$$T \sim \left(\frac{M}{M_p}\right)^{(D-2)/(D-3)} \frac{1}{m}. \quad (3.11)$$

For an initial perturbation of size (3.7), this is

$$T \sim M^3/M_p^4, \quad (3.12)$$

with generalization

$$T \sim M_p^{-1} \left(\frac{M}{M_p}\right)^{(D-1)/(D-3)}. \quad (3.13)$$

For a larger perturbation  $m$ , it occurs even earlier.

Therefore, by the time (3.12) the perturbations  $\sum_{n \geq 3} \mathcal{H}_n$  have an effect that is of order one, by a criterion analogous to (3.6); the same effect also deforms the state outside the black hole. This effect indicates that this perturbative expansion is not valid at such times. One way of thinking of this is that the shift in the Hamiltonian is very small. However, over long times it has a coherent effect on the state, and thus can have a significant net effect. This is a failure of the leading order treatment of the problem in the weak-coupling expansion. As a result, the leading order  $n = 2$  perturbative treatment, which we have seen produces the semiclassical description of [1], does not accurately compute the state at times given by (3.10), or the corresponding density matrix.

Indeed, note that the *net* effect on the nice-slice state becomes important even before the time (3.12). Specifically, given the offset (3.9) and for Gaussian wave packets of width  $M$ , the inner product between the  $n$ th perturbed wave packet and the  $n$ th unperturbed wave packet is thus of order

$$\mathcal{P}_n \approx e^{-(nm/M)^2}. \quad (3.14)$$

The inner product between perturbed and unperturbed states, accounting for the first  $N$  quanta, is thus estimated as

$$\mathcal{P} \approx \prod_{n=1}^N \mathcal{P}_n \sim e^{-N^3(m/M)^2}. \quad (3.15)$$



This product is small past the time

$$T \sim \frac{M}{M_p^2} N \sim \frac{M^{5/3}}{m^{2/3} M_p^2}, \quad (3.16)$$

or  $T \sim M^{7/3}$  in Planck units, for  $m$  given by (3.7). The time estimate (3.10) is an estimate for where the effect on *each quantum* becomes significant, but before this, at the time (3.16), the net effect on the state becomes significant. (Notice that the argument of [29] only tells us *by* what time scale there must be some interactions that relay information if information is to emerge in the Hawking radiation.)

## IV. DISCUSSION

### A. Diagrammatics

To better understand the result of the preceding section, begin by describing it in terms of Feynman diagrammatics, arising from expanding the functional integral (2.20), or equivalently the evolution operator derived from the Hamiltonian (2.34), order by order in  $1/M_p$  about the quadratic approximation, so perturbatively in the terms  $n \geq 3$  of the action or Hamiltonian.

Notice that in the black hole context, there is a generalization of loop diagrams. In particular, simply in the quadratic approximation, one can have a nonzero amplitude of the form

$$\langle \Psi_f | \Delta T_{\mu\nu} | 0 \rangle, \quad (4.1)$$

where the operator  $\Delta T_{\mu\nu}$ , which could be the energy-momentum tensor either for  $\phi$  or for  $h$  fluctuations, is normal ordered as in (2.19) with respect to an in basis, and thus can create a nontrivial out state with an extra pair of particles: one escapes to infinity, and the other falls into the strong-curvature regime.

When one includes the cubic and higher interactions, one also has diagrams with such ‘‘open loops’’ connected by graviton propagators. For example, the second-order term in the expansion of  $\exp\{iS_3\}$  has a contribution corresponding to a pair of vacuum fluctuations connected by a graviton propagator. The case where the vacuum fluctuations are separated by a large time  $T$  is the case described in the preceding section. Specifically, the effect of the first fluctuation on the metric was computed, and then this in turn interacted with the second fluctuation to produce a state that deviates from the state with no interaction terms. This amplitude could equally be represented by a Feynman diagram (illustrated in Fig. 3), of the form

$$\langle \Psi_f' | \Delta T_{\mu\nu} D^{\mu\nu,\lambda\sigma} \Delta T_{\lambda\sigma} | 0 \rangle, \quad (4.2)$$

where we consider states  $\Psi_f'$  on the late-time nice slice that contains both the early and late fluctuations. Here  $D^{\mu\nu,\lambda\sigma}$  is the graviton propagator. The large phase shift described in

the preceding section corresponds to the amplitude (4.2) giving a large correction. Likewise, one could consider higher-order amplitudes, e.g. expanding to higher order in  $S_3$  or in higher terms in the action.

### B. Large boosts

From this perspective an obvious question is the origin of the large effect. One way of thinking of this is to note that for large time separations  $T$ , there is a large relative boost between the two fluctuations. This boost can be estimated by integrating the extrinsic curvature along the nice slice. Specifically, the change in the normal to the slice under a small displacement  $dl$  is

$$dn_i = K_{ij} dl^j. \quad (4.3)$$

Since the extrinsic curvature takes the constant value

$$K_{tt} = -\frac{M}{r_c^2} \sqrt{\frac{2M}{r_c} - 1} \quad (4.4)$$

along the slice, the net boost parameter between two points separated by  $T$  is

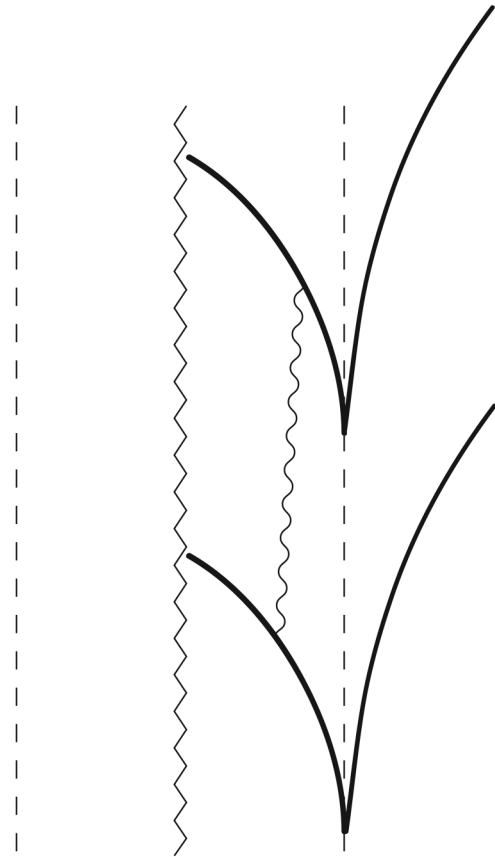


FIG. 3. A Feynman diagram representation of the coupling between fluctuations at different times. The wavy line is the graviton propagator.

$$\theta = \int E^t K_{tt} dt = \frac{M}{r_c^2} T \quad (4.5)$$

where  $E$  is the unit vector in the  $t$  direction. Thus the boost is exponentially large in the time separation  $T$ . This was argued in [16,19,21,32] to explain how small early fluctuations could have large effects on the late metric and late fluctuations, since large boosts magnify gravitational fields. In the preceding section, we have indeed argued that the effect on the late-time state is large. While this is apparently related to the large boost, the relevant calculation of the graviton propagator in nice-slice gauge yields an effect on the state that only grows as  $T$ . In any case, the significant coupling between perturbations appears closely parallel to the origin of strongly coupled gravity in a high-energy collision in a flat background [20].

### C. Lessons and nonlocality

The results of this paper suggest the following conclusions. First of all, the present weak-coupling (or semiclassical) expansion of quantum gravity, which we have shown produces the approximation of quantum field theory in a curved background, apparently breaks down for the purposes of calculating late-time states on nice slices. Validity of the semiclassical approximation has been assumed in [1] and in subsequent calculations of the state. It may be that there is a different way of constructing a perturbative expansion that gives a controlled calculation of the state, but this author does not see the concrete evidence for this at the present.

This approximation may certainly yield valid results for some quantities, like the expectation value of the stress tensor outside the black hole, etc., as such quantities can be computed without explicit reference to the state on the nice slice. However, a calculation critical to the information paradox is that of the density matrix describing the outside world, and its entropy. Existing calculations of this rest heavily on calculating the combined state inside and outside the black hole, and tracing over the internal state, and the most precise extant versions of these apparently rest on logic that amounts to using nice-slice states. For the purposes of computing the density matrix and its entropy, one must do a very fine-grain and precise calculation. In the present approach to this calculation we have exhibited an apparent failure of the semiclassical approximation.

If the state and density matrix cannot be reliably calculated, this calls into serious question the arguments given for an information paradox—and if reliance on an unreliable calculation produces a paradox, the conclusion appears obvious. However, one would like to go further and understand how one could perform a reliable calculation, and how such a calculation would demonstrate information escape from the black hole. The preceding discussion is consistent with the suggestion that such a calculation must be given in the context of a nonperturbative description of

quantum gravity, since the perturbative approach appears to fail over *macroscopic* distances.

In particular, it *could* be that there is a nonperturbative calculation of the precise analog of a nice-slice state and outside density matrix that retains the essential features of local quantum field theory, and, in particular, implies that the nice-slice state has information content deep inside the black hole that is not revealed to the outside world. This author regards this assumption about the nonperturbative dynamics as in doubt, as it returns us to the original paradox.

Thus one might consider other alternative assumptions about nonperturbative gravitational dynamics. There are several reasons to question the assumption of locality. First, there are no known precisely local observables in a quantum theory of gravity; the best one can apparently do is to construct certain relational protocal observables that approximately reduce to local observables in certain states [33,34]. This suggests that there is no precise notion of locality in quantum gravity. Second, if one attempts to probe locality via the  $S$  matrix, in high-energy scattering, one finds characteristic growth of the cross section with energy  $\propto E^2$  in four dimensions, which violates expected bounds on local theories. Third, recent developments in string theory, and, in particular, the AdS/CFT correspondence [11], suggest that there may be a complete “holographic” boundary description of dynamics in a spacetime region, in contrast to the tenets of local field theory. Finally, there is the paradox itself, which seems inevitable if physics is local.

In fact, the effects described in the preceding section, while indicating a breakdown of reliable calculations, appear to hint at a nonlocal effect, as the fluctuations on the nice slice deep within the black hole develop unanticipated correlations with the fluctuations in the vicinity of the horizon. And, as described above, the origin of the effect appears closely connected with the dynamics of high-energy gravitational scattering.

Thus, these arguments suggest a possible entree for an alternative assumption about nonperturbative gravitational dynamics, that of a *nonlocality principle* [19,16,17,20,21], stating that gravitational dynamics is indeed fundamentally quantum mechanical, but that the dynamics that unitarizes gravity at strong coupling does not have the usual locality properties of local quantum field theory. Criteria for operation of such a nonlocality principle were suggested in [21,16,17]; in a flat background these were phrased in terms of a locality bound (with  $n$ -particle generalization in [20]), stating that when two modes with sufficiently high center-of-mass energy (i.e. relative boost) have sufficiently small separation, a description via local quantum field theory (QFT) fails. As we have seen above, the present criterion for the breakdown of the semiclassical approximation appears closely related, and, in particular, the breakdown of the perturbative expansion arising in the

black hole geometry apparently prevents a straightforward semiclassical treatment.

One can guess at how such nonlocal dynamics could yield unitary evolution. Specifically, a nontrivial state which initially appears of the form

$$|\psi_i\rangle = \sum_{\alpha, \hat{\alpha}} \psi_{\alpha \hat{\alpha}} |\hat{\alpha}\rangle_{\text{in}} |\alpha\rangle_{\text{out}}, \quad (4.6)$$

and thus contains entanglement between inside and outside degrees of freedom, might by such nonlocal dynamics evolve to a state of the form

$$|\psi_f\rangle = |BH\rangle_{\text{in}} |\psi_f\rangle_{\text{out}} \quad (4.7)$$

as described from the perspective of the outside observer, where the state  $|BH\rangle$  is an internal state that does not contain information about the initial state. This is essentially the final state proposal of [35]. What was lacking in that proposal was an explanation of how local physics could be violated to give such evolution; a breakdown of the semiclassical approximation together with the proposed nonlocality principle would provide such a rationale.<sup>8</sup>

As stated, a reasonable viewpoint is that the black hole information paradox is a challenge to local quantum field theory of the same magnitude of the challenges to classical physics that ultimately catalyzed the invention of quantum mechanics. Specifically, the classical instability of matter serves as a sharp example of what is wrong with the classical model of the atom. While one could imagine a naive classical physicist approaching this as a mathematical problem to be resolved by some modification of the singularity in classical evolution where the electron reaches the charge center, we know that this approach is far from the mark: the classical description fails to be the correct description, and must be *replaced* by a quantum description, at scales set by the Bohr radius. Classical physics is invalid at the Bohr radius, without there being any explicit signal *within* the classical theory that it fails there.

Likewise, many gravity theorists have been perplexed about the dynamics at a black hole singularity, and have made various proposals about what smooths singularities. But the information paradox suggests that local quantum theory could break down on scales of order the Schwarzschild radius. This could plausibly be the ultimately correct picture, whether or not we had found a signal within existing theory in the form of an explicit breakdown of local QFT at the Schwarzschild radius. The arguments of the preceding section indeed suggest that history has not repeated itself precisely: they are consistent with an actual breakdown of a local quantum analysis. But whether or not this is an ultimate indicator of new nonlocal

quantum physics, such nonlocal physics could be the correct description of quantum dynamics on Schwarzschild scales, just as quantum dynamics replaces classical dynamics at atomic scales.

Independent of the preceding arguments, there is another argument suggesting that one question the argument for information loss. This is Wheeler’s paraphrase of Bohr’s dictum: “no phenomenon is a real phenomenon until it is an observed phenomenon.” The information loss argument relies heavily on the notion of the internal state on a nice slice. But one must be careful in specifying a gauge-invariant description of this internal state. Consider what would be needed to actually measure this state, in the spirit of the discussion of protocol observables of [33,34]. In essence, the state can only be measured by constructing detectors that it correlates with. So, one could construct detectors at infinity, and build them with sensors that activate measurements when they reach the radius corresponding to the nice slice. A family of such “depth-charge” detectors thus in principle could apparently determine the state on the nice slice. However, to do so accurately one would have to make a measurement for each spatial interval of size  $\delta s$  given in (3.4). An absolute (and unrealistically low) minimum for the energy each detector would have is estimated as  $\mathcal{O}(M_p^2/M)$ . A family of such infalling detectors would thus have a huge effect on the black hole—and, in particular, it would deviate from the undisturbed black hole on the  $M^3$  time scale, by failing to evaporate. Notice that these ideas that we must be able to measure the state actually involve a weaker assumption than those of black hole complementarity [37], in that the latter discussion describes both local measurements and moreover requires that the results be relayed to a *common observer*.

#### D. Inflationary cosmology

One anticipates similar effects might play a role in the breakdown of perturbative local effective theory in other spacetimes with horizons, and, in particular, in inflationary cosmologies. This is suggested by the following rough argument. Consider de Sitter space. Here one can think of the spheres of the global description as “nice slices.” Thermal de Sitter radiation corresponds to the de Sitter horizon emitting roughly one quantum of wavelength  $\lambda \sim R_{dS}$  in each e-folding time  $R_{dS}$ .

For simplicity, focus on the  $s$  wave sector with respect to a definite observer taken by convention to reside at the north pole. In analogy with the black hole, pairs of quanta are produced on either side of this observer’s horizon. These follow the trajectories shown in Fig. 4. Regions I and III are de Sitter, but region II should be Schwarzschild-de Sitter, resulting from the energy perturbation. The observer  $O$  thus appears to experience late evolution in this Schwarzschild-de Sitter region. This observer is expected to see a thermal spectrum with a temperature shifted by the

<sup>8</sup>For one discussion of the nonlocality of the final state proposal, see [36].

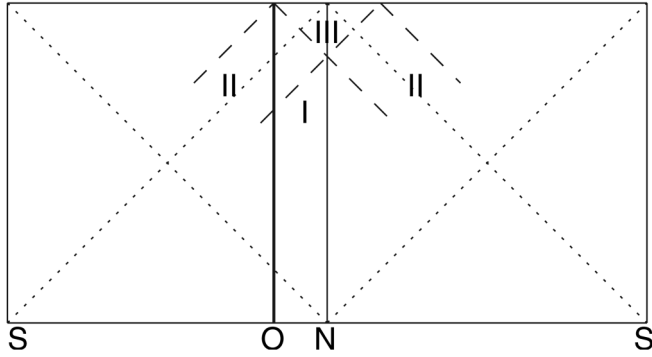


FIG. 4. An extended Penrose diagram for de Sitter space. Horizontal lines correspond to  $D - 1$  spheres which are thought of as nice slices;  $N$  and  $S$  label the north and south poles, respectively. A fluctuation that is  $s$  wave with respect to the observer at  $N$  produces a pair of particles (shown as dashed lines) on either side of the horizon. Regions I and III are thus de Sitter space, but region II is described as Schwarzschild-de Sitter. The observer at  $O$  is as a result expected to see a significant late-time phase shift in the nice-slice state, relative to the corresponding state without backreaction included.

fractional amount  $\delta m/R_{dS}$  where as in (3.7) we expect the minimum such change given by  $\delta m \sim 1/R_{dS}$ . Correspondingly, after a time  $T$ , the state on the nice slice is expected to experience a net shift in the positions of the quanta analogous to (3.9). Thus, for  $T \sim R_{dS}^3$ , the shifts are estimated to become of order  $R_{dS}$ . This would suggest that the perturbative expansion producing local field theory breaks down after  $N \sim S_{dS} \sim R_{dS}^2$  efoldings, and at this time an original region of size  $R_{dS}$  has inflated to size  $\sim R_{dS}e^{S_{dS}}$ . For later times/larger volumes, this argument suggests a perturbative local description is not valid. Similar statements have been inferred from the general analogy between black hole and de Sitter horizons by Arkani-Hamed [38]. There may be other manifestations of breakdown of local field theory in de Sitter space as well; some previous work on this subject includes [39–45] and these issues are presently under investigation [46].

## V. CONCLUSION

This paper has described an attempt to carefully justify the derivation of the state on a set of nice slices through a black hole geometry, in a semiclassical approximation like

that used in [1], and in subsequent improvements, to argue for fundamental information loss. While one can in principle derive local quantum field theory in a leading order approximation, in the limit  $M_p \rightarrow \infty$ , to either the path integral or to a corresponding Hamiltonian evolution, there are apparently circumstances where such a perturbative approach fails. One example is that of high-energy scattering [20,21], and this paper has presented arguments apparently consistent with the statement that perturbative gravity fails to give accurate results in computing precise states on nice slices for times  $T \gtrsim \mathcal{O}(M^3)$ . Thus the present approach to perturbative gravity does not appear to accurately give the density matrix or its entropy. This happens by a time scale of order the time by which information would need to escape the black hole [29] if evolution is indeed unitary.

Since assuming a valid perturbative calculation apparently leads to the information loss paradox, an invalidity offers a welcome possible out. Specifically, if as it appears, there is no reliable calculation of the information destroyed by a black hole, there is no information paradox. Beyond this, to actually explain how information escapes a black hole, one would apparently require some new dynamics of an essentially nonlocal character. There are several reasons to question the essential nature of locality in nonperturbative quantum gravity: rapid growth of high-energy scattering amplitudes, fundamental limitations on local observation [33,34], suggestions from string theory via the AdS/CFT correspondence, and finally the magnitude of the paradox itself. Thus, the combined evidence plausibly indicates the existence of a gravitational nonlocality principle, stating that the nonperturbative dynamics of gravity has a certain nonlocal quality, or alternatively, that the usual notion of locality is not an essential part of the theory and is only recovered in an approximation. If this is correct, a key question is: what is this dynamics?

## ACKNOWLEDGMENTS

The author wishes to thank N. Arkani-Hamed, G. Horowitz, D. Marolf, S. Mathur, and especially J. Hartle for important discussions. This work was supported in part by the Department of Energy under Contract No. DE-FG02-91ER40618, and by Grant No. RFPI-06-18 from the Foundational Questions Institute (fqxi.org).

- [1] S. W. Hawking, *Commun. Math. Phys.* **43**, 199 (1975); **46**, 206(E) (1976).
- [2] S. W. Hawking, *Phys. Rev. D* **14**, 2460 (1976).
- [3] S. W. Hawking, *Phys. Rev. D* **72**, 084013 (2005) [arXiv:hep-th/0507171].

- [4] S. B. Giddings, arXiv:hep-th/9412138; arXiv:hep-th/9508151.
- [5] A. Strominger, arXiv:hep-th/9501071.
- [6] J. R. Ellis, J. S. Hagelin, D. V. Nanopoulos, and M. Srednicki, *Nucl. Phys.* **B241**, 381 (1984).

- [7] T. Banks, L. Susskind, and M.E. Peskin, Nucl. Phys. **B244**, 125 (1984).
- [8] S.B. Giddings, Phys. Rev. D **46**, 1347 (1992) [arXiv:hep-th/9203059].
- [9] G. 't Hooft, arXiv:gr-qc/9310026.
- [10] L. Susskind, J. Math. Phys. (N.Y.) **36**, 6377 (1995) [arXiv:hep-th/9409089].
- [11] J.M. Maldacena, Adv. Theor. Math. Phys. **2**, 231 (1998); Int. J. Theor. Phys. **38**, 1113 (1999) [arXiv:hep-th/9711200].
- [12] D.A. Lowe, J. Polchinski, L. Susskind, L. Thorlacius, and J. Uglum, Phys. Rev. D **52**, 6997 (1995) [arXiv:hep-th/9506138].
- [13] D.N. Page, Phys. Rev. Lett. **44**, 301 (1980).
- [14] G.T. Horowitz, J. High Energy Phys.08 (2005) 091 [arXiv:hep-th/0506166].
- [15] G.T. Horowitz and E. Silverstein, Phys. Rev. D **73**, 064016 (2006) [arXiv:hep-th/0601032].
- [16] S.B. Giddings and M. Lippert, Phys. Rev. D **65**, 024006 (2001) [arXiv:hep-th/0103231].
- [17] S.B. Giddings and M. Lippert, Phys. Rev. D **69**, 124019 (2004) [arXiv:hep-th/0402073].
- [18] D. Lowe and L. Thorlacius, Phys. Rev. D **73**, 104027 (2006) [arXiv:hep-th/0601059].
- [19] S.B. Giddings, Phys. Rev. D **74**, 106005 (2006) [arXiv:hep-th/0605196].
- [20] S.B. Giddings, Phys. Rev. D **74**, 106006 (2006) [arXiv:hep-th/0604072].
- [21] S.B. Giddings, Phys. Rev. D **74**, 106009 (2006) [arXiv:hep-th/0606146].
- [22] J. Polchinski, arXiv:hep-th/9507094.
- [23] T. Banks, Nucl. Phys. **B249**, 332 (1985).
- [24] J.B. Hartle, in *Proceedings of ASI, Cargese, 1986*, edited by B. Carter and J.B. Hartle (Plenum, New York, 1987), p. 329.
- [25] C. Kiefer, Int. Ser. Monogr. Phys. **124**, 1 (2004).
- [26] J.B. Hartle, Phys. Rev. D **29**, 2730 (1984).
- [27] C.G. Callan, S.B. Giddings, J.A. Harvey, and A. Strominger, Phys. Rev. D **45**, R1005 (1992) [arXiv:hep-th/9111056].
- [28] R.M. Wald (unpublished).
- [29] D.N. Page, Phys. Rev. Lett. **71**, 3743 (1993) [arXiv:hep-th/9306083].
- [30] S.D. Mathur, Int. J. Mod. Phys. A **15**, 4877 (2000) [arXiv:gr-qc/0007011].
- [31] P.C. Vaidya, Proc. Indian Acad. Sci. A **33**, 264 (1951).
- [32] T. Banks and W. Fischler, arXiv:hep-th/0606260.
- [33] S.B. Giddings, D. Marolf, and J.B. Hartle, Phys. Rev. D **74**, 064018 (2006) [arXiv:hep-th/0512200].
- [34] M. Gary and S.B. Giddings, Phys. Rev. D **75**, 104007 (2007) [arXiv:hep-th/0612191].
- [35] G.T. Horowitz and J. Maldacena, J. High Energy Phys. 02 (2004) 008 [arXiv:hep-th/0310281].
- [36] D. Ahn, Phys. Rev. D **74**, 084010 (2006) [arXiv:hep-th/0606028].
- [37] L. Susskind, L. Thorlacius, and J. Uglum, Phys. Rev. D **48**, 3743 (1993) [arXiv:hep-th/9306069].
- [38] N. Arkani-Hamed, in Proceedings of the Conference on String Phenomenology, KITP, UC Santa Barbara, 2006 (unpublished).
- [39] T. Banks, arXiv:hep-th/0007146.
- [40] W. Fischler (unpublished); W. Fischler, in Proceedings of the Role of Scaling Laws in Physics and Biology (Celebrating the 60th Birthday of Geoffrey West), Santa Fe, 2000 (unpublished).
- [41] R. Bousso, arXiv:hep-th/0205177.
- [42] L. Dyson, M. Kleban, and L. Susskind, J. High Energy Phys. 10 (2002) 011 [arXiv:hep-th/0208013].
- [43] T. Banks, W. Fischler, and S. Paban, J. High Energy Phys. 12 (2002) 062 [arXiv:hep-th/0210160].
- [44] N. Goheer, M. Kleban, and L. Susskind, J. High Energy Phys. 07 (2003) 056 [arXiv:hep-th/0212209].
- [45] T. Banks, arXiv:astro-ph/0305037.
- [46] S.B. Giddings and D. Marolf (work in progress).