Comparing models for a unitary black hole S matrix

Steven B. Giddings[*]

Department of Physics, University of California, Santa Barbara, California 93106, USA

This paper compares features, challenges, and puzzles of different models for a unitary black hole S-matrix, focusing on both recent nonisometric models, as well as "nonviolent unitarization," which is based on new quantum interactions of a black hole. As a foundation for comparison, the description of real-time Hawking evolution is first overviewed, including leading effects of gravitational dressing and backreaction. Connection is then made to qubit models for evolution, and some technology is outlined to facilitate their description. Important features of both nonisometric models and nonviolent unitarization are investigated in qubit models, which illustrate essential differences between the respective approaches. The nonisometric models present puzzles for understanding evolution of internal outgoing excitations, which can be excited by interactions such as particle decay. Qubit models for nonviolent unitarization are further developed, and nicely illustrate aspects of that approach. Some remaining questions in generalizing to more complete models for evolution are discussed.

I. INTRODUCTION

Consider a cavity with reflecting walls, and a tiny hole through them through which we inject a very high energy photon in a pure state. This photon will generically create additional particles, which evolve towards an approximately thermal distribution inside the cavity. Over a long time, particles will also find the small hole, and leak out. The total state of particles will be pure, but the escaping particles will be entangled with those inside the cavity, and so a density matrix for them will be mixed and initially will have an increasing von Neumann entropy.

After a very long time, all the particles will have leaked out of the cavity, which will now be in its vacuum, and the state of the radiation outside will correspondingly be a pure state if one accounts for all the escaped particles.

If one did not understand the fundamental dynamics but had seen proposals about implementing a final state boundary condition in quantum mechanics [1–3], it might be tempting to propose a new rule, the final state proposal for cavities: that ordinary unitary evolution is modified so that after a long time the fundamental evolution projects on the interior ground state.

However, this is obviously unnecessary, and does not represent the correct fundamental dynamics. It is the small

interactions between cavity modes and modes of the exterior, arising from the hole, combined with the infinite phase space of the exterior of the cavity, that accomplishes the same thing. No new rule is needed.

Black hole (BH) formation and evaporation presents us with a seemingly very similar situation, where an initially simple, even pure, state can form a black hole, in whose interior we expect a complicated state to result. Particles leak out, and ultimately we expect the black hole to disappear, effectively returning its interior to vacuum and the outside state to the same purity to avoid other contradictions arising from information loss [4,5] or remnants [6–9].

What is different is that a description of BH evolution based on that of local quantum field theory (LQFT) on a semiclassical background, while allowing particles to escape [10], does not allow entanglement (which can be used to track quantum information) to escape: its escape is forbidden by locality with respect to the semiclassical background geometry. There are different possible reactions to this.

One is to propose that a fundamentally new principle like the final state proposal [1–3] now comes into play [11], and enforces the ultimate triviality of the BH internal state, while implementing unitarity of the full S-matrix. A newer variant on this theme is discussed in the models of [12], which also successively incorporates final state projections, fundamentally modifying the evolution.

Another is to explore a resolution analogous to evolution of conventional quantum systems like our cavity, arising from small interactions between BH states and those of the surroundings. Of course, the presence of such interactions still appears rather radical, since such interactions would appear to violate the principle of locality—specifically the

requirement that information propagates inside the lightcone—with respect to the background semiclassical geometry. Indeed, if such interactions are present, they may well be evidence of some more fundamentally quantum properties of spacetime itself, and so ultimately are plausibly also associated with a modification of fundamental principles.

Either way, we confront the apparent need for new fundamental principles. This paper will further compare and contrast these two scenarios, treated as modifications to the more concretely understood basic evolution described by LQFT, updating Hawking [10] and many intervening works.

One can consider other possibilities as well; our thinking about them can be organized by considering a "black hole theorem [13]." A first question is whether a BH can be thought of as a quantum subsystem of a larger system. There has been progress on seeing that this is possible, at least to a good approximation [14–16], despite arguments to the contrary [17], but questions remain on the fundamental structural question of how to define subsystems of quantum gravitational systems, and these conceivably present a different out. Or, perhaps other dynamics, e.g. that of fuzzballs [18], could modify that assumption. But, assuming that BHs *do* effectively behave as subsystems, and that we wish to avoid the apparent contradictions of information loss or microscopic remnants, it appears that such a theorem then leaves one with the choices of modification of quantum evolution, or small interactions, or scenarios that are an even greater departure from the tried and tested physics of LQFT, and which also have their own problems.

In order to better understand the nature of modifications to LQFT evolution, this paper will first give in Sec. II a brief overview of aspects of a more detailed description of such real time evolution, including a leading order treatment of quantized general relativity (QGR), incorporating the leading effects of the quantum gravitational field: dressing and backreaction. This provides a basis for comparison of modifications to such evolution, particularly if such modifications can be treated as corrections to such evolution that are in some sense "small."[1] The next section will discuss some basic considerations regarding modifying that evolution. Section IV then investigates qubit models for Hawking evolution and for its modification either via nonisometric evolution, or unitarizing interactions. In the process it develops some useful technology for describing qubit models, and investigates the real-time evolution of the nonisometric models as well as potential issues that they raise. It also further develops nonviolent models for unitarizing interactions and dynamics. Section V concludes with discussion of some questions associated with extension of

such models to a more complete description of evolution, and of further comparison of models in this context.

II. EVOLUTION IN PERTURBATIVE LQFT/QGR: OVERVIEW

This section will briefly review the evolving quantum state of a black hole, as described within perturbative LQFT/QGR. Many features of this have been understood for decades, though aspects of a more complete description, including the role of gravitational dressing/backreaction, have been provided recently in [22,23].

We consider perturbations about a background metric $g_{\mu\nu}$, which we take to describe a BH that formed from collapse. At later times, the classical solution settles down to a stationary black hole; we focus on the simplest static case, with radius R. If we account for Hawking radiation, the mass decreases at the rate $dM/dt \sim 1/R^2$, which can be accounted for by including backreaction of the expectation value of the leading quantum stress tensor in determining $g_{\mu\nu}$.[2] However, since the fractional mass decrease during the typical emission time R of a quantum is $\Delta M/M \sim 1/MR \sim 1/S_{\mathrm{BH}}$, with S_{BH} the Bekenstein-Hawking entropy, this metric is to very good approximation static over many such emission times, during the period after it settles down from initial formation, and before its final demise. We consider quantum perturbations of both infalling and outgoing excitations.

Gauge fixing is needed to describe the evolving quantum state, including matter and metric perturbations, of the BH. Specifically, in QGR this gauge fixing can be accomplished by providing a family of time slices, and spatial coordinates on these slices, described through functions

$$x^\mu = \mathcal{X}^\mu(t, x^i) \qquad (2.1)$$

parameterizing these slices in general coordinates x^μ. Such slices, and evolution on them, are further discussed in [26,27,22]. It is simplest to choose these slices to respect the spherical symmetry of the background metric. Using the standard radial coordinate and an ingoing null coordinate x^+, these simplify to

$$x^+ = \mathcal{X}^+(t, x), \qquad r = \mathcal{X}^r(t, x). \qquad (2.2)$$

On the long timescales over which the static approximation holds, the metric is approximately

$$ds^2 = -f(r)dx^{+2} + 2dx^+dr + r^2d\Omega_{D-2}^2, \qquad (2.3)$$

with

---

[1]Such modifications must of course have a significant effect—the $\mathcal{O}(1)$ departure of the Page curve [19,20] from the rising entropy of Hawking radiation, as also emphasized in [21].

[2]This is clearest in the example of two-dimensional dilaton gravity models [24,25].

$$f = 1 - \left(\frac{R}{r}\right)^{D-3}. \qquad (2.4)$$

It can be advantageous to take the slices to respect the time-translation symmetry $x^+ \to x^+ + a$, and so take the simplified stationary form [28] [26] [22]

$$x^+ = t + s(x), \qquad r = r(x). \qquad (2.5)$$

The Arnowitt-Deser-Misner (ADM) parameterization [29]

$$ds^2 = -N^2 dt^2 + q_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \qquad (2.6)$$

gives the general form of a time-sliced metric, with normal $n^\mu = (1, -N^i)/N$ to the slices. We will describe the evolving quantum state of the BH using these variables. We consider the small G limit while keeping $G^{1/(D-2)}M$ large to maintain a large black hole in Planck units, and work first to zeroth order in Newton's constant G, and then include perturbative corrections in G.

A. Vanishing G limit

For simplicity we consider scalar matter with Lagrangian

$$S = -\frac{1}{2} \int d^D x \sqrt{|g|} (\nabla \phi)^2; \qquad (2.7)$$

we expect the treatment to generalize to matter with higher spin. In the vanishing G limit we neglect the backreaction of matter fluctuations on the metric. The conjugate momentum to ϕ is

$$\pi = \frac{1}{\sqrt{q}} \frac{\delta S}{\delta \dot{\phi}} = n^\mu \partial_\mu \phi, \qquad (2.8)$$

in terms of which the canonical form of the action becomes

$$S = \int dt d^{D-1} x \sqrt{q} (\pi \dot{\phi} - \mathcal{H}), \qquad (2.9)$$

with Hamiltonian

$$H = \int d^{D-1} x \sqrt{q} \mathcal{H}$$
$$= \int d^{D-1} x \sqrt{q} \left[\frac{N}{2} (\pi^2 + q^{ij} \partial_i \phi \partial_j \phi) + \pi N^i \partial_i \phi\right]. \qquad (2.10)$$

The form of the classical solutions for ϕ plays an important role in quantization. Working in the static background (2.3), these may be separated into two orthogonal spaces [22]. The first are the "in" solutions, $\Phi_{\tilde{A}}(t, x^i)$. These are defined such that as $t \to -\infty$, well-localized wavepacket superpositions are incoming and supported only in the asymptotic large-r region. In the future, they have components that are both outgoing, and that travel towards $r = 0$. (In the 2d case of conformal matter, these are left movers.) The other "up" space consists of modes $\Phi_A(t, x^i)$ such that as $t \to -\infty$ wavepackets become increasingly localized at $r = R$, and outgoing[3]; in the actual limit they become singular at $r = R$. In the future, they also have both outgoing and infalling components. (In the 2d conformal case, these are the right movers.) A basis for these modes can be chosen that is regular at the horizon, but if one instead chooses an energy eigenbasis with respect to a stationary slicing, that basis is singular at $r = R$ and can be furthermore separated into orthogonal subspaces of modes with outgoing component near $r = R$ restricted either to $r < R$ or $r > R$, respectively [22].

Orthogonality of these modes is with respect to the inner product

$$(\Phi_1, \Phi_2) = i \int d^{D-1} x \sqrt{q} n^\mu \Phi_1^* \overset{\leftrightarrow}{\partial}_\mu \Phi_2, \qquad (2.11)$$

which is conserved by the equations of motion. This means it can for example be evaluated at $t \to -\infty$, demonstrating orthogonality of the different subspaces as defined above. Given a time slice, a solution $\Phi(t, x^i)$ corresponds to Cauchy data $\gamma(x^i) = (\phi(x^i), \pi(x^i))$ on that slice, and (2.11) induces the inner product

$$(\gamma_1, \gamma_2) = i \int d^{D-1} x \sqrt{q} (\phi_1^* \pi_2 - \pi_1^* \phi_2) \qquad (2.12)$$

on that Cauchy data, which is also conserved. If we consider the fully extended BH solution (2.3), Cauchy slices can be defined by taking the interior portion of the slices to asymptote to $x^+ = -\infty$, for example as $r \to R_n < R$; otherwise one must confront the question of the form of the dynamics near $r = 0$.

Description of the evolving wavefunction begins with the canonical commutation relations

$$[\pi(x^i, t), \phi(x^{i\prime}, t)] = -i \frac{\delta^{D-1}(x - x')}{\sqrt{q}}. \qquad (2.13)$$

We consider quantization with respect to a time slicing (2.1) or (2.2).[4] Assume that there is a separation of the modes into complex bases (Φ_A, Φ_A^*), $(\Phi_{\tilde{A}}, \Phi_{\tilde{A}}^*)$, with conjugates orthogonal,[5] and such that the modes Φ_A, $\Phi_{\tilde{A}}$ are positive norm under (2.11). Then, in terms of the corresponding Cauchy data on some fiducial $t = $ constant slice, expand

---

[3]For $r < R$, outgoing means asymptotically tracks the "outward" null direction.

[4]Due to additional subtleties of time-dependent metrics [30–34], we typically also restrict to stationary slicings (2.5).

[5]This can be described in terms of a complex structure on the space of solutions.

$$\phi(x^i) = \sum [a_A \phi_A(x^i) + a_A^\dagger \phi_A^*(x^i) + a_{\tilde{A}} \phi_{\tilde{A}}(x^i) + a_{\tilde{A}}^\dagger \phi_{\tilde{A}}^*(x^i)],$$

$$\pi(x^i) = \sum [a_A \pi_A(x^i) + a_A^\dagger \pi_A^*(x^i) + a_{\tilde{A}} \pi_{\tilde{A}}(x^i) + a_{\tilde{A}}^\dagger \pi_{\tilde{A}}^*(x^i)].$$

$$(2.14)$$

If the mode bases are taken to be orthonormal, $(\gamma_A, \gamma_B) = (\gamma_{\tilde{A}}, \gamma_{\tilde{B}}) = \delta_{AB}$, the canonical commutators become

$$[a_A, a_B^\dagger] = [a_{\tilde{A}}, a_{\tilde{B}}^\dagger] = \delta_{AB},$$

$$[a_A, a_B] = [a_A^\dagger, a_B^\dagger] = [a_{\tilde{A}}, a_{\tilde{B}}] = [a_{\tilde{A}}^\dagger, a_{\tilde{B}}^\dagger] = 0, \quad (2.15)$$

along with vanishing commutators between in and up operators, and Fock space states may be constructed by acting with raising operators on the vacuum $|0\rangle$ annihilated by a_A and $a_{\tilde{A}}$. In occupation number basis, these take the form $|\{n_A\}, \{n_{\tilde{A}}\}\rangle$.

For a "gauge-fixing" specified by a given slicing, determining (N, N^i, q_{ij}), the Hamiltonian (2.10) can be rewritten in terms of the annihilation and creation operators by substitution of (2.14).[6] This gives an expression

$$H = \sum_{A,A'} (A_{AA'} a_A^\dagger a_{A'} + B_{AA'} a_A^\dagger a_{A'}^\dagger + \text{c.c.}) + (A \to \tilde{A}), \quad (2.16)$$

where c.c. denotes conjugation that does not change operator ordering, and there is no mixing between A and $\tilde{A}$ modes. The coefficients $A_{AA'}$, $B_{AA'}$, $A_{\tilde{A}\tilde{A}'}$, $B_{\tilde{A}\tilde{A}'}$ may be expressed in terms of the mode basis and the metric coefficients, with explicit expressions given in [26,27] [22]. A specified initial state is then evolved by

$$U(t_2, t_1) = e^{-i \int_{t_1}^{t_2} dt H}. \quad (2.17)$$

In addition to standard propagation described by $a^\dagger a$ terms, H creates paired excitations through the $a^\dagger a^\dagger$ terms. One way to understand the detailed structure of this pairing is to rewrite the state in a basis of energy eigenmodes, as described above. Then the Hamiltonian takes the form [22]

$$H = \sum_{lm} \int \frac{d\omega}{4\pi\omega} \omega (b_{\omega lm}^\dagger b_{\omega lm} - \hat{b}_{\omega lm}^\dagger \hat{b}_{\omega lm} + \tilde{b}_{\omega lm}^\dagger \tilde{b}_{\omega lm}),$$

$$(2.18)$$

up to a normal-ordering constant, with the annihilation operators $b_{\omega lm}$, $\hat{b}_{\omega lm}$, $\tilde{b}_{\omega lm}$ corresponding to external up modes, internal up modes, and in modes, respectively. Such a basis is singular, and makes contact with Hawking's original arguments [10]. In such a basis, a regular state such

as $|0\rangle$ has pairing between excitations on either side of the horizon, with thermal coefficients; at high wavenumbers such a regular state has behavior (specializing to $D = 4$)

$$|\psi\rangle \sim \sum_{\{n_{\omega lm}\}} e^{-2\pi R \int d\omega \omega n_{\omega lm}} |\{\hat{n}_{\omega lm}\}\rangle |\{n_{\omega lm}\}\rangle. \quad (2.19)$$

The diagonalized Hamiltonian (2.18) means its evolution is trivial, with inside modes having negative energies. If one works in a wavepacket basis of these modes, one can follow an outside mode into the asymptotic region, where it can alternately be described in terms of a regular mode. This mode is paired with internal excitations, through (2.19); more discussion of this argument appears in [22]. All initially regular states are argued to have the same late time behavior; differences correspond to transitory excitations that fall deep into the BH or escape to infinity, and continuing long-time production of paired excitations then follows from the uniform vacuumlike structure of such states at short distances near $r = R$.

In principle, one should be able to see much of this structure directly in terms of a regular basis of modes. Since localization plays an important role, we can imagine choosing bases Φ_A, $\Phi_{\tilde{A}}$ that are approximately localized both in space and in wavenumber. One such basis comes from window functions in frequency [10,36], although one may wish to optimize localization even further by using other bases, for example a wavelet basis. In addition, with such a basis of localized modes (modulo some rapidly falling tails), one may separate the up modes Φ_A into sets $\{\Phi_a\}, \{\Phi_{\hat{a}}\}$ of modes dominantly localized in the regions $r > R$ and $r < R$ respectively. In such a basis it is in practice much harder to evaluate the A and B coefficients in the Hamiltonian, and the unitary evolution operator (2.17), although in principle they can be calculated. But due to the approximate correspondence of these wavepackets to those of the singular basis in regions well away from the horizon, we should approximately have the same entanglement structure as in (2.19), with approximately thermal weights, when the outside modes are at $r \gg R$; the corresponding inside modes will then typically be deep inside the BH. Of course, in the region $r \sim R$ this description of the state differs significantly—in the regular basis, the state looks like the vacuum for modes very close to the horizon [26,27].

This discussion should also straightforwardly generalize to other types of matter, with spin. The choice of the mode basis of course becomes a little more intricate, and one may also need to specify some appropriate gauge fixing. But otherwise, one may similarly canonically quantize, and for a free field, find a Hamiltonian with the same structure as (2.16), resulting in a similar structure for the evolving quantum state, with particle production and internal/external pairing, and approximately thermal weights. This includes

---

[6]Different such descriptions of the state arising from different gauge fixings are then expected to be related by unitarity transformations, with an analogy to change of picture. One approach to arguing for such equivalence is expected to be based on the Hadamard behavior of the state [35].

the case of spin two perturbations, corresponding to fluctuations of the metric.

The approach described here also has the virtue of readily (in principle) generalizing to the case of interacting fields. This is in contrast to conventional derivations of Hawking radiation. There, modes are traced back via free evolution to the near-horizon region, in order to match onto the short-distance behavior there, and this procedure requires significant modification in an interacting theory. In the present treatment using regular modes, interactions simply lead to higher-order terms in (2.16), whose contribution to evolution through (2.17) may in principle be followed. Of course this is no simpler than the description of any other interacting QFT system, but does have the virtue of being placed on a foundation involving regular rather than singular modes.

B. Leading $\mathcal{O}(G)$ corrections: Dressing and backreaction

While the full quantum description of the evolving gravitational state remains somewhat mysterious, one can apparently characterize the perturbative behavior of its small fluctuations, for example to leading order in G, working in QGR. The starting point is the action

$$S = \int d^D x \sqrt{|g|} \left(\frac{1}{16\pi G} R + \mathcal{L}_m \right) + S_\partial \qquad (2.20)$$

where $\mathcal{L}_m$ is the matter Lagrangian and S_∂ is a surface term. The conjugate momenta to q_{ij} are

$$P^{ij} = \frac{\delta S_g}{\delta \dot{q}_{ij}} = -\frac{\sqrt{q}}{16\pi G} (K^{ij} - q^{ij} K), \qquad (2.21)$$

given in terms of the extrinsic curvature

$$K_{ij} = \frac{1}{2N} (-\dot{q}_{ij} + D_i N_j + D_j N_i) \qquad (2.22)$$

of the slices, with D_i the covariant derivative with respect to q. The conjugate momenta to N and N^i vanish, corresponding to their role as Lagrange multipliers for constraints. These result in the canonical form for the gravitational part of the action

$$S_g = \int d^D x (P^{ij} \dot{q}_{ij} - \mathcal{H}_g). \qquad (2.23)$$

The full Hamiltonian may be written in two ways,[7] as

$$H = \int d^{D-1} x (\mathcal{H}_g + \mathcal{H}_m) = \int d^{D-1} x (N \mathcal{C}_n + N^i \mathcal{C}_i) + H_\partial. \qquad (2.24)$$

Here $\mathcal{H}_g$ is quadratic in momenta and spatial derivatives of the metric variables. In the second expression, $\mathcal{C}_n, \mathcal{C}_i$ are the constraints,

$$\mathcal{C}_\mu = \sqrt{q} \left(-\frac{G_{\mu\nu}}{8\pi G} + T_{\mu\nu} \right) n^\nu, \qquad (2.25)$$

with $\mathcal{C}_n = n^\mu \mathcal{C}_\mu$. H_∂ is a boundary term containing the ADM energy and momentum, and when the constraints are satisfied, $\mathcal{C}_\mu = 0$, the Hamiltonian equals this boundary term.

There are different approaches to quantization; here we take that of "gauge-invariant canonical quantization," which is closely related to covariant gauge fixing (breaking) [23]. The canonical commutators are taken to be

$$[P^{ij}(x, t), q_{kl}(x', t)] = -i\delta^i_{(k} \delta^j_{l)} \delta^{D-1}(x - x'), \qquad (2.26)$$

with $\delta^i_{(k} \delta^j_{l)} = (\delta^i_k \delta^j_l + \delta^i_l \delta^j_k)/2$, and evolution is generated by the Hamiltonian, as summarized by the Heisenberg equations,

$$\partial_t q_{ij} = i[H, q_{ij}], \qquad \partial_t P^{ij} = i[H, P^{ij}]. \qquad (2.27)$$

Since the constraints generate gauge transformations, gauge-invariant operators are those that commute with the constraints,

$$[\mathcal{C}_\mu(x), O] = 0. \qquad (2.28)$$

However, the constraints should not annihilate physical states, as that would be inconsistent with the equations of motion (2.27). Instead, the constraints are decomposed into pieces

$$\mathcal{C}_\mu(x) = \mathcal{C}_\mu^+(x) + \mathcal{C}_\mu^-(x), \qquad (2.29)$$

corresponding for example to positive and negative frequency contributions, and physical states are taken to obey

$$\mathcal{C}_\mu^+(x)|\psi\rangle = 0. \qquad (2.30)$$

Given a vacuum state satisfying (2.30), other physical states may then be constructed by acting with a gauge-invariant operator, $|\psi\rangle = O|0\rangle$. With such states, evolution of matrix elements of gauge-invariant operators,

$$\partial_t \langle \psi'|O|\psi\rangle = i\langle \psi'|[H, O]|\psi\rangle, \qquad (2.31)$$

is gauge invariant, i.e. independent of N, N^i.

---

[7]For more detailed expressions, see [23].

The gauge-invariance conditions (2.28), (2.30) may be solved perturbatively by constructing a gravitational dressing [37–41] [14,15].[8] Defining $\kappa^2 = 32\pi G$, the metric variables are expanded about a background metric as

$$\tilde{g}_{\mu\nu} = g_{\mu\nu} + \kappa h_{\mu\nu}, \quad \tilde{q}_{ij} = q_{ij} + \kappa h_{ij}, \quad \tilde{P}^{ij} = P^{ij} + \frac{p^{ij}}{\kappa},$$

(2.32)

and the commutators (2.26) become

$$[p^{ij}(x,t), h_{kl}(x',t)] = -i\delta^i_{(k}\delta^j_{l)}\delta^{D-1}(x-x').$$ (2.33)

The constraints may then be expanded in κ, as described in [23]. For a given operator O_0 of the underlying field theory (here, of ϕ), a gauge-invariant solution to (2.28), to first order in κ, takes the form

$$O = e^{i\int d^{D-1}x\sqrt{q}V^\mu(x)T_{n\mu}} O_0 e^{-i\int d^{D-1}x\sqrt{q}V^\mu(x)T_{n\mu}};$$ (2.34)

this may also be generalized to include dressing of gravitational perturbations by including a gravitational pseudostress tensor contribution to the stress tensor $T_{n\mu}$. Here $V^\mu(x)$ are functionals of the metric perturbation. Leading order gauge invariance follows if these functionals take the form

$$V^n(x) = \frac{\kappa}{2}\int d^{D-1}x'\left[\check{h}_{ij}(x',x)p^{ij}(x') - \check{p}^{ij}(x',x)h_{ij}(x')\right],$$

$$V^i(x) = \kappa\int d^{D-1}x'\left[G^{ijk}(x',x)h_{jk}(x') + H^i_{jk}(x',x)p^{jk}(x')\right]$$

(2.35)

where $\check{h}_{ij}$, $\check{p}^{ij}$, G^{ijk}, H^i_{jk} are Green's function solutions of the linearized metric perturbation problem, detailed in [23]. These are not uniquely specified, in the absence of boundary conditions: to any such solution, we can add a solution of the homogeneous equations, corresponding to a freely propagating gravitational wave. This leads to a corresponding nonuniqueness in the gravitational dressing.

A matter state $|\psi\rangle_0$ (or one of perturbative gravitons) may be likewise dressed

$$|\psi\rangle = e^{i\int d^{D-1}x\sqrt{q}V^\mu(x)T_{n\mu}}|\psi\rangle_0,$$ (2.36)

and will solve (2.30) if the underlying state $|\psi\rangle_0$ is also taken to solve the $\kappa = 0$ version of equation (2.30).

---

[8]Earlier work on this question includes [42] and [43]. The first derived nontrivial commutators are as arising from the constraints but did not give the dressed operators, and the second focused on deriving *commuting* bulk operators.

The dressing in (2.36) has the anticipated effect of creating the linearized quantum gravitational field arising from the stress energy of the underlying state. Again, this is not unique since it can be augmented by freely propagating gravitational waves. Examples of this nonuniqueness are provided in [38], where one may for example consider initial conditions where the gravitational field configuration is the analog of the Coulomb field, or is instead concentrated in a single gravitational line stretching to infinity.

Evolution of an initial dressed state (2.36) will then be governed by the Hamiltonian (2.24). The leftmost form of this Hamiltonian exhibits the terms responsible for evolving the metric perturbation and the matter perturbation, in the standard LQFT fashion. For example, an initial linelike gravitational field like that just described will transition to an energetically favorable field like that of Coulomb, with gravitational radiation escaping to infinity. This treatment of the leading gravitational field thus also describes the leading quantum backreaction of the underlying matter state, that is it accounts for the leading correction to the gravitational background $g_{\mu\nu}$ driven by the stress tensor of that matter state.

C. General features of the perturbative quantum state

Beginning with an initial state $|\psi(0)\rangle$, the perturbatively evolved $G = 0$ Fock space state can be expanded in the general form

$$|\psi(t)\rangle = \sum_{\{n_a\}\{n_{\hat{a}}\}\{n_{\tilde{A}}\}} \mathcal{A}(\{n_a\}\{n_{\hat{a}}\}\{n_{\tilde{A}}\}, t)|\{n_a\}\{n_{\hat{a}}\}\{n_{\tilde{A}}\}\rangle;$$

(2.37)

for example, by relating this to the evolution of (2.19), as described in Sec. II A, we find entanglement between the internal $a^\dagger_{\hat{a}}$ and external $a^\dagger_a$ excitations that grows with time. This can be characterized by introducing an appropriate regulator, e.g. cutting off near horizon excitations with wavelengths $\ll R$, and tracing out internal excitations to find a density matrix with a growing von Neuman entropy.[9] Continued growth of this during the BH evolution to a value comparable to the Bekenstein-Hawking entropy S_{BH} is evidence of the missing information.

The leading perturbative correction to this arises from evolving the dressed version (2.36) of the state (2.37) by the Hamiltonian (2.24); as described, this evolution incorporates the gravitational constraints. The gravitational dressing *does* introduce additional subtleties in giving a precise characterization of the BH subsystem and calculation of its entropy. An open question is how to sharply characterize gravitational subsystems [14–16], taking into account such effects at this and higher orders in κ. It has even been argued that the dressing eliminates the missing

---

[9]See e.g. [36] for a simple two-dimensional example.

information [17,44]. However, it can be seen that the effects of the leading perturbative dressing at best give access to the missing information that is exponentially suppressed with distance [16], and there does not appear to be a controlled approximation in which an asymptotic observer can access internal information, either with Minkowski or AdS asymptotics. Moreover, this evolution does not appear to significantly alter the trans-horizon entanglement evident in (2.37), nor does incorporating the leading semiclassical backreaction arising from $\langle T_{\mu\nu} \rangle$.[10]

This discussion is of course expected to also readily extend to other kinds of matter fields coupled to gravity. Moreover, treatment of evolution using a Hamiltonian of the form (2.10), (2.16) readily generalizes to incorporate interactions between fields, which e.g. introduce higher-order terms in (2.16).

The evolution can be described as producing outgoing Hawking quanta at a rate $\mathcal{O}(1/R)$, for each species of particle. Correspondingly, the growth in the entanglement or von Neumann entropy of the BH grows also at a rate $\mathcal{O}(1/R)$, and so grows to a value $\sim S_{\mathrm{BH}} = A/4G$ by the time a significant fraction of the initial BH mass has been lost.

III. MODIFICATION OF LQFT/QGR EVOLUTION

The "missing information" internal to the BH thus appears to persist in the presence of at least the leading-order gravitational corrections that are described by combined evolution via LQFT and QGR. If so, that returns us to the apparent inconsistency that has been previously explored: either the information is just lost [4], ultimately resulting in a quantum description which unacceptably conflicts with energy conservation [5], or a BH remnant remains, apparently inconsistent with basic stability [6–8] and other [9] constraints. These problems thus strongly suggest that there must be some modification to the preceding description of growth of missing information, which is characterized by growing entanglement entropy.

Taking a step back [41,45], if we assume that the theory is fundamentally quantum mechanical, then we need a description of the quantum state and its evolution that is fundamentally unitary. The question is what form this state and its evolution take, and how those depart from the LQFT/QGR description.

One feature of the basic kinematic description of many other quantum systems is division of those into quantum subsystems. As briefly described above, the question of a precise such division in quantum gravity remains. However, at least at perturbative level an approximate such division appears possible, and moreover there is no obvious indication that perturbative QGR corrections to this approximation are responsible for solving the missing

information problem. In particular, we have outlined the leading LQFT/QGR description of such evolution in the preceding section, and have seen that if the leading gravitational dressing is included, one still has a description of the state with an approximate subsystem division, and consequent buildup of entanglement.

Thus, if there is a meaningful approximate division into subsystems, one is led [13] to consider modifications to the evolution of a combined system of a BH and its environment. One feature of the LQFT/QGR description is that it can be given in different gauges, and so a specific description involves choosing a particular gauge, represented by choice of slicing and spatial coordinates [27]. One might likewise anticipate that the more fundamental description has analogous gauge freedom, and likewise describe that in a specific gauge.

Given the wide-ranging success of LQFT and GR, extending both into the regime of predictions from early cosmology and now strong gravitational fields of black holes, it seems plausible that the more fundamental description involves departures from these that are small in most circumstances, in particular plausibly including in describing observations of observers falling into a BH. Thus, we seek to parameterize the more complete dynamics in terms of corrections to LQFT/QGR that for most purposes remain small.

One possible correction to the LQFT/QGR description is the presence of new quantum interactions between a BH and its surroundings, which are capable of transferring information, yet have limited effect on observers near or falling into the BH. This is the basic idea of "nonviolent unitarization."

Another recent idea [12] is that there is a new form of evolution of the BH and its surroundings, which acts in a nonunitary (or nonisometric) fashion on the degrees of freedom of the LQFT description, yet preserves ultimate unitarity of scattering. Reference [12] advocates that this be understood in terms of a map to a "holographically dual" theory, on whose Hilbert space the evolution is unitary.

In either case, we do not yet have complete models, but we do have models capturing some of the important features. One approach to understanding the basics of such models, and in particular some essential questions of how information flows or transfers, is to work with simplified qubit models. We next turn to their description, and comparison, before returning to the question of evolution in a more complete model that matches onto the evolution of LQFT in familiar regimes.

IV. QUBIT MODELS FOR EVOLUTION

An approach to describing important features of black hole evolution is to simplify, and describe qubit (or more generally qudit) models for their states. An early such description of Hawking evolution was given in [21], based partly on the simple form of the more complete Hawking

---

[10]For some investigation of the latter in two-dimensional models, see [36].

state in two dimensions [36]. Qubit models modifying Hawking's description then have been considered in [21,46–48] and other subsequent works.

A. Hawking evolution

Instead of the states $|\{n_a\}\{n_{\hat{a}}\}\{n_{\tilde{A}}\}\rangle$ of the full G^0 Fock space, one can model Hawking evolution in terms of qubit states, at the same time discretizing time evolution in terms of steps, $T = NR$, for integer N. The pairwise emergence of inside and outgoing excitations of (2.19), (2.37) can be described by the qubit state

$$|\psi(T)\rangle = \frac{1}{2^{N/2}} \prod_{i=1}^{N} (|\hat{0}\rangle|0\rangle + |\hat{1}\rangle|1\rangle)_i, \qquad (4.1)$$

which lives in the Hilbert space $\hat{\mathcal{H}}_T \otimes \mathcal{H}_T = \hat{\mathcal{Q}}^{\otimes N} \otimes \mathcal{Q}^{\otimes N}$, where $\mathcal{Q}$ denotes a basic qubit Hilbert space. We may include infalling matter by also tensoring in a space $\tilde{\mathcal{H}}_T = \hat{\tilde{\mathcal{H}}}_T \otimes \tilde{\mathcal{H}}_T^{env}$. Here the factorization at time T is that corresponding to describing an excitation as inside the BH, or in its exterior environment, and each of the factors is likewise a product of qubit spaces $\tilde{\mathcal{Q}}$.

Description of these states and their evolution can be facilitated by introducing fermionic annihilation/creation operators ψ_i, $\psi_i^\dagger$, satisfying the basic relations

$$|1\rangle_i = \psi_i^\dagger|0\rangle_i, \quad |0\rangle_i = \psi_i|1\rangle_i, \quad \psi_i|0\rangle_i = \psi_i^\dagger|1\rangle_i = 0, \quad (4.2)$$

and likewise for the hatted and tilded spaces.

A model for the energy-eigenbasis Hamiltonian (2.18) is

$$H_0 = \sum_i \omega_0 (\psi_i^\dagger \psi_i - \hat{\psi}_i^\dagger \hat{\psi}_i), \qquad (4.3)$$

with $\omega_0 \sim 1/R$ a fixed energy of the Hawking quanta. This is supplemented by a "left-moving" Hamiltonian $\tilde{H}_0$ which translates qubits in $\tilde{\mathcal{H}}$. The Hamiltonian H_0 annihilates the Hawking state (4.1). However, the form of the state indicates that at each stage an additional pair of entangled qubits is added. Alternately, one could, as with (2.19), think of those factors as being present throughout the evolution, but inaccessible to low-energy observables until after timestep $N = i$.

The growing number of accessible qubits leads to growing entanglement of the environment of the BH with its internal states, and this entanglement is part of the description of the missing information. As the model makes clear, and as holds in more general BH evolution beyond such models, there is a second source of BH entanglement: one may entangle an excitation in $\tilde{\mathcal{H}}$ with another outside excitation, and then allow the first to evolve into the interior space via $\tilde{H}$. Then, if the BH evaporates and disappears, both types of entanglement are associated to problematic

unitarity violation. We refer to these as type I and type II entanglement, respectively. (Interactions in more complete models may mix these.)

The large entropy associated with type I entanglement may be illustrated by introducing a new basis of states $|K\rangle_N \in \mathcal{Q}^{\otimes N}$:

$$|K\rangle_N = |0\rangle|0\rangle \cdots |0\rangle \quad , \quad |0\rangle|0\rangle \cdots |1\rangle \quad , \quad \ldots \quad , \quad |1\rangle|1\rangle \cdots |1\rangle. \qquad (4.4)$$

In terms of these $\mathcal{N}_N = 2^N$ states, and hatted counterparts, the Hawking state may be written

$$|\psi(T)\rangle = \frac{1}{2^{N/2}} \sum_K^{\mathcal{N}_N} |\hat{K}\rangle_N |K\rangle_N. \qquad (4.5)$$

The corresponding von Neumann entropy of the exterior density matrix is $N \ln 2$.

The growth in both types of entanglement during evolution is a central feature of the information problem of Hawking evolution; in order to instead describe unitary evolution, in which the BH is permitted to disappear at the end, the entanglement should begin to decrease as the black hole loses energy, reaching zero at the end—as in the case of the cavity described in the Introduction. An early description of this behavior appeared in the works [19,20]. We next explore different kinds of models for such reduction in the entanglement.

B. Nonisometric evolution

Recent work [12] has given qubit models for BH evolution that are based on a certain form of nonisometric evolution. In order to better understand the meaning and implications of these models, and their comparison with other kinds of models, we will connect them to the language used in this paper to describe real-time evolution.

The models of [12] are based on the idea that there are two "holographically equivalent" descriptions that they call the "effective" and "fundamental" descriptions. Since our focus will be on the description of a "bulk" observer— which e.g. could be a detector falling into a BH, and we will be agnostic about the meaning and origin of a possibly more fundamental description, we will simply describe these in terms of Hilbert spaces $\tilde{\mathcal{H}}$ and $\mathcal{H}'$, respectively, referring to the latter as the prime description, and exploring the relation between them.

Specifically, we examine the time evolution of the state between timesteps, as a modification to the Hawking evolution described above. Before modification, the state at time T is taken to live in the Hilbert space

$$\bar{\mathcal{H}}_T = \hat{\mathcal{H}}_T \otimes \hat{\tilde{\mathcal{H}}}_T \otimes \mathcal{H}_T \otimes \tilde{\mathcal{H}}_T^{env}, \qquad (4.6)$$

where again $\hat{\mathcal{H}}_T \otimes \tilde{\mathcal{H}}_T$ describe the ingoing and outgoing internal states of the BH, respectively, and the evolution was described in the preceding section.

Reference [12] modifies this evolution according to the following description, after translating from their language. First, at each timestep one qubit is shifted from the space $\tilde{\mathcal{H}}_T^{env}$ to join $\tilde{\mathcal{H}}_T$. Then, a unitary U_T acts on the latter combined space; different scenarios are considered for how strongly this scrambles the states. At the same time, two new pairs of qubits are appended to $\hat{\mathcal{H}}_T \otimes \mathcal{H}_T$, in the Hawking state

$$\frac{1}{2} \left(|\hat{0}\rangle|0\rangle + |\hat{1}\rangle|1\rangle \right)_{2N+1} \left(|\hat{0}\rangle|0\rangle + |\hat{1}\rangle|1\rangle \right)_{2N+2}. \quad (4.7)$$

Finally, a projection operator

$$P_T = 2 \left(\langle \tilde{0}|_{N+1} \langle \hat{0}|_{2N+1} + \langle \tilde{1}|_{N+1} \langle \hat{1}|_{2N+1} \right)$$
$$\times \left(\langle \tilde{0}|_N \langle \hat{0}|_{2N+2} + \langle \tilde{1}|_N \langle \hat{1}|_{2N+2} \right) \quad (4.8)$$

acts on two qubits of the output of U_T and on the two internal qubits that were added. The net result is a non-isometric map M_T from the space

$$\tilde{\mathcal{H}}_T \otimes \tilde{\mathcal{Q}} \otimes \hat{\mathcal{Q}}^{\otimes 2} \otimes \mathcal{Q}^{\otimes 2} \quad (4.9)$$

to the space

$$\tilde{\mathcal{H}}_{T+R} \otimes \mathcal{Q}^{\otimes 2}, \quad (4.10)$$

where $\tilde{\mathcal{H}}_{T+R}$ has one fewer qubit than $\tilde{\mathcal{H}}_T$, and where $\mathcal{Q}^{\otimes 2}$ augments the outgoing radiation space $\mathcal{H}_T$.

Reference [12] provides additional interpretation for this by defining the Hilbert space ("fundamental" Hilbert space) $\mathcal{H}_T' = \tilde{\mathcal{H}}_{T+R}$. Then, evolution is described as the unitary U_T acting on $\mathcal{H}_{T-R}' \otimes \tilde{\mathcal{Q}}$, followed by identifying two of the output qubits as radiation qubits, to map to the space $\mathcal{H}_T' \otimes \mathcal{Q}^{\otimes 2}$ in (4.10). As part of this interpretation, they define a "holographic map" V_T. This is related to the preceding by first defining an intermediate "holographic map" v_T from $\tilde{\mathcal{H}}_T \otimes \tilde{\mathcal{Q}} \otimes \hat{\mathcal{Q}}^{\otimes 2}$ to the space $\mathcal{H}_T'$; v_T arises from combining U_T with the operator P_T of (4.8). Then, V_T is defined iteratively by $V_T = v_T v_{T-1} \cdots v_0$.

We next compare this evolution with Hawking evolution, to investigate its features. A first small difference is that two radiation quanta are emitted at each step, while one infaller enters. A larger difference is that the evolution does not build up type I entanglement, as a result of the projection P_T; it also transfers the type II entanglement out of the BH. Correspondingly, there is no increase in the dimension of $\hat{\mathcal{H}}_T$ at each step.

It does accomplish this at the price of introducing fundamentally nonlocal evolution: the projection P_T acts in a nonlocal way on the quantum state. The combined evolution M_T is then both nonlocal and nonunitary on the underlying Hilbert space corresponding to the QFT description. This forces the type II entanglement out of the BH in a different way than entanglement escapes via interactions in systems like the cavity discussed in the Introduction.

To explore this further, note that M_T (and correspondingly the "holographic map" v_T) has null states. The first factor in P_T annihilates the two states $|\hat{0}\rangle_{N+1}|\hat{1}\rangle_{2N+1}$ and $|\tilde{1}\rangle_{N+1}|\hat{0}\rangle_{2N+1}$, and similarly for the second. Null states in evolution seemingly present a "nowhere to go" problem: a branch of the wavefunction evolves to nothing.

One might think this might be ameliorated by the fact that it doesn't restrict the infalling matter, but only states of the Hawking quanta, which are not input from the exterior. Specifically, if the internal evolution U_T outputs $|\tilde{0}\rangle_{N+1}$ the projection requires that the outgoing quanta be in the state $|\hat{0}\rangle_{2N+1}|0\rangle_{2N+1}$; if it outputs $|\tilde{1}\rangle_{N+1}$, it requires $|\hat{1}\rangle_{2N+1}|1\rangle_{2N+1}$.

This does seem to raise potential issues, however, when extended to a more complete model. A first one is a potential firewall problem. Once the Hawking state is projected as in (4.8), it no longer has the entanglement structure needed for infalling observers to see vacuumlike behavior. So, if this step in the evolution acts while excitations are near the horizon, it will yield high-energy excitations as seen by those observers. Possibly, this could be avoided as in nonviolent unitarization, if the evolution M_T acts once the Hawking quanta have reached large, e.g. $\mathcal{O}(R)$ wavelengths, but that does also extend the nonlocality and nonunitarity over those scales.

But beyond that, the internal "outgoing" modes, modeled by the qubits $\hat{\mathcal{Q}}$, can participate in interactions in a more complete theory, raising the question of the role of the projections in that context. A simple example (see [27] for discussion) is decay behind the horizon of a massive infalling particle into lighter particles, one of which is inward going and one of which is outward going, which will then populate the modes corresponding to $\hat{\mathcal{Q}}$. Then, there is the question of how to describe consistent evolution, in view of the "nowhere to go" problem. Or, more picturesquely, an infalling observer could fire their rocket engines after crossing the horizon, in a vain attempt to escape. They would not do so, but could become an internal outward mover. This again raises the question of a consistent extension of the projections (4.8), acting on outward movers. Additional questions raised by interactions, e.g. regarding ultimate unitarity, have also been recently considered in [49].

C. Nonviolent unitarization

The approach of nonviolent unitarization (NVU) [46,50–52] [28,53] considers interactions between the BH Hilbert space and the environment, which can transfer

information or entanglement out of the BH, but in a way that "minimally" affects the standard description of BHs, e.g. as described by infalling or other exterior observers. Accomplishing the information transfer via interactions is similar to the physics by which the cavity described in the introduction reduces its entanglement.

Qubit models for NVU are found by augmenting the Hamiltonian (4.3): $H = H_0 + \Delta H$, with two additional pieces. The first is a term ΔH_I which describes the interactions between BH and environment states, and so can transfer entanglement from BH to environment. The second is a term ΔH_{bh} which alters the evolution of the internal state of the BH. The internal part of the state (4.1) has a "frozen" character, in that the internal excitations simply persist unchanged; similar behavior is found for more complete states (2.37) based on nice slicings that avoid the singularity [26,27] [22]. This is a feature of a particular gauge or picture for describing the state, which artificially retards evolution into the region near $r = 0$. In a more complete quantum description of the interior evolution, it is quite plausible that the natural dynamics ultimately leads to nontrivial evolution; e.g. it might begin mixing or scrambling the internal state on some timescale. ΔH_{bh} is introduced to describe this aspect of evolution, which is analogous to mixing of internal states in the cavity described in the Introduction.

At the level of detailed qubit models, there are multiple possibilities. ΔH_{bh} could induce general transitions between BH states $|\hat{K}\rangle$, as well as mixing with the infalling states $|\tilde{K}\rangle$. A particularly simple possibility is if ΔH_{bh} conserves H_0. It could couple multiple internal qubits, or even have a form as simple as $\Delta H_{bh} = h_{ij}\hat{\psi}_i^\dagger\hat{\psi}_j$ and/or $\Delta H_{bh} = h_{ij}\hat{\psi}_i\tilde{\psi}_j$, rearranging qubits. We will also assume that there is a "latency time" $\Delta T = R\Delta N$, such that $\Delta H_{bh}(T)$ only acts nontrivially on the first $N - \Delta N$ qubits; before the latency time, the internal qubits in (4.1) are unscrambled. Different forms for ΔH_{bh} may also follow from different choices of gauge to describe internal evolution.

The transfer term ΔH_I—which would be forbidden by locality in a QFT description—plays the key role of transferring entanglement to the environment, and the structure of this term is particularly interesting since it bears on how the BH appears from the outside. The simplest form for such a term is

$$\Delta H_I \sim \gamma_{\hat{\alpha}\alpha}\hat{\mathcal{O}}_{\hat{\alpha}}\mathcal{O}_\alpha \qquad (4.11)$$

with some coefficients $\gamma_{\hat{\alpha}\alpha}$, and with general operators acting on the internal excitations and external excitations, respectively; if one includes infalling matter, there could likewise be terms with internal operators $\tilde{\mathcal{O}}_{\tilde{\alpha}}$. These two kinds of terms can then transfer type I or type II entanglement, respectively, to the BH's environment.

For a simple model, consider the case where ΔH_I couples to a single qubit in the interior, $\hat{\mathcal{O}} = \hat{\psi}_i^\dagger$. Recall from (4.3) that $\psi_i^\dagger$ lowers the BH energy; a particularly simple model would have interactions $\gamma_{ij}\hat{\psi}_i^\dagger\psi_j^\dagger$, conserving energy, although we will also consider coupling to a more general external operator. We can also generalize these interactions to couple to the operators $\tilde{\psi}_i$ in the interior.

To model the possibly delayed turn on of such interactions, append a qubit $\mathcal{Q}_0$ to the qubits of the Hawking state (4.1), and consider concretely

$$\Delta H_I = \gamma\hat{\psi}_0^\dagger\mathcal{O} + \text{H.c.} \qquad (4.12)$$

for some exterior operator $\mathcal{O}$. Then, assume that ΔH_{bh} acts to mix $\mathcal{Q}_0$ and qubits with $i = 1, \cdots N - \Delta N$, including the latency time parameter ΔT. After this time, the qubits of the Hawking state (4.1), mix with $\mathcal{Q}_0$; the interaction (4.12) is then sensitive to their state, and can transfer their information into the state of the environment via the exterior operator $\mathcal{O}$.

Key questions, then, are whether ΔH_I transfers information at a sufficient rate to address the unitarity problem, what effects it has on dynamics in the vicinity of the BH, and what effect dynamics has on observers who have fallen in.

Estimating the rate at which (4.12) induces transitions gives an estimate of the rate at which it transfers information (more precisely, entanglement) [28,54]. Consider its action on a state which is a general (for simplicity) N-qubit state $|\psi_{bh}\rangle$ of the interior, and $|\psi_e\rangle$, which could be vacuum, for the exterior. Transition rates to different final states can be estimated from the matrix elements

$$\gamma\langle\psi'_{bh}|\psi_0^\dagger|\psi_{bh}\rangle\langle\psi'_e|\mathcal{O}|\psi_e\rangle \qquad (4.13)$$

of ΔH_I.

Suppose that $|\psi_{bh}\rangle = \sum_K c_K|\hat{K}\rangle$ is a generic unit-normalized BH state, arising from the evolution with $H_0 + \Delta H$ acting on the Hawking state (4.5), and projecting on one of the internal components. Then, the matrix element to a generic final state is of the same size as the generic sizes of the c_K,

$$\langle\psi'_{bh}|\psi_0^\dagger|\psi_{bh}\rangle \sim \frac{1}{2^{N/2}} = e^{-S_{bh}/2}. \qquad (4.14)$$

Here we have defined a black hole entropy by $\exp\{S_{bh}\} = 2^N$, for comparison with other discussions. While this effectively weak coupling to a generic state suggests a tiny transition rate, there are $\mathcal{O}(2^N) = e^{S_{bh}}$ final states which one can transition to. So, an argument based on Fermi's golden rule [28] suggests that if the matrix element of the operator $\gamma\mathcal{O}$ in (4.13) is $\mathcal{O}(1/R)$, the total transition rate, and hence the information transfer rate, will be the same size, $\mathcal{O}(1/R)$. This, in order of magnitude, is the needed size

to compensate for the growth of type I entanglement due to Hawking evolution, and in a more general scenario, can also transfer any type II entanglement that has been injected into the BH.

This rate can be explained from a different argument. For generic qubit state $|\psi_{bh}\rangle$, $\psi_0^\dagger$ acts to create *some* state with $\mathcal{O}(1)$ norm. Then, the transition rate to that state will not be exponentially suppressed.

This kind of argument—which can be refined[11]—sets the size of the couplings to the operators $\mathcal{O}$, which we think of as acting on a "quantum atmosphere" of the BH. The same methods can be used to estimate the amplitude if the initial external state $|\psi_e\rangle$ is not vacuum, but instead something else, e.g. an incoming gravitational wave state [53]; that just changes which matrix element between external states enters in (4.13). In either case, the "nonviolence" assumption [46,50–52] [28,53] is that the operator $\mathcal{O}$ only transfers soft momenta and energies to the external state, e.g. with $\mathcal{O}(1/R)$ sizes, although more general powers of R can also be considered. In qubit models, this is naturally implemented since the qubit excitations have energies $\sim 1/R$, as long as $\mathcal{O}$ only acts on qubits that have begun to exit the horizon region. The final result is that the interactions (4.12) not only yield entanglement transfer rates of size $\mathcal{O}(1/R)$, but also can give $\mathcal{O}(1)$ modifications to scattering amplitudes which have energy and momentum transfers $\mathcal{O}(1/R)$.

One can ask about other effects besides corrections to these amplitudes, e.g. on an infalling external observer. From this outside perspective, a simple test is to inquire whether (4.12) can give a significant coherent shift to the external Hamiltonian. To estimate the size of such a shift, consider the expectation value in a generic BH state,

$$\Delta H_I^{\text{eff}} = \langle \psi_{bh}|\hat{\psi}_0^\dagger|\psi_{bh}\rangle \gamma \mathcal{O} + \text{H.c.} \qquad (4.15)$$

The size can be estimated by expanding

$$|\psi_{bh}\rangle = |\hat{0}\rangle_0 |\phi_1\rangle + |\hat{1}\rangle_0 |\phi_2\rangle, \qquad (4.16)$$

where for a generic state $|\psi_{bh}\rangle$, the states $|\phi_1\rangle$, $|\phi_2\rangle$ are generic states of the qubits excluding $\hat{\mathcal{Q}}_0$. Then the expectation value in (4.15) is $\langle \phi_2|\phi_1\rangle$, which is also of order $2^{-N/2} = e^{-S_{bh}/2}$, and so such a coherent effect is exponentially small, as also found in [28].

Another question is that of the effect on internal observers. Note that the Hamiltonian (4.12), or its generalization to the case where ΔH_{bh} also mixes the infalling modes in $\tilde{\mathcal{H}}$, has *no* effect on internal excitations until after the latency time ΔT. This could be very long as compared to the expected infall time, $\sim R$, of an infalling observer, and as noted may only become important at much longer times where deviations from nice slice evolution become

important. Once internal excitations do become mixed with the excitations on which ΔH_I acts (in the simple model, just $\hat{\mathcal{Q}}_0$), those excitations are nontrivially acted on, corresponding to, e.g., particle creation or annihilation in the relevant modes.

Finally, we have not discussed the second term in (4.12), arising from the Hermitian conjugate. Note that if the first term can lower the energy of the BH and produce an excitation in the atmosphere, this term has the opposite effect. However, we can compare this situation to that of the cavity. There, the interaction term between internal and external excitations, arising from the presence of the small hole, has the same nature: it can either transfer particles (and entanglement) out of the cavity, or back into the cavity. The reason that transfer out dominates is due to phase space considerations: there are many more states outside the cavity than inside, and when one accounts for the full Hamiltonian, excitations will dominantly fill those external states after sufficiently long time. We expect that a more complete dynamical model should exhibit such behavior.

V. COMPARING MODELS, AND TOWARDS A MORE COMPLETE DESCRIPTION

We conclude with further discussion of the comparison of the two types of models, and of possible features of their generalizations to a more complete evolutionary law, modifying the LQFT/QGR evolution described in Sec. II.

In the qubit models, the nonisometric evolution of [12] has been seen to combine unitary evolution of the internal state of the BH with a fundamentally new kind of evolution, which involves projections correlating the quanta of the would-be Hawking state with other internal excitations. This evolution accomplishes transfer of entanglement from the BH internal state, but thus in a way that is different from the quantum dynamics of other known fundamental systems with subsystems, like the cavity example described in the Introduction. These operations are neither local nor unitary, when acting on the degrees of freedom of the familiar LQFT description, and in particular nonlocally transfer entanglement to degrees of freedom near the horizon. Reference [12] suggests that the evolution be reinterpreted as unitary evolution in a holographically dual fundamental evolution.

In particular, the projection eliminates certain states of the outgoing internal excitations, which, if excited, then have "nowhere to go" in the evolution, and this also prevents buildup of type I entanglement due to the Hawking evolution. In the simplest models, only the Hawking evolution produces such excitations. However, in a more realistic model interactions can also populate these modes, raising the question of how to describe a more complete consistent evolution incorporating the nonisometric projections. A very simple example [27] is decay of an infalling particle, producing decay products in outgoing internal modes. Other questions regarding interactions are explored in [49].

---

[11]For *some* further discussion, see [54].

The models also eliminate excitations from the internal matter, through the projections; for example, after the projection (4.8), the two qubits $\tilde{\mathcal{Q}}_N$ and $\tilde{\mathcal{Q}}_{N+1}$ are no longer part of the internal system, and so one likewise expects "disappearance" of excitations or particles from the internal state in a more complete model, having an effect on internal observers. The models do however offer the apparent virtue of hiding this elimination of excitations, due to the complexity of the unitary internal mixing [12].

If the Hilbert space structure of quantum gravity is such that the BH and environment are distinct subsystems, and evolution both describes ultimate BH disappearance and is instead unitary, one requires [13] interactions to transfer the information to the external state. Scenarios for this have been investigated, with the goal of describing interactions with "minimal" effect on infalling or other near-BH observers, in the approach of "nonviolent unitarization" [46,50–52] [28,53]; such interactions have a different effect on the respective subsystems than the nonisometric models. Models for nonviolent unitarization that go beyond the qubit models surmount some important challenges, although there are also remaining questions regarding a more complete dynamics.

As with the nonisometric models, one does pay the price of apparent nonlocality with respect to the semiclassical background BH geometry. However, the interactions of nonviolent unitarization, while they can give order one corrections to amplitudes, only do so for situations with soft momentum and energy transfer, for example with characteristic wavenumber or energy $1/R$. This has nearly negligible effect on small observers falling into a BH; put differently, the departures from the equivalence principle are likewise extremely modest, and vanish in the large-R limit.

Such interactions do not have the same kind of puzzling effects on the internal outgoing excitations as the non-isometric evolution. However, through the combination of modified internal evolution due to ΔH_{bh} and the inter-actions ΔH_I, this evolution is expected to have effects on both ingoing and outgoing inside modes that depart from evolution described by LQFT, and which, e.g., lead to disappearance of excitations. Depending on the details of the model, this may only affect inside excitations after long times. One motivation for this is if the departure from the nice-slice Hamiltonian and evolution of LQFT is a small effect, and so only builds up after long-time evolution, where the nice-slice gauge for early excitations becomes increasingly extreme. This evolution moreover could share the feature of hiding the effects of information transfer [12] on inside excitations if the internal evolution via ΔH_{bh} is sufficiently complex.

One important question in the models we have described is whether the entanglement transfer from the BH is complete by the time that the BH disappears; otherwise one has a potentially problematic remnant. The simple evolution described above, with combined scrambling of excitations and transfer via interactions, is analogous to the cavity model described in the Introduction. However, as a BH nears the end of its evolution, its rate of energy loss accelerates, which is different from the slowdown of information transfer expected in such a cavity model. One expects that a more complete Hamiltonian would describe an effect analogous to squeezing of the cavity to accelerate information transfer. One effect that could at least partly account for this is an increase in the internal energy levels with decreasing BH size. Further aspects of a more complete description of evolution are to be explored in future work.

[1] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz, Time symmetry in the quantum process of measurement, Phys. Rev. **134**, B1410 (1964).

[2] R. B. Griffiths, Consistent histories and the interpretation of quantum mechanics, J. Stat. Phys. **36**, 219 (1984).

[3] M. Gell-Mann and J. B. Hartle, Time symmetry and asymmetry in quantum mechanics and quantum cosmology, in *Proceedings of the 4th International Conference on Ion Sources* (1991); arXiv:gr-qc/9304023.

[4] S. W. Hawking, Breakdown of predictability in gravitational collapse, Phys. Rev. D **14**, 2460 (1976).

[5] T. Banks, L. Susskind, and M. E. Peskin, Difficulties for the evolution of pure states into mixed states, Nucl. Phys. **B244**, 125 (1984).

[6] R. D. Carlitz and R. S. Willey, The lifetime of a black hole, Phys. Rev. D **36**, 2336 (1987).

[7] J. Preskill, Do black holes destroy information?, in *Proceedings of the International Symposium on Black Holes, Membranes, Wormholes and Superstrings* (1992); arXiv:hep-th/9209058.

[8] S. B. Giddings, Why aren't black holes infinitely produced?, Phys. Rev. D **51**, 6860 (1995); arXiv:hep-th/9412159.

[9] L. Susskind, Trouble for remnants, arXiv:hep-th/9501106.

[10] S. W. Hawking, Particle creation by black holes, Commun. Math. Phys. **43**, 199 (1975); **46**, 206(E) (1976).

[11] G. T. Horowitz and J. M. Maldacena, The black hole final state, J. High Energy Phys. 02 (2004) 008; arXiv:hep-th/0310281.

[12] C. Akers, N. Engelhardt, D. Harlow, G. Penington, and S. Vardhan, The black hole interior from non-isometric codes and complexity, arXiv:2207.06536.

[13] S. B. Giddings, A 'black hole theorem,' and its implications, Classical Quantum Gravity **40**, 085002 (2023); arXiv:2110.10690.

[14] W. Donnelly and S. B. Giddings, Gravitational splitting at first order: Quantum information localization in gravity, Phys. Rev. D **98**, 086006 (2018); arXiv:1805.11095.

[15] S. B. Giddings, Gravitational dressing, soft charges, and perturbative gravitational splitting, Phys. Rev. D **100**, 126001 (2019); arXiv:1903.06160.

[16] S. B. Giddings, On the questions of asymptotic recoverability of information and subsystems in quantum gravity, J. High Energy Phys. 08 (2022) 227; arXiv:2112.03207.

[17] C. Chowdhury, V. Godet, O. Papadoulaki, and S. Raju, Holography from the Wheeler-DeWitt equation, J. High Energy Phys. 03 (2022) 019; arXiv:2107.14802.

[18] S. D. Mathur, The fuzzball proposal for black holes: An elementary review, Fortschr. Phys. **53**, 793 (2005); arXiv:hep-th/0502050.

[19] D. N. Page, Average entropy of a subsystem, Phys. Rev. Lett. **71**, 1291 (1993); arXiv:gr-qc/9305007.

[20] D. N. Page, Information in black hole radiation, Phys. Rev. Lett. **71**, 3743 (1993); arXiv:hep-th/9306083.

[21] S. D. Mathur, The information paradox: A pedagogical introduction, Classical Quantum Gravity **26**, 224001 (2009); arXiv:0909.1038.

[22] S. B. Giddings and J. Perkins, Quantum evolution of the Hawking state for black holes, Phys. Rev. D **106**, 065011 (2022); arXiv:2204.13126.

[23] S. B. Giddings and J. Perkins, Perturbative quantum evolution of the gravitational state and dressing in general backgrounds, arXiv:2209.06836.

[24] C. G. Callan, Jr., S. B. Giddings, J. A. Harvey, and A. Strominger, Evanescent black holes, Phys. Rev. D **45**, R1005 (1992); arXiv:hep-th/9111056.

[25] S. B. Giddings, Quantum mechanics of black holes, in *ICTP Summer School in High-energy Physics and Cosmology* (1994); arXiv:hep-th/9412138.

[26] S. B. Giddings, Schrödinger evolution of the Hawking state, Phys. Rev. D **102**, 125022 (2020); arXiv:2006.10834.

[27] S. B. Giddings, Schrödinger evolution of two-dimensional black holes, J. High Energy Phys. 12 (2021) 025; arXiv:2108.07824.

[28] S. B. Giddings, Nonviolent unitarization: Basic postulates to soft quantum structure of black holes, J. High Energy Phys. 12 (2017) 047; arXiv:1701.08765.

[29] R. L. Arnowitt, S. Deser, and C. W. Misner, The dynamics of general relativity, Gen. Relativ. Gravit. **40**, 1997 (2008); arXiv:gr-qc/0405109.

[30] C. G. Torre and M. Varadarajan, Functional evolution of free quantum fields, Classical Quantum Gravity **16**, 2651 (1999); arXiv:hep-th/9811222.

[31] J. Cortez, G. A. Mena Marugan, J. Olmedo, and J. M. Velhinho, A uniqueness criterion for the Fock quantization of scalar fields with time dependent mass, Classical Quantum Gravity **28**, 172001 (2011); arXiv:1106.5000.

[32] J. Cortez, L. Fonseca, D. Martín-de Blas, and G. A. Mena Marugán, Uniqueness of the Fock quantization of scalar fields under mode preserving canonical transformations varying in time, Phys. Rev. D **87**, 044013 (2013); arXiv:1212.3947.

[33] I. Agullo and A. Ashtekar, Unitarity and ultraviolet regularity in cosmology, Phys. Rev. D **91**, 124010 (2015); arXiv:1503.03407.

[34] A. Much and R. Oeckl, Self-adjointness in Klein-Gordon theory on globally hyperbolic spacetimes, Math. Phys. Anal. Geom. **24**, 5 (2021); arXiv:1804.07782.

[35] B. S. Kay and R. M. Wald, Theorems on the uniqueness and thermal properties of stationary, nonsingular, quasifree states on space-times with a bifurcate killing horizon, Phys. Rep. **207**, 49 (1991).

[36] S. B. Giddings and W. M. Nelson, Quantum emission from two-dimensional black holes, Phys. Rev. D **46**, 2486 (1992); arXiv:hep-th/9204072.

[37] S. B. Giddings, Hilbert space structure in quantum gravity: An algebraic perspective, J. High Energy Phys. 12 (2015) 099; arXiv:1503.08207.

[38] W. Donnelly and S. B. Giddings, Diffeomorphism-invariant observables and their nonlocal algebra, Phys. Rev. D **93**, 024030 (2016); arXiv:1507.07921; **94**, 029903(E) (2016).

[39] W. Donnelly and S. B. Giddings, Observables, gravitational dressing, and obstructions to locality and subsystems, Phys. Rev. D **94**, 104038 (2016); arXiv:1607.01025.

[40] W. Donnelly and S. B. Giddings, How is quantum information localized in gravity?, Phys. Rev. D **96**, 086013 (2017); arXiv:1706.03104.

[41] S. B. Giddings, Quantum gravity: A quantum-first approach, Lett. High Energy Phys. **3**, 1 (2018).

[42] I. Heemskerk, Construction of bulk fields with gauge redundancy, J. High Energy Phys. 09 (2012) 106; arXiv:1201.3666.

[43] D. Kabat and G. Lifschytz, Decoding the hologram: Scalar fields interacting with gravity, Phys. Rev. D **89**, 066010 (2014); arXiv:1311.3020.

[44] A. Laddha, S. G. Prabhu, S. Raju, and P. Shrivastava, The holographic nature of null infinity, SciPost Phys. **10**, 041 (2021); arXiv:2002.02448.

[45] S. B. Giddings, Quantum-first gravity, Found. Phys. **49**, 177 (2019); arXiv:1803.04973.

[46] S. B. Giddings, Models for unitary black hole disintegration, Phys. Rev. D **85**, 044038 (2012); arXiv:1108.2015.

[47] S. B. Giddings, Black holes, quantum information, and unitary evolution, Phys. Rev. D **85**, 124063 (2012); arXiv:1201.1037.

[48] S. B. Giddings and Y. Shi, Quantum information transfer and models for black hole mechanics, Phys. Rev. D **87**, 064031 (2013); arXiv:1205.4732.

[49] I. H. Kim and J. Preskill, Complementarity and the unitarity of the black hole *S*-matrix, J. High Energy Phys. 02 (2023) 233; arXiv:2212.00194.

[50] S. B. Giddings, Nonviolent nonlocality, Phys. Rev. D **88**, 064023 (2013); arXiv:1211.7070.

[51] S. B. Giddings, Nonviolent information transfer from black holes: A field theory parametrization, Phys. Rev. D **88,** 024018 (2013); arXiv:1302.2613.

[52] S. B. Giddings, Modulated Hawking radiation and a non-violent channel for information release, Phys. Lett. B **738,** 92 (2014); arXiv:1401.5804.

[53] S. B. Giddings, Black holes in the quantum universe, Phil. Trans. R. Soc. A **377,** 20190029 (2019); arXiv:1905 .08807.

[54] S. B. Giddings and M. Rota, Quantum information or entanglement transfer between subsystems, Phys. Rev. A **98,** 062329 (2018); arXiv:1710.00005.