Unitarity and Holography in Gravitational Physics

Donald Marolf
Physics Department, UCSB, Santa Barbara,
CA 93106, USA
marolf@physics.ucsb.edu

ABSTRACT: 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 non-perturbative 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 non-perturbative level, 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 non-perturbatively, 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 \( t_2 \).

KEYWORDS: Gravity, Black Holes, Unitarity.
1. 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 $\ell_p$, where back-reaction is ignored. However, strict locality explicitly fails at first interacting order. A clean signal is the fact that a form of time evolution is generated by a boundary term at spacelike infinity (e.g., the ADM energy in asymptotically flat space [3]). This feature is related to the lack of local observables in diffeomorphism-invariant theories.

We show below that this simple observation leads to interesting results. Even in the context of perturbation theory about asymptotically flat collapsing black hole backgrounds, at first interacting order the algebra generated by fields on future null infinity ($I^+$) in any neighborhood of spacelike infinity ($i^0$) contains a complete set of observables. In the asymptotically AdS context, the algebra of boundary observables defined by any neighborhood of a boundary Cauchy surface is similarly complete. As a result, full information about the quantum state is always contained in the asymptotic fields.

The reader may object that perturbative quantum gravity appears both local and causal in, say, de Donder gauge. However, locality of observables does not follow due to the presence of longitudinal gravitons and the associated residual gauge symmetries. As is familiar from the Coulomb gauge in Maxwell theory, gauge fixing all residual symmetries removes the apparently manifest locality. In Yang-Mills theory one can avoid these issues by constructing Wilson loops which provide a complete set of compactly supported observables. However, no such compactly supported observables are available in a diffeomorphism-invariant theory. Instead, observables invariant under the full set of gauge symmetries have the non-local properties we require.

We refer to the completeness results above as ‘perturbative holography,’ though 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].

We begin with the asymptotically flat context in section 2. After deriving perturbative completeness of the algebra near \( i^0 \), we consider implications for the non-perturbative theory. Assuming that the same algebra remains complete at the non-perturbative level, 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 section 3. We also note that, whether or not the stated algebra continues to be complete non-perturbatively, 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 section 4.

2. Quantum Gravity in Asymptotically Flat Space

To avoid making detailed assumptions about the quantum nature of gravity, it natural to proceed using either semi-classical 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 section 2.1. Section 2.2 then studies the implications for the non-perturbative theory and discusses unitarity of the S-matrix.

2.1 The Holographic nature of perturbative gravity

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 \( I^+ \). I.e., 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 figure 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
The number of dimensions would not significantly change the analysis.\footnote{Except that higher dimensions improve the infrared behavior. 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 \cite{8} allow us to use gravitational perturbation theory and to speak of an S-matrix. In higher dimensions, no such divergences arise.} The essential inputs are only 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 $\kappa$ 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 non-trivial 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 non-local, 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 Arnowitt-Deser-Misner (ADM) boundary term at spatial infinity ($i^0$). We denote this boundary term $\Phi$ as it will be convenient to think of this term as a gravitational flux. We have

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

(2.1)

where $r^a$ is a radial unit normal, $C$ is a cut of $i^0$ as defined e.g. in \cite{9}, $dA$ is the area element on $C$, and $D_a$ is the covariant derivative defined by the fixed flat metric $\eta_{ab}$ 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. \cite{10}). 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. \cite{11}). Either expression is purely quadratic in $g_{ab} - \eta_{ab}$, where $\eta_{ab}$ is a fixed flat metric at infinity. This calculation shows explicitly that $\Phi$ 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 $H_{\text{lin}}^{\tilde{h} + h}$. 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 $H_{h+h}^{\text{lin}}$ is quadratic, it is straightforward to expand in powers of $h_{ab}$:

$$H_{h+h}^{\text{lin}} = \tilde{E} + S + H_{h}^{\text{lin}}. \tag{2.2}$$

Here $\tilde{E}$ is the 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 $H_{h}^{\text{lin}}$ is just the integral of the (quadratic) stress tensor for perturbations $h_{ab}$ propagating on the flat metric $\eta_{ab}$.

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 $\Phi$ non-trivial. 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

$$\Phi = H_{h+h}^{\text{lin}} \tag{2.3}$$

holds as an equality of Heisenberg-picture quantum operators. Furthermore, we see that $\Phi$ generates $v$-translations of $\tilde{h}_{ab} + h_{ab}$ in the sense that

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

The terms involving $\tilde{h}_{ab}$ on the final right-hand-side are associated with the source terms in (2.2), or equivalently with the difference between $\Phi$ and $H_{h}^{\text{lin}}$. Equation (2.4) is a key result which we will use liberally. Note that while $h_{ab}(v)$ 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,$^2$ we consider $i^0$ to be a point on $I^+$ with $u = -\infty$ so that $\mathcal{A}_{u_1}$ contains $\Phi$. Since we may use (2.4), it remains only to approximate $h_{ab}(v - \tau)$ by operators in $\mathcal{A}_{u_1}$.

$^2$We could also have used $\Phi(u)$ associated with a cut of $I^+$ at retarded time $u$ to approximate $\Phi$ as $u \to -\infty$, but our argument loses nothing by making the above simplifying convention.
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 ($\tilde{g}_{ab}$ 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 $A^+_{u_1}$. Beyond the geometric optics approximation, and taking into account non-linear corrections at some fixed order of perturbation theory, we may use the equations of motion to write

$$h_{ab}(v - \tau) = O_{ab}(v - \tau, u_1) + \Delta_{ab}(v - \tau, u_1),$$

(2.5)

where $O_{ab}(v - \tau, u_1) \in A^+_{u_1}$ 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 in the limit $v_1 - (v - \tau) \to \infty$.

Using the same perturbative equations of motion, $\Delta_{ab}(v - \tau, u_1)$ may be expressed as an operator on $I^-$ for which the largest contributions come from the region around $v_1(u_1, L)$. As a result, the operator $e^{-i\tau\Phi} \Delta_{ab}(v - \tau, u_1)e^{i\tau\Phi}$ is largely associated with the region around $v_1(u_1, L) + \tau$. Since as $v \to +\infty$ the correlation functions in any Fock space state approach those of the vacuum, we see that all matrix elements of $e^{-i\tau\Phi} \Delta_{ab}(v - \tau, u_1)e^{i\tau\Phi}$ agree with the vacuum expectation value as $\tau \to \infty$; i.e., for large $\tau$ the operator becomes effectively a c-number determined by the background $\tilde{g}_{ab}$. We write $e^{-i\tau\Phi} \Delta_{ab}(v - \tau, u_1)e^{i\tau\Phi} \to c_{ab}(v - \tau, u_1, \tau)$.

Combining the above results we have

$$h_{ab}(v) = \lim_{\tau \to \infty} \left[ e^{-i\tau\Phi} O_{ab}(v - \tau, u_1)e^{i\tau\Phi} + \tilde{h}_{ab}(v - \tau) - \tilde{h}_{ab}(v) + c_{ab}(v - \tau, u_1, \tau) \right].$$

(2.6)

Since any c-number lies in $A^+_{u_1}$, the right-hand side contains only elements of $A^+_{u_1}$ as desired. We have shown that any fundamental field on $I^-$ can be expressed with arbitrary accuracy as an element of $A^+_{u_1}$. 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 $A^+_{u_1}$. For convenience, we used a Coulomb-like gauge, but the corresponding result for gauge-invariant observables follows immediately in any gauge.

### 2.2 Non-perturbative gravity and Unitarity of the S-matrix

We saw above that perturbative gravity in asymptotically flat space is holographic in the sense that the algebra of observables generated by $\Phi$ 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 time-like infinity in a black hole spacetime. As a result, in perturbation theory about such a background, the total gravitational flux $\Phi$ 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 $i^-$. However, in the non-perturbative theory one expects any black hole that forms to decay by Hawking evaporation. The end products may then plausibly be well-described by perturbative quantum gravity about flat spacetime, in which case $i^+$ is regular. Let us therefore suppose that, in any state of the non-perturbative theory, perturbative quantum gravity about flat spacetime becomes an arbitrarily good approximation for studying 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 $\Phi$ and asymptotic fields on $I^+$ in any $A_{i^+}$ again contains a complete set of observables. Since we have a regular $i^+$, the gravitational flux $\Phi$ 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 $\Phi$ 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 $I^\pm$. Since local operators can be translated between $I^+$ and $I^-$, these constructions merely define two bases for the same Hilbert space. Because the inner product of any two such states (say, in the $I^-$ basis) can be expressed in terms of correlators of local fields in the Fock vacuum, and because the Fock vacuum defines the same state on fields at $I^-$ as at $I^+$, it follows that this inner product agrees with the corresponding inner product of states in the $I^+$ basis. The S-matrix is unitary.

\[^3\text{Note that this is necessarily an assumption. In particular, it does not follow from the assumption that perturbation theory is arbitrarily good near infinity. Our previous perturbative argument required us to propagate fields from $I^-$ to $I^+$ through the bulk of the spacetime where non-perturbative effects can be important. The purpose of our perturbative argument is to render this assumption plausible by removing objections based on perturbative fields falling into semi-classical black holes.}\]
3. 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 \[ g_{ab} = \frac{\ell^2}{r^2} dr^2 + \left( g^{(0)}_{ij} \frac{r^2}{\ell^2} + g^{(1)}_{ij} \frac{r}{\ell} + g^{(2)}_{ij} + g^{(3)}_{ij} \frac{\ell}{r} + \cdots \right) dx^i dx^j, \] (3.1)

for some fixed boundary metric \( g^{(0)}_{ij} \). Here \( \ell \) is the AdS scale, the \( x^i \) are coordinates on \( S^2 \times \mathbb{R} \), and the \( \cdots \) represent higher order terms in \( r/\ell \) which may include cross terms of the form \( dr dx^i \). 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.

Certain simplifications arise if we couple the gravitational field to a conformally coupled scalar field \( \phi \), 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. \[ \phi = \frac{\alpha}{r} + \frac{\beta}{r^2} + \cdots, \] (3.2)

where \( \alpha \) 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 \( \alpha \) will be chosen to generate scalar radiation which collapses to form a black hole\(^4\). 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 \( \Sigma \) in the initial pure AdS region. It is clear that any field at any later time can be expressed in terms of fields on \( \Sigma \). Similarly, in the linearized approximation, any field on \( \Sigma \) can be expressed in terms of the boundary fields \( \tilde{g}^{(3)}_{ab} \) and \( \beta \) at earlier times. Some explicit formulae for the scalar case\(^5\) appear in e.g. \[15\], but the fact that this is possible follows immediately from the observation that any linearized solution with given \( \delta g^{(0)}_{ab} \) and \( \alpha \) is determined by the values of \( \delta g^{(0)}_{ab}, \alpha, g^{(3)}_{ab}, \) and \( \beta \) to the past of \( \Sigma \). This in turn follows from a simple argument: Suppose that two such solutions have

\(^4\)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.

\(^5\)The explicit formula in \[15\] express local bulk fields in terms of boundary fields in a compact region of 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.
the the same values of $\delta g_{0(ab)}$, $\alpha$, $g_{3(ab)}$, and $\beta$ to the past of $\Sigma$, 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, its energy vanishes in the distant past. But by construction this solution conserves energy, so it must vanish at all times. We thus conclude that any linearized field on $\Sigma$ can be expressed in terms of the boundary fields $g_{(3)ab}$ and $\beta$ at earlier times. It follows that the same result holds at each order in perturbation theory.

To complete the main argument, simply note that the algebra of boundary operators $\mathcal{A}_{t,\Delta t}$ supported within any time $\Delta t$ of the boundary time $t$ contains the Hamiltonian. Thus we may in fact express any perturbative field on $\Sigma$ as an element of $\mathcal{A}_{t,\Delta t}$ for any $t, \Delta t$, 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. 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, this assumption is manifestly false at any finite order in perturbation theory about a black hole background.

Let us now briefly consider a non-perturbative theory. In asymptotically flat space we assumed that perturbative 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 non-standard boundary conditions that allow the particles to leave the original AdS space. E.g., we could consider the evaporon model of [16]. 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 $U(t_1, t_2) = \mathcal{P} \exp \left( -i \int_{t_1}^{t_2} H(t) dt \right)$, 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, or that all information is present at the boundary. However, we can immediately conclude that each $\mathcal{A}_{t,\Delta t}$ contains the same set of observables. In this sense, any information which happens to present at the boundary at any time $t_1$ remains present at any other time $t_2$.

This result is naturally called ‘boundary unitarity.’ To provide some physical interpretation, 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 all boundary observables, boundary unitarity will allow her to extract at any time $t_2$ any information which she may have 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 $\Phi$ 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 [17].

4. Discussion

We have shown 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 full algebra is contained in the observable algebra within 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 our analysis, confirming the scenario outlined in [6].

If this same algebra remains complete at the non-perturbative level, and if perturbative quantum gravity about flat space is a good approximation to some asymptotically flat non-perturbative 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. It is interesting to classify failures of this latter assumption 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 might be the case in so-called third-quantized theories [18], 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 [19].

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 non-perturbative 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 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 [20, 21].
The reader may worry that the presence of so much information near infinity might violate the “no quantum xerox theorem” [22]. 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 non-commuting 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. In the contexts discussed here, the fact that the boundary observer has access to all degrees of freedom, including the measuring devices of the local observer, ensures that 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.

A related and even more interesting context arises when we add a second boundary to the spacetime and suppose the two boundaries to be in causal contact. 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. Because we wish the boundaries to be in causal contact, the new boundary will be at finite distance.

In the asymptotically flat version, this finite boundary may prohibit a regular $i^+$ and may also interfere with the scattering of wavepackets 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 appear to involve extremely precise measurements of the gravitational flux $\Phi$ at infinity. For now, we merely note that such experiments are very difficult. Indeed, we expect that the coarse-graining which leads to semi-classical black hole thermodynamics is mostly a lack of precision in measuring $\Phi$. In this way, our perspective is consistent with that of [23], where information is also lost simply by the erasure of quantum mechanical detail in semiclassical measurements. We will explore this issue and the associated possible paradoxes further in the near future [17].

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 [20, 24].

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, we briefly summarize the picture of this process suggested by our arguments in the asymptotically flat context. We first argued that the algebra of observables near $i^0$ is always complete, and contains full information. The most important observable was the gravitational flux $\Phi$, 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 $\Phi$ directly. Instead, the flux of
stress-energy in the Hawking radiation is related (via the gravitational Gauss’ law) to the difference between $\Phi$ at $i^0$ and the corresponding gravitational flux $\Phi_{\text{horizon}}$ at the black hole horizon. If one assumes that the density of states associated with $\Phi_{\text{horizon}}$ 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 [25]. 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 $\Phi$ directly to the stress tensor. At this point there is no analogue of $\Phi_{\text{horizon}}$ 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 $\Phi$), 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 non-perturbative quantum theory – even if the concepts of spacetime and diffeomorphism-invariance themselves break down. This conjecture seems to hold, for example, in AdS/CFT [26], see [13, 27].

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 $\Phi$ 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.

Acknowledgements

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 Alejandro Castro, Sergei Gukov, Steve Giddings, Gary Horowitz, Veronika Hubeny, Joe Polchinski, Mukund Rangamani, Rafael Sorkin, and 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. This work was supported in part by the US National Science Foundation under Grant No. PHY05-55669, and by funds from the University of California. The au-
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.

References

[1] S. W. Hawking, Particle Creation by Black Holes, Commun. Math. Phys. 43, 199 (1975).

[2] R. Wald, “The thermodynamics of black holes,” R. Wald, Living Rev. Rel. 4, 6 (2001) [arXiv:gr-qc/9912119].

[3] R. Arnowitt, S. Deser and C. W. Misner, Nuovo Cimento, 19 (1961) 668; R. Arnowitt, S. Deser and C. W. Misner, J. Math. Phys. 1 (1960) 434; R. Arnowitt, S. Deser and C. W. Misner, “The Dynamics Of General Relativity,” [arXiv:gr-qc/0405109].

[4] L. Susskind, “The World As A Hologram,” J. Math. Phys. 36, 6377 (1995) [arXiv:hep-th/9409089]; D. Bigatti and L. Susskind, “TASI lectures on the holographic principle,” arXiv:hep-th/0002044.

[5] T. Banks, L. Susskind and M. E. Peskin, “Difficulties For The Evolution Of Pure States Into Mixed States,” Nucl. Phys. B 244, 125 (1984).

[6] V. Balasubramanian, D. Marolf and M. Rozali, “Information recovery from black holes,” Gen. Rel. Grav. 38, 1529 (2006) [Int. J. Mod. Phys. D 15, 2285 (2006)] [arXiv:hep-th/0604045].

[7] M. K. Parikh, “Energy conservation and Hawking radiation,” arXiv:hep-th/0402166; “A secret tunnel through the horizon,” Int. J. Mod. Phys. D 13, 2351 (2004) [Gen. Rel. Grav. 36, 2419 (2004)] [arXiv:hep-th/0405160].

[8] S. Weinberg, “Infrared photons and gravitons,” Phys. Rev. 140, B516 (1965).

[9] A. Ashtekar and R. O. Hansen, “A Unified Treatment of Null and Spatial Infinity in General Relativity. I. Universal Structure, Asymptotic Symmetries, and Conserved Quantities at Spatial Infinity,” J. Math. Phys., 19 (1978) 1542; A. Ashtekar, “Asymptotic structure of the gravitational field at spatial infinity” in General relativity and gravitation : one hundred years after the birth of Albert Einstein, edited by A. Held (New York, Plenum Press, 1980); A. Ashtekar and J. D. Romano, “Spatial infinity as a boundary of space-time,” Class. Quant. Grav. 9 (1992) 1069.

[10] A. Ashtekar, L. Bombelli and O. Reula, “The Covariant Phase Space Of Asymptotically Flat Gravitational Fields,” in Analysis, Geometry and Mechanics: 200 Years After Lagrange, edited by M. Francaviglia and D. Holm (North-Holland, Amsterdam, 1991).

[11] A. Cresswell and R. L. Zimmerman, “Angular Momentum in General Relativity,” Gen. Rel. Grav. 3 (1986) 1221.

[12] C. Fefferman and C.R. Graham, “Conformal Invariants,” in Élie Cartan et les Mathématiques d’Aujourd’hui (Asterisque, 1985) 95.

[13] M. Henningson and K. Skenderis, “The holographic Weyl anomaly,” JHEP 9807, 023 (1998) [arXiv:hep-th/9806087].

[14] P. Breitenlohner and D. Z. Freedman, “Stability In Gauged Extended Supergravity,” Annals Phys. 144, 249 (1982). P. Breitenlohner and D. Z. Freedman, “Positive Energy In Anti-De Sitter Backgrounds And Gauged Extended Supergravity,” Phys. Lett. B 115, 197 (1982).
[15] A. Hamilton, D. N. Kabat, G. Lifschytz and D. A. Lowe, “Local bulk operators in AdS/CFT: A boundary view of horizons and locality,” Phys. Rev. D 73, 086003 (2006) [arXiv:hep-th/0506118]; A. Hamilton, D. N. Kabat, G. Lifschytz and D. A. Lowe, “Holographic representation of local bulk operators,” Phys. Rev. D 74, 066009 (2006) [arXiv:hep-th/0606141].

[16] J. V. Rocha, “Evaporation of large black holes in AdS: coupling to the evaporon,” arXiv:0804.0055 [hep-th].

[17] D. Marolf, “Holographic Thought Experiments,” to appear slightly later in today’s arxiv listing. Tentatively arXiv: 0808.2845 [gr-qc].

[18] T. Banks, “Prolegomena to a Theory of Bifurcating Universes: A Nonlocal Solution to the Cosmological Constant Problem Or Little Lambda Goes Back to the Future,” Nucl. Phys. B 309, 493 (1988); S. B. Giddings and A. Strominger, “Baby Universes, Third Quantization and the Cosmological Constant,” Nucl. Phys. B 321, 481 (1989).

[19] S. R. Coleman, “Black Holes as Red Herrings: Topological Fluctuations and the Loss of Quantum Coherence,” Nucl. Phys. B 307, 867 (1988); I. R. Klebanov, L. Susskind and T. Banks, “Wormholes and the Cosmological Constant,” Nucl. Phys. B 317, 665 (1989). J. Preskill, “Wormholes in space-time and the constants of nature,” Nucl. Phys. B 323, 141 (1989).

[20] J. M. Maldacena, “Eternal black holes in Anti-de-Sitter,” JHEP 0304, 021 (2003) [arXiv:hep-th/0106112].

[21] B. Freivogel, V. E. Hubeny, A. Maloney, R. C. Myers, M. Rangamani and S. Shenker, “Inflation in AdS/CFT,” JHEP 0603, 007 (2006) [arXiv:hep-th/0510046].

[22] W. K. Wootters and W. H. Zurek, “A single quantum cannot be cloned,” Nature 299 (1982) 802; D. Dieks, “Communication By Epr Devices,” Phys. Lett. A 92, 271 (1982).

[23] V. Balasubramanian, V. Jejjala and J. Simon, The library of Babel, Int. J. Mod. Phys. D 14, 2181 (2005) [arXiv:hep-th/0505123]; V. Balasubramanian, J. de Boer, V. Jejjala and J. Simon, The library of Babel: On the origin of gravitational thermodynamics, JHEP 0512, 006 (2005) [arXiv:hep-th/0508023].

[24] S. W. Hawking, “Information loss in black holes,” Phys. Rev. D 72, 084013 (2005) [arXiv:hep-th/0507171].

[25] D. N. Page, “Expected Entropy Of A Subsystem,” Phys. Rev. Lett. 71, 1291 (1993) [arXiv:gr-qc/9305007]. D. N. Page, Black hole information, in Proceedings of the 5th Canadian Conference on General Relativity and Relativistic Astrophysics, ed. R. B. Mann and R. G. McLenaghan (1994) [arXiv:hep-th/9305040].

[26] J. M. Maldacena, “The large N limit of superconformal field theories and supergravity,” Adv. Theor. Math. Phys. 2, 231 (1998) [Int. J. Theor. Phys. 38, 1113 (1999)] [arXiv:hep-th/9711200].

[27] V. Balasubramanian and P. Kraus, “A stress tensor for anti-de Sitter gravity,” Commun. Math. Phys. 208, 413 (1999) [arXiv:hep-th/9902121].

– 13 –