PHYSICAL REVIEW D **79,** 044010 (2009)

Unitarity and holography in gravitational physics

Donald Marolf*

Physics Department, UCSB, Santa Barbara, California 93106, USA

(Received 17 September 2008; published 9 February 2009)

Because the gravitational Hamiltonian is a pure boundary term on shell, asymptotic gravitational fields store information in a manner not possible in local field theories. This fact has consequences for both perturbative and nonperturbative quantum gravity. In perturbation theory about an asymptotically flat collapsing black hole, the algebra generated by asymptotic fields on future null infinity within any neighborhood of spacelike infinity contains a complete set of observables. Assuming that the same algebra remains complete at the nonperturbative quantum level, we argue that either (1) the S matrix is unitary or (2) the dynamics in the region near timelike, null, and spacelike infinity is not described by perturbative quantum gravity about flat space. We also consider perturbation theory about a collapsing asymptotically anti-de Sitter (AdS) black hole, where we show that the algebra of boundary observables within any neighborhood of any boundary Cauchy surface is similarly complete. Whether or not this algebra continues to be complete nonperturbatively, the assumption that the Hamiltonian remains a boundary term implies that information available at the AdS boundary at any one time t_1 remains present at this boundary at any other time t_2.

DOI: 10.1103/PhysRevD.79.044010 PACS numbers: 04.70.Dy

I. INTRODUCTION

Arguments for information loss in black hole evaporation are typically based on locality and causality in quantum field theory on a fixed background (see e.g. [1,2]). In perturbative quantum gravity these properties also hold at zeroth order in the Planck length ℓ_p, where backreaction is ignored. However, strict locality explicitly fails at the first interacting order. A clean signal of this failure is the fact that a form of time evolution is generated by a boundary term at spacelike infinity (e.g., the Arnowitt-Deser-Misner (ADM) energy in asymptotically flat space [3]). This feature is closely related to the lack of local observables in diffeomorphism-invariant theories.

We show below that this simple observation leads to interesting results. For example, consider the context of perturbation theory about asymptotically flat collapsing black hole backgrounds. There we will show that, at first interacting order and beyond, a complete set of observables is contained in the algebra generated by fields on future null infinity (I^+) within any neighborhood of spacelike infinity (i^0). In the asymptotically anti-de Sitter (AdS) context, the algebra of boundary observables defined by any neighborhood of a boundary Cauchy surface is similarly complete in perturbation theory. As a result, in both cases full information about the quantum state is contained in the asymptotic fields.

We refer to the above completeness results as ''perturbative holography.'' However, we caution the reader that, in contrast to [4], our use of this term does not directly imply any particular limit on the number of degrees of freedom. The centrality of energy conservation to any discussion of unitarity was previously emphasized in [5], while the representation of gravitational energy as a boundary term and the associated ability of the long-range gravitational fields to store information was emphasized in [6]. The arguments below stem from a fusion of these ideas. Other works connecting energy conservation to black hole unitarity include [7].

Before beginning the main arguments, it is appropriate to briefly address three common objections that the reader may already hold:

Objection #1, Locality via gauge fixing.—The reader may object that perturbative quantum gravity appears both local and causal in, say, de Donder gauge. However, it is important to recall that such gauges contain propagating longitudinal gravitons associated with residual gauge symmetries. As is familiar from the Coulomb gauge in Maxwell theory, gauge fixing all residual symmetries removes the apparently manifest locality so that no immediate conclusions can be drawn regarding the nature of observables. In Yang-Mills theory, one can avoid these issues by constructing Wilson loops which provide a complete set of compactly supported observables. In contrast, no such compactly supported observables are available in diffeomorphism-invariant theories of gravity.

Objection #2, The characteristic initial value problem.—Section II A considers perturbations about an asymptotically flat collapsing black hole spacetime. At the level of rigor used below, the characteristic initial value theorem states that the radiative parts of metric perturbations on the future horizon (H^+) and future null infinity (I^+) form a complete set of independent operators. As a result, the radiative parts of metric perturbations on I^+ cannot, by themselves, define a complete set of observables in this context. It is important to note that this statement

*marolf@physics.ucsb.edu

does not contradict our claims. Indeed, our argument below makes explicit use of *both* the radiative and the nonradiative (Coulomb) parts of the metric perturbations on I^+. These Coulomb parts are *not* independent of the radiative parts of metric perturbations on H^+, but are instead related to the full set of radiative perturbations by the gravitational constraints.

Objection #3, Comparison with classical physics.— While we use a quantum-mechanical language below, replacing certain commutators by Poisson brackets suffices to recast our perturbative arguments in the language of classical gravitational physics. As a result, our arguments imply that in *classical* perturbative gravity the Poisson algebra[1] generated by fields on future null infinity (I^+) in any neighborhood of spacelike infinity (i^0) contains a complete set of observables, and that a similar result holds in the anti-de Sitter context.

The reader may feel that this statement should contradict the fact the black holes lose information in classical gravity. That no such contradiction arises can be illustrated using the SO(3) angular momentum generators J_x, J_y, J_z. Of course, J_z lies in the Poisson algebra generated by J_x, J_y. Nevertheless, at the classical level, the ability to measure J_x and J_y imparts no knowledge of J_z. Full information is obtained only about *algebraic* functions $f(J_x, J_y)$. It is only at the quantum level that the situation changes, and that alternating measurements of J_x and J_y can indeed provide information about J_z. This last point will be emphasized in a companion paper [8], which also resolves a number of possible paradoxes that the reader may fear might be associated with such measurements.

Having dispensed with the above objections, we may now turn to the main arguments. At the quantum level our discussion is somewhat formal. However, at least in the perturbative context, mathematically rigorous results can be obtained by reinterpreting the arguments below in terms of classical gravity, replacing commutators with Poisson brackets and (where appropriate) with finite flows along Hamiltonian vector fields. As briefly discussed under objection #3 above, while such results have minimal implications for classical physics, it is clear that they set the stage for more interesting effects at the quantum level.

We begin with the asymptotically flat context in Sec. II. After deriving perturbative completeness of the algebra near i^0, we consider implications for the nonperturbative theory. Assuming that the same algebra remains complete in the nonperturbative quantum theory, we show that either (1) the S matrix is unitary or (2) the dynamics in the region near timelike, null, and spacelike infinity is not described by perturbative quantum gravity about flat space.

We then derive perturbative holography for asymptotically anti-de Sitter (AdS) quantum gravity in Sec. III. We also note that, whether or not the stated algebra continues to be complete nonperturbatively, the assumption that the Hamiltonian remains a boundary term implies a form of boundary unitarity. In particular, information available at the AdS boundary at any one time t_1 remains present at this boundary at any other time t_2. We close with some final discussion in Sec. IV.

II. QUANTUM GRAVITY IN ASYMPTOTICALLY FLAT SPACE

To avoid making detailed assumptions about the quantum nature of gravity, it is natural to proceed using either semiclassical methods or perturbation theory. We choose the latter here, where we have in mind treating perturbative gravity as an effective field theory (in which appropriate new parameters may need to be added at each order). This is the setting for Sec. II A. Section II B then studies the implications for the nonperturbative theory and discusses the unitarity of the S matrix.

A. The holographic nature of perturbative gravity

It is useful to begin with a brief summary of the argument: We consider perturbation theory around an asymptotically flat classical solution which is flat in the distant past but contains a black hole in the distant future. The argument below simply uses the Hamiltonian (an operator at i^0) to translate any operator on past null infinity (I^-) into the distant past, deep into the flat region before the black hole forms. The perturbative equations of motion then express any such operator in terms of operators on I^+. That is, since the black hole does not form until much later, very little of the operator falls into the black hole. Furthermore, since we translated the operator on I^- into the distant past, the support on I^+ is concentrated near i^0. Taking a limit yields the desired result.

It is convenient to perturb about a background solution which is exactly flat space before some advanced time v_0 (see Fig. 1). For familiarity and concreteness, we consider pure Einstein-Hilbert gravity in 3 + 1 dimensions so that the black hole forms from gravitational waves arriving from past null infinity (I^-). Adding matter fields or changing the number of dimensions would not significantly change the analysis.[2] The essential inputs are only

[1] In fact, one requires a certain closure of the usual Poisson algebra which allows one to flow any element A of the algebra by a finite amount along the Hamiltonian vector field defined by any other element B.

[2] The sole exception is that the infrared behavior improves in higher dimensions. In 3 + 1 dimensions, our argument is rather formal in that it ignores infrared divergences associated with soft gravitons. While it may be interesting to examine the detailed effect of IR divergences on the argument below, here we simply assume that the usual techniques [9] allow us to use gravitational perturbation theory and to speak of an S matrix. In higher dimensions, no such divergences arise.

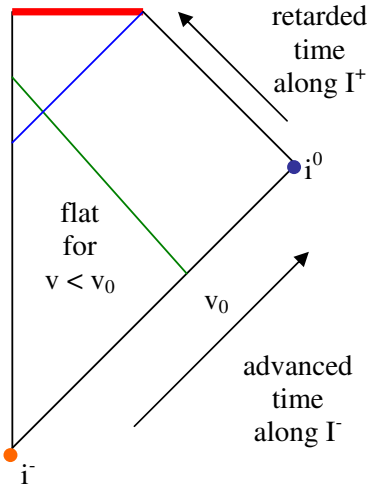

FIG. 1 (color online). The spacetime is flat before advanced time v_0, but the formation of a black hole prohibits a regular i^+.

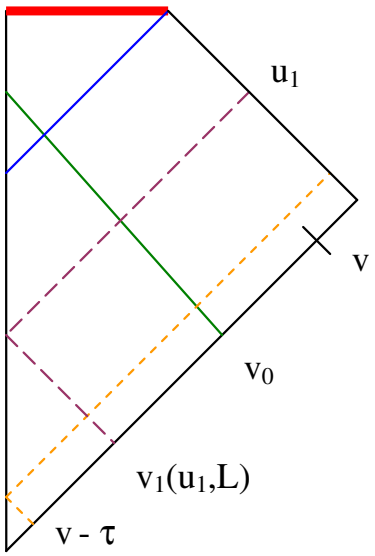

FIG. 2 (color online). As $\tau \to +\infty$, operators at $v - \tau$ on I^- can be written in terms of operators on I^+ before retarded time u_1.

diffeomorphism-invariance (so that the Hamiltonian is indeed a boundary term) and our choice of boundary conditions.

To begin the main argument, let $\tilde{g}_{ab}$ denote the metric of the background spacetime and write the dynamical metric as $g_{ab} = \tilde{g}_{ab} + \kappa h_{ab}$ where $\kappa^2 = 8\pi G$ so that the action for h_{ab} has canonical kinetic term. As usual, we work to some finite order in κ and discard terms of higher order. We will not need to be explicit about the details below; all that is important is that we work to some order in which interactions are relevant so that the gravitational version of Gauss' law leads to a nontrivial gravitational flux [see (2.1) below] at spacelike infinity (i^0). For later use it will also be convenient to expand the background about flat space by writing $\tilde{g}_{ab} = \eta_{ab} + \kappa \tilde{h}_{ab}$. The latter expansion is useful near infinity where $\tilde{h}_{ab}$ is small.

The perturbations h_{ab} may be quantized in any gauge for which all propagating modes are physical; e.g. a Coulomb-like gauge. The Hamiltonian in such gauges is necessarily nonlocal, but this will not be a complication. The advantage of such gauges is that all equations of motion hold at the level of the Heisenberg operators. For example, the gravitational equivalent of Gauss' law holds as an operator identity and need not be imposed as a constraint on physical states.

We now remind the reader of several facts from classical general relativity. First, recall that the total energy of the full metric g_{ab} is given by the ADM boundary term at spatial infinity (i^0). We denote this boundary term Φ as it will be convenient to think of this term as a gravitational flux. We have

$$\Phi = \frac{1}{2\kappa} \int_C dA (r^a P^{bc} D_b - r^b P^{ac} D_b)(\tilde{h}_{ac} + h_{ac}), \quad (2.1)$$

where r^a is a radial unit normal, C is a cut of i^0 as defined e.g. in [10], dA is the area element on C, and D_a is the covariant derivative defined by the fixed flat metric η^{ab} which also defines the spatial projection P_{ab} orthogonal to the chosen time direction.

Second, if past timelike infinity (i^-) is regular, then the ADM energy can also be expressed as the integral over past null infinity (I^-) of the flux of stress-energy through I^- due to gravitational radiation (see e.g. [11]). This flux is given by the news tensor, but may be equally well thought of as the integral of the appropriate component of the stress tensor of linearized gravity integrated along I^- (see e.g. [12]). Either expression is purely quadratic in $g_{ab} - \eta_{ab}$, where η_{ab} is a fixed flat metric at infinity. This calculation shows explicitly that Φ agrees near I^- with the Hamiltonian of linearized gravity about flat space, where the linearized field is $\tilde{h}_{ab} + h_{ab}$. We denote this Hamiltonian $H^{\rm lin}_{\tilde{h}+h}$. Note that since the perturbations $\tilde{h}_{ab}$, h_{ab} fall off at I^-, this linearized Hamiltonian also generates translations along I^- in the full theory (and, in particular, at any order in perturbation theory).

Since $H^{\rm lin}_{\tilde{h}+h}$ is quadratic, it is straightforward to expand in powers of h_{ab}:

$$H^{\rm lin}_{\tilde{h}+h} = \tilde{E} + S + H^{\rm lin}_h. \quad (2.2)$$

Here $\tilde{E}$ is the v-dependent energy of the background metric $\tilde{g}_{ab}$, S denotes a set of "source terms" linear in both $\tilde{h}_{ab}$ and h_{ab}, and $H^{\rm lin}_h$ is just the integral of the (quadratic) stress tensor for perturbations h_{ab} propagating on the flat metric η_{ab}.

Most importantly for our purposes, the above results can be derived using the equations of motion near $I^\pm$ expanded only to second order in h_{ab}. As a result, they hold in perturbative classical gravity at any order beyond the free

linear theory; i.e., at any order where the gravitational Gauss' law makes Φ nontrivial. Furthermore, the results also hold in perturbative quantum gravity as the only operator that requires regularization is the (quadratic) stress tensor for gravitons propagating in flat space.

Below, it will be convenient to denote operators on I^- as $h_{ab}(v)$, and to speak as if they are well-defined operators. In doing so we choose a notation which suppresses several details. First, some rescaling with r is required to define finite objects on I^-. Second, we implicitly assume that the operators have been smeared with appropriate test functions. Third, at certain points below it will be convenient to assume that an expansion in spherical harmonics has been performed and that each $h_{ab}(v)$ has a definite angular momentum.

Since we consider perturbations about a background $\tilde{g}_{ab}$ which is flat before the advanced time v_0, past timelike infinity is regular. As a result, the relation

$$\Phi = H^{\rm lin}_{\tilde{h}+h} \tag{2.3}$$

holds as an equality of Heisenberg-picture quantum operators. This relation is somewhat subtle, however, since Φ as defined in (2.1) is linear in $\tilde{h}_{ab} + h_{ab}$ while $H^{\rm lin}$ is quadratic. The point here is simply that $H^{\rm lin}$ is defined by linearizing about a certain background (η_{ab}). As a result, the relationship between Φ and $H^{\rm lin}$ is sensitive to this choice of background. In particular, subtracting the (v-dependent) energy $\tilde{E}(v)$ of the background metric yields $\Phi - \tilde{E}(v) = S(v) + H^{\rm lin}_h$, where $S(v)$ is an operator linear in both $\tilde{h}_{ab}(v)$ and $h_{ab}(v)$ as in (2.2). Thus, $S(v)$ has an explicit v dependence though the background $\tilde{h}_{ab}(v)$. In contrast, the operator $H^{\rm lin}_h$ is just what would appear in linearized gravity about flat space; $H^{\rm lin}_h$ has no explicit v dependence.

From (2.3) we see that Φ generates v translations of $\tilde{h}_{ab} + h_{ab}$ in the sense that

$$(\tilde{h}_{ab} + h_{ab})(v) = e^{-i\tau\Phi}(\tilde{h}_{ab} + h_{ab})(v-\tau)e^{i\tau\Phi},$$

or

$$h_{ab}(v) = e^{-i\tau\Phi}h_{ab}(v-\tau)e^{i\tau\Phi} + \tilde{h}_{ab}(v-\tau) - \tilde{h}_{ab}(v). \tag{2.4}$$

The terms involving $\tilde{h}_{ab}$ on the final right-hand side are associated with the source terms $S(v)$ in (2.2), or equivalently with the difference between Φ and $H^{\rm lin}_h$. The role of these c-number terms is to compensate for the fact that Φ effectively translates both the perturbation and the background. Equation (2.4) is a key result which we will use liberally. Note that while $\tilde{h}_{ab}(v)$ is formally of order $1/\kappa$, its effects become arbitrarily small at sufficiently large r; i.e., near infinity terms involving $\tilde{h}_{ab}(v)$ need not interfere with our perturbative treatment.

We now proceed to our main argument. Choose any retarded time u_1 along I^+ and any operator $h_{ab}(v)$ at any advanced time v on I^-. We wish to show that, in any state, the operator $h_{ab}(v)$ can be arbitrarily well approximated by elements of the algebra $\mathcal{A}^+_{u_1}$ generated by operators at I^+ supported at retarded times $u < u_1$. By convention,[3] we consider i^0 to be a point on I^+ with $u = -\infty$ so that $\mathcal{A}^+_{u_1}$ contains Φ. Since we may use (2.4), it remains only to approximate $h_{ab}(v-\tau)$ by operators in $\mathcal{A}^+_{u_1}$.

To do so, note that since $\tilde{g}_{ab}$ is flat for $v < v_0$, there is some advanced time $v_1(u_1, L)$ such that all null geodesics (with angular momentum L) launched from I^- before $v_1(u_1, L)$ arrive at I^+ before retarded time u_1. As a result, in the geometric optics approximation to the linearized theory, the equations of motion relate operators $h_{ab}(v-\tau)$ with angular momentum L and $v - \tau < v_1(u_1, L)$ to an operator in $\mathcal{A}^+_{u_1}$. This situation is summarized in Fig. 2. Beyond the geometric optics approximation, and taking into account nonlinear corrections at some fixed order of perturbation theory, we may use the equations of motion to write

$$h_{ab}(v-\tau) = \mathcal{O}_{ab}(v-\tau, u_1) + \Delta_{ab}(v-\tau, u_1), \tag{2.5}$$

where $\mathcal{O}_{ab}(v-\tau, u_1) \in \mathcal{A}^+_{u_1}$ and $\Delta_{ab}(v-\tau, u_1)$ is an error term. Because all corrections are determined by Green's functions peaked on the light cone, in any fixed state (having a finite number of particles on I^-) the error $\Delta_{ab}(v-\tau, u_1)$ will vanish as some power law in the limit $v_1 - (v-\tau) \to \infty$.

This is nearly the desired result. For the final step of the argument, it is useful to express $\Delta_{ab}(v-\tau, u_1)$ in terms of the operators $h_{ab}(v)$ on I^- using the same perturbative equations of motion. The largest contributions will come from the region near $v_1(u_1, L)$, but there will be power-law suppressed contributions from other regions as well. Now, since we observed above that matrix elements of $\Delta_{ab}(v-\tau, u_1)$ must vanish as a power law in all states having a finite number of particles on I^- in the limit $(v-\tau) \to \infty$, we may expand $\Delta_{ab}(v-\tau, u_1)$ in powers of $(v-\tau)^{-1}$; i.e., we write

$$\Delta_{ab}(v-\tau, u_1) \approx \sum_{n>0} (v-\tau, u_1)^{-n} \Delta^{(n)}_{ab}(u_1), \tag{2.6}$$

where the operators $\Delta^{(n)}_{ab}(u_1)$ are independent of v, τ. We will use (2.6) only as an asymptotic series and do not require convergence. At any fixed order in perturbation theory, the operators $\Delta^{(n)}_{ab}(u_1)$ are simply integrals over products of operators h_{ab} on I^- with a weighting function determined by u_1. Using (2.6), consider now the contribution

$$e^{-i\tau\Phi}\Delta_{ab}(v-\tau, u_1)e^{i\tau\Phi} = \sum_{n>0} (v-\tau, u_1)^{-n} \times e^{-i\tau\Phi}\Delta^{(n)}_{ab}(u_1)e^{i\tau\Phi} \tag{2.7}$$

[3] We could also have used the Bondi energy associated with a cut of I^+ at retarded time u to approximate Φ as $u \to -\infty$, but our argument loses nothing by making the above simplifying convention.

of $\Delta_{ab}(v-\tau, u_1)$ to (2.4). We wish to take the limit $\tau \to \infty$. This has two effects. First, the factors of $e^{\pm i\tau\Phi}$ translate each $\Delta^{(n)}_{ab}(u_1)$ toward i^0. Since correlation functions in any Fock space state approach those of the vacuum at large times, the large τ limit of each $e^{-i\tau\Phi}\Delta^{(n)}_{ab}(u_1)e^{i\tau\Phi}$ is a (finite) c number determined by the background metric $\tilde{g}_{ab}$. It follows that the large τ limit of (2.7) must vanish due to the factors of $(v-\tau, u_1)^{-n}$.

Combining the above results we have

$$h_{ab}(v) = \lim_{\tau\to\infty}[e^{-i\tau\Phi}\mathcal{O}_{ab}(v-\tau, u_1)e^{i\tau\Phi} + \tilde{h}_{ab}(v-\tau) - \tilde{h}_{ab}(v)]. \tag{2.8}$$

Since any c number [e.g., $\tilde{h}_{ab}(v)$ or $\tilde{h}_{ab}(v)$] lies in $\mathcal{A}^+_{u_1}$, the right-hand side contains only elements of $\mathcal{A}^+_{u_1}$ as desired. Thus we have shown that any fundamental field on I^- can be expressed with arbitrary accuracy as an element of $\mathcal{A}^+_{u_1}$. Similarly, any product of such fields can be expressed (with arbitrary accuracy) by taking the above limit separately for each operator in the product.

We conclude that a complete set of operators on I^- is contained in the weak closure of $\mathcal{A}^+_{u_1}$. For convenience, we used a Coulomb-like gauge, but the corresponding result for gauge-invariant observables follows immediately in any gauge.

B. Nonperturbative gravity and unitarity of the *S* matrix

We saw above that perturbative gravity about an asymptotically flat spacetime is holographic in the sense that the algebra of observables generated by the ADM Hamiltonian Φ and the usual asymptotic fields within any neighborhood of i^0 in I^+ contains a complete set of observables. Thus, all of the information present at I^- is encoded in observables in the stated region of I^+. However, discussions of black hole unitarity typically focus on unitarity of the S matrix. This is a somewhat different question, defined in terms of the Fock spaces at $I^\pm$. In particular, it is manifestly clear that, at a finite order in perturbation theory about a collapsing black hole, the Fock spaces at $I^\pm$ do not encode the same degrees of freedom.

From our point of view, this difference arises because there is no regular future timelike infinity in a black hole spacetime. As a result, in perturbation theory about such a background, the total gravitational flux Φ cannot be expressed solely in terms of the stress tensor at I^+, and thus cannot be expressed in terms of creation and annihilation operators at I^+. This was possible at I^- only due to the particular boundary conditions chosen at i^-.

On the other hand, one expects any black hole that forms to decay by Hawking evaporation. While this process cannot be fully described in perturbation theory, perturbative quantum gravity (say, about flat spacetime) may well be a good description of the end products resulting from the decay. In this case, i^+ is regular. Let us therefore suppose that, in any asymptotically flat state of the nonperturbative theory, perturbative quantum gravity about flat spacetime becomes an arbitrarily good approximation for field operators near past (i^- and I^-), future (i^+ and I^+), and spacelike infinity (i^0). Let us also extrapolate our perturbative result and assume that the algebra generated by Φ and asymptotic fields on I^+ in any $\mathcal{A}^+_{u_1}$ again contains a complete set of observables, at least within an appropriate superselection sector.[4] Since we have a regular i^+, the gravitational flux Φ *can* be expressed as the integral of the linearized stress tensor over I^+. It follows that any observable can indeed be expressed in terms of creation and annihilation operators on I^+. Our discussion is tailored to settings with no stable massive particles but, since we assume that physics is perturbative near i^+, allowing stable massive particles would merely require Φ to be expressed in terms of the stress tensor at both I^+ and i^+, and for the corresponding creation and annihilation operators at i^+ to be included in our discussion.

Note that the other Poincaré generators on I^- can be related to those on I^+ in precisely the same manner as was done for time translations. Thus the Poincaré-invariant vacuum on I^- also defines a Poincaré-invariant state on I^+. Since such a state is unique in perturbative quantum field theory, the Fock vacua on $I^\pm$ coincide.

The unitarity of the S matrix now follows in the usual way. N-particle states are defined by the action of local operators at $I^\pm$ on the Fock vacuum. Since local operators can be translated between I^+ and I^-, and since the vacua at $I^\pm$ coincide, these constructions merely define two bases for the same Hilbert space. The S matrix is then nothing more than the expression of the dictionary between I^- and I^+. Since the two bases define the same Hilbert space, the S matrix is unitary.

III. ASYMPTOTICALLY ADS QUANTUM GRAVITY

We saw above that there is a sense in which perturbative gravity is holographic in asymptotically flat space. As we now show, similar methods lead to an analogous result in the context of (e.g., 3 + 1) AdS asymptotics. To be specific, we require that the metric has a Fefferman-Graham expansion [14] (see also [15]) of the form

[4]Note that this is necessarily a new assumption. In particular, it does not follow from the assumption that perturbation theory is arbitrarily good near infinity. Our previous perturbative argument required us to propagate fields from I^- to I^+ through the bulk of the spacetime where nonperturbative effects can be important. The purpose of mentioning our perturbative argument here is only to render this assumption plausible by removing objections based on perturbative fields falling into semiclassical black holes. See e.g. [13] for further discussion of the idea that this assumption may hold only within an appropriate superselection sector.

$$g_{ab} = \frac{\ell^2}{r^2} dr^2 + \left(g_{(0)ij} \frac{r^2}{\ell^2} + g_{(1)ij} \frac{r}{\ell} + g_{(2)ij} + g_{(3)ij} \frac{\ell}{r} + \cdots \right) dx^i dx^j, \tag{3.1}$$

for some fixed boundary metric $g_{(0)ij}$. Here ℓ is the AdS scale, the x^i are coordinates on $S^2 \times \mathbb{R}$, and the $\cdots$ represent higher order terms in r/ℓ which may include cross terms of the form $drdx^i$. The coefficients $g_{(1)ij}$, $g_{(2)ij}$ are determined by the choice of $g_{(0)ij}$ (and any matter fields, see below) via the Einstein equations. In contrast, $g_{(3)ij}$ depends on the propagating degrees of freedom in the bulk. For convenience below, we will take one of the coordinates to be some t such that the intersection of each $t =$ constant surface with the boundary spacetime is a Cauchy surface of the boundary spacetime.

Certain simplifications arise if we couple the gravitational field to a conformally coupled scalar field ϕ, though this does not appear to be essential to the argument. In $3 + 1$ dimensions we take the scalar to have the standard asymptotic behavior (see e.g. [16])

$$\phi = \frac{\alpha}{r} + \frac{\beta}{r^2} + \cdots, \tag{3.2}$$

where α will be a fixed scalar function on the boundary. In this context, we may fix $g_{(0)ab}$ to be the metric on the Einstein static universe. We also take $\alpha = 0$ before some time t_i and again after some time t_f. In particular, we take the background metric $\tilde{g}_{ab}$ to describe empty AdS space to the past of some boundary time t_f. For $t_i < t < t_f$, the time dependence of α will be chosen to generate scalar radiation which collapses to form a black hole.[5] Note that for such boundary conditions we may define a time-dependent Hamiltonian which differs from the Hamiltonian for $\alpha = 0$ by the addition of certain source terms for the scalar field in the region $t_i < t < t_f$.

Now consider any spacelike surface Σ in the initial pure AdS region. It is clear that any field at any later time can be expressed in terms of fields on Σ. Similarly, *in the linearized approximation*, any field on Σ can be expressed in terms of the boundary fields $g_{(3)ab}$ and β at earlier times. Some explicit formulas for the scalar case[6] appear in e.g. [17], but the fact that this is possible follows immediately from the observation that any linearized solution with given $\delta g_{0(ab)}$ and α is determined by the values of $\delta g_{0(ab)}$, α, $g_{3(ab)}$, and β to the past of Σ. This in turn follows from a simple argument: Suppose that two such solutions have the same values of $\delta g_{0(ab)}$, α, $g_{3(ab)}$, and β to the past of Σ, so that their difference has $\delta g_{0(ab)} = \alpha = g_{3(ab)} = \beta = 0$. This solution also satisfies ingoing boundary conditions, and so must vanish in the distant past. In particular, the energy function defined by Dirichlet boundary conditions vanishes in the distant past when evaluated on this solution. But by construction our difference solution conserves this notion energy, so that it must vanish at all times; i.e., the solution must vanish identically. We conclude that any linearized field on Σ is determined by the boundary fields $g_{(3)ab}$ and β at earlier times. As a result, any operator in the linearized theory may be expressed in terms of the boundary operators $g_{(3)ab}$ and β.

It follows that the same result holds at each order in perturbation theory. However, we stress that since we have used $g_{(3)ab}$ and β at all times, this statement does yet not constitute "holography." Instead, it merely notes certain properties of wave equations in anti-de Sitter space.[7]

To complete the argument for our perturbative holography, simply note that the algebra of boundary operators $\mathcal{A}_{t,\Delta t}$ supported within any time Δt of any boundary time t contains the Hamiltonian. Thus we may in fact express any perturbative field on Σ as an element of $\mathcal{A}_{t,\Delta t}$ for *any* t, Δt, including those times in the distant future. For $t > t_i$, we need merely include the effects of the source terms in the time-dependent Hamiltonian. Since the coordinate t is arbitrary, it follows that the algebra generated by boundary fields within any neighborhood of any boundary Cauchy surface is similarly complete.

At least at the level of perturbation theory, we have expressed any observable in terms of the boundary fields at an arbitrary time t. In this sense, perturbative gravity in AdS may be called "holographic." However, as in the case of asymptotically flat space, this observation does not immediately allow us to express our observable as a set of standard creation and annihilation operators at the desired late time. As in flat space, it is manifestly clear that such an expression is *not* possible at any finite order in perturbation theory about a black hole background.

Let us therefore briefly consider a nonperturbative theory. In asymptotically flat space we assumed that perturba-

[5]It is straightforward to find such boundary conditions. Consider for the moment a solution to the free conformally coupled scalar wave equation on the $3 + 1$ Einstein static universe in which $\phi = 0$ in the northern hemisphere at some time t_i, but in which a large spherically symmetric pulse of short-wavelength scalar radiation crosses the equator a short time later. Now restrict this solution to the northern hemisphere and conformally map the result to a solution of the free scalar equation on AdS. The $\alpha(x)$ defined by this solution generates a large spherical pulse of scalar radiation which enters the AdS space through the boundary shortly after time t_i. For large enough amplitude, this pulse will collapse to form a black hole.

[6]The explicit formula in [17] expresses local bulk fields in terms of boundary fields in a compact region of the boundary causally disconnected from the point at which the local bulk field is defined. A small additional time translation will reexpress this result in terms of fields at earlier times.

[7]This result is similar to certain consequences of Holmgren's uniqueness theorem [18], though in our context we find global uniqueness of the solution.

tive quantum gravity was a good approximation at both early and late times in order to derive unitarity of the S matrix. We could give a similar argument in the AdS case, but it would require nonstandard boundary conditions that allow the particles to leave the original AdS space. For example, we could consider the evaporon model of [19]. However, it is perhaps more enlightening to maintain standard AdS boundary conditions and to derive a more restrictive result. To proceed, we assume only that

(i) There is a well-defined, perhaps time-dependent, family of self-adjoint operators $H(t)$.
(ii) Each $H(t)$ is a member of the corresponding algebra $\mathcal{A}_{t,\Delta t}$ of boundary observables.
(iii) This family of operators generates time evolution in the usual sense associated with time-dependent Hamiltonians; i.e., the time translation is $U(t_1, t_2) = \mathcal{P}\exp(-i\int_{t_1}^{t_2} H(t)dt)$, where $\mathcal{P}$ denotes path ordering.

From these assumptions alone we cannot conclude that $\mathcal{A}_{t,\Delta t}$ contains the full set of observables, nor can we conclude that all information is present at the boundary. However, given any observable $\mathcal{O}_{t_0} \in \mathcal{A}_{t_0,\Delta t}$, we can use (i) and (ii) to define a one-parameter family of operators $\mathcal{O}_t \in \mathcal{A}_{t,\Delta t}$ which satisfy

$$\frac{d}{dt}\mathcal{O}_t = i[H(t), \mathcal{O}_t]. \tag{3.3}$$

It then follows from (ii) that $\frac{d}{dt}\mathcal{O}_t$ *also* lies in $\mathcal{A}_{t,\Delta t}$. Since this holds for each possible $\mathcal{O}_t \in \mathcal{A}_{t,\Delta t}$, the algebra does not change with time. That is, each $\mathcal{A}_{t,\Delta t}$ contains the *same* set of observables. In this sense, any information which happens to be present at the boundary at any time t_1 remains present at any other time t_2. This result is naturally called "boundary unitarity."

We again stress that the above argument does not assume completeness of the boundary observables. In particular, assumption (i) does not specify the Hilbert space on which $H(t)$ is self-adjoint. We leave thus open the possibility of new nonperturbative bulk observables, or perhaps even of new observables corresponding to "baby universes." The role of assumption (i) is merely to ensure that the path-ordered exponential of $\int H(t)dt$ is well defined. In a corresponding argument at the classical level, all that would be required is that one be able to flow any boundary observable $\mathcal{O}$ by any finite amount of time along the (time-dependent) Hamiltonian vector field generated by $H(t)$; i.e., one simply requires time evolution to be well defined along the asymptotic boundary. Such a requirement would amount to a rather weak form of cosmic censorship.

To provide some physical interpretation of the above result, consider a hypothetical observer who lives outside the spacetime but who can interact with our spacetime through the boundary observables. If the observer has complete control over the full algebra $\mathcal{A}_{t,\Delta t}$ of boundary observables at each t, then at any time t_2 boundary unitarity will allow her to extract any information which she has encoded in the spacetime at any earlier time t_1.

Physically, the point is that particles which travel inward from the boundary at time t_1 leave an imprint on the boundary fields: the gravitational constraints precisely encode the total energy in the gravitational flux Φ at the boundary. Because energy is the generator of time translations, the boundary observer can recover the desired information at any later time through appropriate couplings to this energy. Such processes will be explored in detail in [8].

IV. DISCUSSION

We have argued that perturbative quantum gravity about a collapsing black hole background is, in a certain sense, holographic. By this we mean that, in the asymptotically flat context, the algebra generated by asymptotic fields on I^+ within any neighborhood of i^0 contains a complete set of observables. In the AdS context, the algebra of boundary observables associated with any neighborhood of any Cauchy surface of the boundary spacetime is similarly complete. The fact that the gravitational Hamiltonian is a pure boundary term played a key role, in a manner similar to that predicted in [6].

If this same algebra remains complete at the nonperturbative level, and if perturbative quantum gravity about flat space is a good approximation to some asymptotically flat nonperturbative quantum gravity theory near past infinity (i^- and I^-), future infinity (i^+ and I^+), and spacelike infinity (i^0), it follows that the S matrix is unitary. This is again true if the completeness holds only in some appropriate superselection sector, as it would in an asymptotically flat analogue of the scenario outlined in [13].

It is interesting to classify possible failures of the assumption that perturbative gravity describes physics near $I^\pm$, $i^\pm$, and i^0 into two types. First, the physics might be described by perturbative quantum gravity about some different background. This might occur if the original boundary conditions are somehow unstable and if additional boundaries arise dynamically. The other sort of failure would preserve the boundary conditions but not allow a good approximation by perturbative quantum gravity. This might occur if, for example, strongly coupled regions continue to interact with perturbative fields at all times. This could be the case in so-called third-quantized theories [20], in which a given universe continually interacts with a bath of baby universes. However, in such cases a form of unitarity may nevertheless hold due to the superselection effects discussed in [21].

In the AdS context, much weaker assumptions imply that similar superselection effects *must* occur. Specifically, whether or not the set of boundary observables is complete, boundary unitarity follows directly from the assumption that, in the nonperturbative theory, the algebra of boundary observables again contains a self-adjoint Hamiltonian.

While complete information may never be present at the boundary, any information present there at one time t_1 is also contained in boundary observables at any other time t_2. Any independent observables that may exist do not affect the evolution of boundary observables, though a given quantum state might contain interesting correlations. We note briefly that this fits well with the picture of certain extensions of AdS/CFT discussed e.g. in [22,23] and with the general picture of AdS/CFT described in [13].

A number of possible objections were already addressed in the introduction. Nonethless, the reader may have certain further concerns. For example, one may worry that the presence of so much information near infinity might violate the ''no quantum Xerox theorem'' [24]. However, the original quantum state has in no way been copied to new degrees of freedom. Instead, the equations of motion imply operator identities which require two *a priori* different operators to be sensitive to the same qubit of quantum information.

One might also worry that our scenario may lead to paradoxes associated with noncommuting measurements of some qubit being performed by spacelike-separated observers: one in the interior of the spacetime who measures local degrees of freedom, and one at the boundary who makes use of the holographic encoding in the algebra of boundary observables. However, in a context where the boundary observables are complete, the boundary observer has access to *all* degrees of freedom, including the measuring devices of the local observer. As a result, no paradoxes can arise. Any measurement made by a local observer can always be undone by the boundary observer, though it would of course be interesting to understand the details.

An interesting, if perhaps somewhat artificial, context where the usual algebra of boundary observables is *not* complete can be constructed by adding a second boundary to the spacetime. We may then place one observer outside each boundary, so that there is no danger of the local observer's devices being holographically encoded at the other boundary. Since the interesting case arises when the two boundaries are in causal contact, we take this new boundary to be at finite distance (i.e., it is not an asymptotic boundary).

In the asymptotically flat version, this finite boundary may prohibit i^+ from being regular and may also interfere with the scattering of wave packets at early times. As a result, we cannot conclude that complete information is contained in a neighborhood of i^0. However, at least in the AdS case our notion of boundary unitarity will remain. Attempts to make use of this effect to extract *a priori* ''lost'' information appear to involve extremely precise measurements of the gravitational flux Φ at infinity. For now, we merely note that such experiments are very difficult. Indeed, we expect that the coarse graining which leads to semiclassical black hole thermodynamics is mostly a lack of precision in measuring Φ. In this way, our perspective is consistent with that of [6], and also with [25] (where information is also lost simply by the erasure of quantum-mechanical detail in semiclassical measurements). This issue and the associated possible paradoxes will be explored further in [8].

There are many interesting issues that we have not addressed in this work. For example, we have in no way suggested a microscopic mechanism that would determine the entropy of black holes, or even to render it finite. As a result, we do not address the sort of unitarity questions raised in [22,26].

Even under the assumptions which led to unitarity of the S matrix, a second (related) issue that we have not addressed is the *rate* at which information is transferred to the Hawking radiation. To see the relation to the density of states, let us briefly summarize the picture of this process suggested by our arguments in the asymptotically flat context. Motived by our perturbative results, we first assumed that the algebra of observables near i^0 is complete, and contains full information (at least in some superselection sector). The most important observable was the gravitational flux Φ, which led to completeness when combined with the usual perturbative observables. However, an observer outside the black hole who uses, say, a set of particle detectors to extract information from the outgoing Hawking radiation does not measure Φ directly. Instead, the flux of stress energy in the Hawking radiation is related (via the gravitational Gauss' law) to the difference between Φ at i^0 and the corresponding gravitational flux $\Phi_{\rm horizon}$ at the black hole horizon. If one assumes that the density of states associated with $\Phi_{\rm horizon}$ is given by the Bekenstein-Hawking formula, then one can predict the rate at which information is transferred to the Hawking radiation. This amounts essentially to the classic analysis of [27]. However, we again emphasize that we have provided no detailed justification for this assumption here.

What we have done is to point out that, if the black hole evaporates completely, the constraints then relate Φ directly to the stress tensor. At this point there is no analogue of $\Phi_{\rm horizon}$ and the information has become fully encoded in the Hawking radiation. Furthermore, even before the black hole evaporates fully, we see that the horizon need not limit the transfer of information to outgoing radiation. Since information associated with particle degrees of freedom inside the black hole is also encoded in the gravitational field outside the black hole (e.g., in Φ), local physics outside the horizon is in principle sufficient to imprint this information on the Hawking radiation.

The essential point in our discussion was that the Hamiltonian of a classical diffeomorphism-invariant theory is a pure boundary term. A similar feature holds in quantum perturbation theory, and it seems reasonable to conjecture this property to hold in a nonperturbative quantum theory—even if the concepts of spacetime and

diffeomorphism-invariance themselves break down. This conjecture seems to hold, for example, in AdS/CFT [28], see [15,29].

As we have seen, the logical consequence of this property is that the asymptotic fields store information in a way that would not be possible in a local quantum field theory. It is clear that such arguments can be generalized to many other boundary conditions. A generalization may also hold for the case of closed cosmologies. There one imagines that a physical clock might play the role of the boundaries used above. In perturbation theory, the gravitational constraints will tie the energy of such a clock to the integral of the linearized stress tensor of the gravitational degrees of freedom, so that it might be used much like the gravitational flux Φ in our work above. Indeed, one might model such an observer by replacing their worldline with an interior boundary. We will save the detailed exploration of such ideas for future work.

ACKNOWLEDGMENTS

The author has benefited from discussions with numerous physicists at UCSB, the Perimeter Institute, the ICMS workshop on Gravitational Thermodynamics and the Quantum Nature of Space Time in Edinburgh, and the ICTS Monsoon Workshop on String theory in Mumbai. In particular, he is grateful to Alejandra Castro, Sergei Gukov, Steve Giddings, Gary Horowitz, Veronika Hubeny, Joe Polchinski, Mukund Rangamani, Rafael Sorkin, Lenny Susskind, Aron Wall, and especially to Ted Jacobson and Simon Ross for their thought-provoking comments and questions and to Maulik Parikh for a careful reading of an earlier draft. We also thank Stefan Hollands for pointing out the connection to Holmgren's uniqueness theorem mentioned in footnote 7. This work was supported in part by the U.S. National Science Foundation under Grant No. PHY05-55669, and by funds from the University of California. The author thanks the Tata Institute for Fundamental Research and the International Center for Theoretical Sciences for their hospitality and support during the final stages of this project.

[1] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).
[2] R. Wald, Living Rev. Relativity **4**, 6 (2001).
[3] R. Arnowitt, S. Deser, and C. W. Misner, Nuovo Cimento **19**, 668 (1961); J. Math. Phys. (N.Y.) **1**, 434 (1960); arXiv: gr-qc/0405109.
[4] L. Susskind, J. Math. Phys. (N.Y.) **36**, 6377 (1995); D. Bigatti and L. Susskind, arXiv:hep-th/0002044.
[5] T. Banks, L. Susskind, and M. E. Peskin, Nucl. Phys. **B244**, 125 (1984).
[6] V. Balasubramanian, D. Marolf, and M. Rozali, Gen. Relativ. Gravit. **38**, 1529 (2006); Int. J. Mod. Phys. D **15**, 2285 (2006).
[7] M. K. Parikh, arXiv:hep-th/0402166; Int. J. Mod. Phys. D **13**, 2351 (2004); Gen. Relativ. Gravit. **36**, 2419 (2004).
[8] D. Marolf, Phys. Rev. D **79**, 024029 (2009).
[9] S. Weinberg, Phys. Rev. **140**, B516 (1965).
[10] A. Ashtekar and R. O. Hansen, J. Math. Phys. (N.Y.) **19**, 1542 (1978); A. Ashtekar, in *General Relativity and Gravitation: One Hundred Years after the Birth of Albert Einstein*, edited by A. Held (Plenum Press, New York, 1980); A. Ashtekar and J. D. Romano, Classical Quantum Gravity **9**, 1069 (1992).
[11] A. Ashtekar, L. Bombelli, and O. Reula, in *Analysis, Geometry and Mechanics: 200 Years After Lagrange*, edited by M. Francaviglia and D. Holm (North-Holland, Amsterdam, 1991).
[12] A. Cresswell and R. L. Zimmerman, Gen. Relativ. Gravit. **3**, 1221 (1986).
[13] D. Marolf, arXiv:0810.4886.
[14] C. Fefferman and C. R. Graham, in *Élie Cartan et les Mathématiques d'Aujourd'hui* (Asterisque, Paris, 1985), p. 95.
[15] M. Henningson and K. Skenderis, J. High Energy Phys. 07 (1998) 023.
[16] P. Breitenlohner and D. Z. Freedman, Ann. Phys. (N.Y.) **144**, 249 (1982); Phys. Lett. **115B**, 197 (1982).
[17] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, Phys. Rev. D **73**, 086003 (2006); **74**, 066009 (2006).
[18] L. Hormander, J. Diff. Geom. **6**, 129 (1971).
[19] J. V. Rocha, J. High Energy Phys. 08 (2008) 075.
[20] T. Banks, Nucl. Phys. **B309**, 493 (1988); S. B. Giddings and A. Strominger, Nucl. Phys. **B321**, 481 (1989).
[21] S. R. Coleman, Nucl. Phys. **B307**, 867 (1988); I. R. Klebanov, L. Susskind, and T. Banks, Nucl. Phys. **B317**, 665 (1989); J. Preskill, Nucl. Phys. **B323**, 141 (1989).
[22] J. M. Maldacena, J. High Energy Phys. 04 (2003) 021.
[23] B. Freivogel, V. E. Hubeny, A. Maloney, R. C. Myers, M. Rangamani, and S. Shenker, J. High Energy Phys. 03 (2006) 007.
[24] W. K. Wootters and W. H. Zurek, Nature (London) **299**, 802 (1982); D. Dieks, Phys. Lett. **92A**, 271 (1982).
[25] V. Balasubramanian, V. Jejjala, and J. Simon, Int. J. Mod. Phys. D **14**, 2181 (2005); V. Balasubramanian, J. de Boer, V. Jejjala, and J. Simon, J. High Energy Phys. 12 (2005) 006.
[26] S. W. Hawking, Phys. Rev. D **72**, 084013 (2005).
[27] D. N. Page, Phys. Rev. Lett. **71**, 1291 (1993).
[28] J. M. Maldacena, Adv. Theor. Math. Phys. **2**, 231 (1998); Int. J. Theor. Phys. **38**, 1113 (1999).
[29] V. Balasubramanian and P. Kraus, Commun. Math. Phys. **208**, 413 (1999).