Decoding the Apparent Horizon: A Coarse-Grained Holographic Entropy

Netta Engelhardt
Department of Physics, Princeton University, Princeton NJ 08544 USA

Aron C. Wall
Institute for Advanced Study, Einstein Drive, Princeton NJ 08540 USA

When a black hole forms from collapse in a holographic theory, the information in the black hole interior remains encoded in the boundary. We prove that the area of the black hole’s apparent horizon is precisely the entropy associated to coarse graining over the information in its interior, subject to knowing the exterior geometry. This is the maximum holographic entanglement entropy that is compatible with all classical measurements conducted outside of the apparent horizon. We identify the boundary dual to this entropy and explain why it obeys a Second Law of Thermodynamics.

INTRODUCTION

The Second Law of Thermodynamics states that entropy increases with time. One natural notion of entropy is the von Neumann entropy:

$$S[\rho] = -\text{tr}(\rho \ln \rho),$$

(1)

where $\rho$ is the density matrix of a quantum system. However, this quantity is conserved under unitary time evolution, in apparent tension with the Second Law. To obtain an increasing entropy, it is necessary to coarse grain $S$ by “forgetting” certain information, since the vast majority of microscopic data in a thermal system is inaccessible to macroscopic observations. One common coarse-graining method is the maximization of the system’s entropy subject to fixed the values of a set of feasible macroscopic measurements $M(t)$ at a moment in time [13]:

$$S^{\text{coarse}}(t) = \max_{\rho'} (S[\rho'] : M(t)).$$

(2)

Assuming that any ordered information inaccessible at early times remains so at later times, $S^{\text{coarse}}$ should increase with time, defining a nontrivial Second Law.

The most mysterious application of the Second Law is to black holes. Stationary black holes (e.g. Kerr) have entropy, which is proportional to the area of their horizon $H$ [4, 5]:

$$S_{BH} = \frac{\text{Area}[H]}{4G\hbar},$$

(3)

as suggested by the Laws of Black Hole Mechanics [4]. However, despite some clues from string theory and other approaches (reviewed in [8]), it is still unclear in general what microscopic quantum-gravitational degrees of freedom are counted by this entropy. Dynamically evolving black holes such as those formed from stellar collapse are even more controversial, since there are multiple possible definitions of a horizon, e.g. the event horizon and the apparent horizon [9] — and correspondingly, multiple area increase theorems [6, 10–14].

In holographic models of quantum gravity, a black hole is dual to some boundary state $\rho$ whose von Neumann entropy $S[\rho]$ can be computed from a compact extremal (HRT) surface in the bulk [15, 18]:

$$S[\rho] = \frac{\text{Area}[X_{\text{HRT}}]}{4G\hbar}. $$

(4)

A surface is extremal if its area is unchanged by any first order perturbation to the surface’s location; if there is more than one, $X_{\text{HRT}}$ is the one with the minimal area extremal surface (and homologous to the boundary [16, 19]). This quantity is independent of time, so it is not suitable for describing the entropy increase of a growing black hole. Unitarity of the boundary theory implies that no information is lost, but this is not enough: to account for the increase of black hole entropy, a coarse graining scheme must be specified.

Even though black hole thermodynamics was the original motivation for the holographic principle [20, 21], no one has yet given a clear explanation of the role of the black hole horizon as a repository of information about the interior. Indeed, it was recently shown [22] that if we know the outcome of all classical measurements $M(t)$ outside of the event horizon $H$, then $S^{\text{coarse}} < \text{Area}[H]/4G\hbar$: we have access to too much information for our remaining ignorance to be given by the event horizon’s area (thus refuting a broad class of proposals relating entropy to area, including [23, 26]).

We therefore look for alternatives to the event horizon. An appealing option is the apparent horizon $\mu$, the outermost compact surface (at a moment of time) which is marginally outer trapped [9], i.e. the the expansion $\theta_k \equiv \nabla_k \ln(\text{Area}[n]) = 0$, where $k$ is a future-outwards null vector, and $n$ is a small pencil of lightrays shot out in the $k$-direction from a small neighborhood of a point on
μ. In the case of a black hole that forms from collapse, such marginally trapped surfaces form behind the event horizon, even though the HRT surface is the empty set (so that the boundary state is pure).

In this Letter, we give a classical geometric proof that the area of the apparent horizon μ does play the role of a coarse-grained entropy:

\[ S_{\text{coarse}} = \frac{\text{Area}[\mu]}{4G\hbar}, \]  

where we coarse grain over the region behind the apparent horizon (the “microstates”) while holding all classical measurements in the exterior fixed. This makes it plausible that the interior is encoded holographically by set of independent qubits, one per 4/ln 2 Planck-areas, on the apparent horizon (but not the event horizon!) [27][30]. Our classical proof explicitly constructs the entropy-maximizing geometry, which would correspond to maximally scrambling all of these qubits. If our result can be extended to the quantum regime (along the lines of [31][35] it might provide insight into the firewalls paradox [36][39], a puzzle about whether maximally scrambling all of these qubits. If our result can be extended to the quantum regime (along the lines of [31][35] it might provide insight into the firewalls paradox [36][39], a puzzle about whether maximally scrambling all of these qubits.

Hence, the area of the apparent horizon has a statistical interpretation as the maximum boundary entropy that is compatible with the geometry of its exterior. This provides a holographic answer to the disputed question: what does the Bekenstein-Hawking entropy of a black hole count? [41][46]

Outline of Proof: Let k (respectively ℓ) be the orthogonal future-directed null vectors pointing outward (respectively inward) from a surface. An extremal surface X satisfies \( \theta_k = \theta_\ell = 0 \). An HRT surface additionally must be the minimal area surface (homologous to the boundary) on some spatial slice \( \Sigma \) [47].

An apparent horizon \( \mu \) (an outermost marginally trapped surface) satisfies \( \theta_k = 0, \theta_\ell \leq 0 \), and (generically) \( \nabla_k \theta_\ell < 0 \) [48][49]. We assume that \( \mu \) is homologous to the boundary, i.e. there exists a spatial slice \( \Sigma \) connecting \( \mu \) to the boundary, and moreover that there exists a \( \Sigma \) such that the area of any surface circumscribing \( \mu \) is larger than the area of \( \Sigma \). These requirements are reasonable for black hole horizons.

In any spacetime, \( \text{Area}[X_{\text{HRT}}] \leq \text{Area}[\mu] \); this can be proven by a simple focusing argument: in a spacetime satisfying the Null Energy Condition (\( T_{\nu\nu} \geq 0 \) for any null vector \( \nu \)), a null surface \( N_{\pm k}[\mu] \) shot out along the ±k-direction of \( \mu \) has monotonically decreasing area, where we truncate the surface when generators intersect [9][50][51]. We extend \( N_{\pm k} \) to the slice \( \Sigma \) on which \( X_{\text{HRT}} \) is minimal [47].

\[ \text{Area}[\mu] \geq \text{Area}[\Sigma \cap N_{\pm k}[\mu]] \geq \text{Area}[X_{\text{HRT}}]. \]  

Hence the entropy \( S[\rho'] \) cannot exceed \( \text{Area}[\mu]/4G\hbar \).

To prove that this inequality is saturated, we construct a bulk spacetime \( M' \) (with the same outer wedge \( O_W[\mu] \)) satisfying \( \text{Area}[X_{\text{HRT}}] = \text{Area}[\mu] \). To specify
the interior data in \( M' \), we impose initial data on the boundary of \( O_W[\mu] \), a null surface. We shall refer to this null surface as \( N_{-k} \), as it is the null surface fired from \( \mu \) in the \(-k\) direction. We choose our initial data so that the surface \( N_{-k} \) is stationary; every cross-section has the same geometry. (The Appendix shows this construction satisfies the constraint equations, so that a spacetime solution \( M' \) exists.)

By following \( N_{-k} \) far enough, we eventually come to an extremal surface \( X \). Since \( N_{-k} \) is stationary, \( \text{Area}[X] = \text{Area}[\mu] \). We can complete the spacetime by requiring it to be invariant under a CPT-reflection about \( X \) (i.e. we reflect space and time about \( X \) while exchanging matter with antimatter). See Fig. 1. The resulting bulk \( M' \) has two asymptotic boundaries, and therefore represents a pure state (analogous to the thermal field double wormhole construction [52]). When the state \( \rho' \) is restricted to a single boundary, the entropy \( S[\rho'] = X_{HRT}[M'] \). (Note that the region \( O_W[X_{HRT}] \) agrees with the original bulk geometry dual to \( \rho' \) [47, 53–57].)

**FIG. 1.** The coarse-grained spacetime dual to the state \( \rho' \) with maximal \( S[\rho'] \) and fixed \( O_W[\mu] \) (shaded gray). The null congruence \( N_{-k} \) (red) is fired from \( \mu \) towards the \(-k\) direction and is stationary. The congruence \( N_{-l} \), the past boundary of \( O_W[\mu] \), is fired in the \(-l\) direction from \( \mu \). \( X \) is the HRT surface of the coarse-grained spacetime. Tilded quantities represent the CPT mirror reverse.

Because \( N_{-k} \) is stationary and \( \mu \) is minimal on a slice of \( O_W[\mu] \), we now have an initial data slice \( \Sigma \) on which \( X \) is the minimal cross-section. Any other extremal surface \( X' \) (even if it is not on the initial data Cauchy slice) has greater area than \( X \):

\[
\text{Area}[X'] \geq \text{Area}[\Sigma \cap N_{\pm k}[X']] \geq \text{Area}[X],
\]

where the first inequality comes from focusing of a null surface \( N_{\pm k}[X'] \) shot out from \( X' \). Hence \( X = X_{HRT} \), proving Eq. 7.

**SIMPLE ENTROPY**

Thus far, our coarse-grained entropy has been defined from the bulk point of view. We now identify the boundary dual to the outer entropy, which we call the simple entropy, as it relies on “simple operators”.

In AdS/CFT, single trace operators on the boundary correspond to locally propagating fields in the bulk. More generally, we expect that the product of a small number of single trace operators also propagates locally in the bulk. However, it is known that sufficiently complicated operators (known as precursors [58, 59]) can change the deep bulk region acausally; hence to define a coarse graining that is dual to \( O_W[\mu] \), we must avoid such complicated operations. We therefore define a “simple” experiment as a procedure performed after a moment of time \( t_i \), in which we measure a local operator \( \mathcal{O}(t > t_i) \) after having turned on a set of local sources \( J(t > t_i) \); we require that these sources propagate causally into the bulk. For classical solutions, we can restrict attention to one-point operators and sources, since the higher-point functions are determined from them. (The “one-point entropy” [26], proposed as a holographic dual to the area of the event horizon, did not allow sources.) To prevent recurrences, we implicitly include a late time cutoff \( t_f \) prior to exponentially large values of \( t \).

The simple entropy is now defined as the maximum entropy of a state \( \rho' \) compatible with the outcomes of all such simple experiments:

\[
S^{\text{(simple)}}(t_i) = \max_{\rho'} \left( S[\rho'] : \langle E^\dagger \mathcal{O}(t) E \rangle \right),
\]

where

\[
E = \mathcal{T} \exp[-i \int_{t_i}^{t} J(t') \mathcal{O}(t') dt']
\]

is the time-ordered insertion of sources \( J(t) \) used to prepare the simple experiment by which \( \mathcal{O}(t) \) is measured.

A simple experiment, by definition, can only access the subset of the bulk \( F(t_i) \) that is to the future of the boundary time \( t_i \). When the spacetime has a black hole, turning on simple sources can shift the location of any event horizon \( H \) in the spacetime [60]. However, the event horizon must always remain outside of any marginally trapped surface (assuming the Null Energy Condition) [9, 51]. Therefore, if \( \mu \) is a marginally trapped surface on \( N(t_i) \), the boundary of \( F(t_i) \), a simple experiment can access at most the outer wedge \( O_W[\mu] \). Note that by causality, turning on simple sources cannot modify the fact that \( \mu \) is marginally trapped (a similar argument was given for extremal sur-
faces in \([33]\). See Fig. 2(a). It immediately follows that

\[ S^{(\text{simple})}(t_i) \geq S^{(\text{outer})}[\mu]. \tag{12} \]

If \(N(t_i)\) contains more than one marginally trapped surface, we restrict attention to the earliest (i.e., outermost) one. This guarantees that \(\mu\) is in fact an apparent horizon. We propose that in this case, the inequality \([13]\) is saturated. In other words the simple entropy is the holographic dual of the area of the apparent horizon.

![Figure 2](image)

**FIG. 2.** (a) We fire a null congruence \(N_{-k}\) into the bulk from time \(t = t_i\). The surface \(\mu\) is the first cross-section of \(N_{-k}\) with vanishing \(k\) expansion. We can recover all the data in \(O_W[\mu]\), at least when the black hole is near equilibrium, by means of a “simple experiment” performed after time \(t_i\). (b) A spacelike holographic screen (purple) has increasing area in a spacelike direction, going from 1 to 3. The corresponding outer wedges are nested, implying that the outer entropy must increase outwards. Similarly, the simple entropy must increase with \(t\) from \(t_1\) to \(t_3\).

We now show that this is true for a black hole that is approaching thermal equilibrium after time \(t_i\). We may use the “HKLL” procedure \([61, 67]\) to reconstruct the “causal wedge” \(C_W[t_i]\) of \(t_i\), i.e., the subset of \(F(t_i)\) outside of the event horizon \([68, 69]\). If no matter or gravitational radiation were falling across the event horizon \(H\), it would be stationary; there would be no separation between \(H\) and \(\mu\), and we would be done. In order to reconstruct the data in \(O_W[\mu]\), we must ensure that no matter falls across \(H\) after \(\mu\).

Since \(\mu\) is perturbatively close to the event horizon, \([63, 70]\) allows us to map the matter fields falling across the event horizon to data on the boundary. We can therefore turn these fields “off” by adding suitable sources to the boundary after \(t_i\). This has the effect of shifting the event horizon to the location of \(\mu\), so that \(C_W[t_i] = O_W[\mu]\). This shows that we can use HKLL to reconstruct the spacetime data arbitrarily close to \(\mu\). (Although to reconstruct points a distance \(\epsilon\) from \(\mu\), we need to wait a time of order \(\ln(\epsilon^{-1})\) for the signal to reach the boundary.) This shows that, order-by-order in small perturbation to a stationary black hole,

\[ S^{(\text{simple})}(t_i) = \frac{\text{Area}[\mu]}{4G\hbar}. \tag{13} \]

This is a new entry in the holographic dictionary, which we conjecture also holds for finite deviations from thermality.

**AN EXPLANATION FOR THE SECOND LAW**

A surface \(\mathcal{H}\) foliated by marginally trapped surfaces and satisfying certain regularity conditions obeys an area law: the area of marginally trapped surfaces increases with evolution along \(\mathcal{H}\) \([10,14]\). In the case where the marginally trapped surfaces foliating \(\mathcal{H}\) are apparent horizons, \(\mathcal{H}\) must be spacelike \([10]\), and are called trapping horizons \([10]\). Dynamical horizons \([11, 72]\), or spacelike future holographic screens \([13]\). The area law for these surfaces says that the area of slices of \(\mathcal{H}\) increases going in an outward direction.

The spacelike holographic screen \(\mathcal{H}\) is illustrated in Fig. 2(b) in a collapsing black hole, where such objects are ubiquitous. The area increases in outwards evolution along apparent horizon slices of \(\mathcal{H}\). The corresponding outer wedges are nested: evolving in the direction of increasing area corresponds to computing the outer entropy of progressively smaller outer wedges. This provides an immediate explanation for why the outer entropy increases along \(\mathcal{H}\): evolution along \(\mathcal{H}\) is the equivalent of maximizing the von Neumann entropy with progressively fewer constraints.

From a boundary perspective, the simple entropy increases for much the same reason, since as \(t_i\) is increased, there are fewer simple experiments available. It may seem odd that the simple entropy also allows measurements to be made at times after \(t_i\), but this is equivalent to saying that, for a coarse-graining scheme to have a Second Law, information cannot be discarded if it is going to become available later. (Our very late time cutoff \(t_f\), which is held constant as \(t_i\) is increased, prevents us from having to worry about recurrences.)

**Acknowledgments:** It is a pleasure to thank R. Bousso, X. Dong, G. Horowitz, J. Maldacena, D. Marolf, M. Mirbabayi, F. Pretorius, J. Santos, D. Stanford, H. Verlinde, S. Weinberg, B. White, and E. Witten for helpful discussions. The work of NE is supported in part by NSF grant PHY-1620059, while AW was supported by the Institute for

---

1 When light rays in \(N(t_i)\) intersect before reaching \(\mu\), \(O_W[\mu] \supset C_W[t_i]\) since the past boundaries do not coincide. However, \(O_W[\mu]\) still lies in the domain of dependence of \(C_W[t_i]\) allowing reconstruction of the full data. \([41]\)
APPENDIX: CONSTRAINT EQUATIONS

Since we are imposing data on $N_{-k}$, we need to use the “characteristic initial data formalism” [73][79], which guarantees the existence of a solution if we satisfy the following constraint equations on $N_{-k}$ (one for each spacetime dimension $D$):

$$\nabla_k \theta_k = -\frac{1}{D-2} \theta_k^2 - \zeta_k^2 - 8\pi G T_{kk},$$

$$\nabla_k \chi_i = 8\pi G T_{ik} - (D^\chi \cdot \zeta_k)_i,$$

$$\nabla_k \theta_\ell = -\frac{1}{2} R - 2 \nabla \cdot \chi - \theta_k \theta_k + 2 \chi^2 + 8\pi G T_{k\ell},$$

as well as the corresponding junction conditions which require $\theta_k$, $\chi_i$, and $\theta_\ell$ to be continuous. Here $\zeta_k$ is the shear tensor, which is free on $N_{-k}$; $\chi$ is the intrinsic Ricci curvature of cross-sections of $N_{-k}$; $\chi_i$ is a $D-2$ component twist 1-form gauge field that tells you how much a normal vector gets boosted when transplanted in the transverse $i$-direction; $T_{ab}$ is the stress tensor; $D^\chi = \nabla - \chi$ is the twist-covariant derivative. All quantities are defined on constant $v$-slices, where $v$ is an affine parameter defined on each geodesic of $N_{-k}$, normalized so that $\nabla_k = \nabla_v$, and $k \cdot \ell = -1$.

We can solve these constraint equations for stationary $N_{-k}$ by stipulating that $\zeta = \theta_k = T_{kk} = T_{ki} = 0$, while $R$, $\chi_i$, $T_{ik}$ are constant along $v$. The marginality condition $\theta_k[\mu] = 0$ ensures continuity of $\theta_k$ and $\nabla_k \theta_\ell$ on the junction between $N_{-k}$ and $O_W[\mu]$. The shear may be discontinuous across the junction, but that is fine [80][81].

The above conditions on the stress tensor can be satisfied by reasonable matter fields. For a minimally coupled scalar field $\phi$, take $\phi$ constant in the $k$-direction; for a Maxwell field $A_k$, impose $\nabla_k A_k = 0$ in the gauge $A_k = 0$. In the Maxwell case there is one additional constraint equation for $\nabla_k F_{kk}$ that is satisfied if the current $j_k = 0$.

Because $\mu$ is an apparent horizon, generically $\nabla_k \theta_\ell < 0$ on $N_{-k}$ and $\theta_\ell[\mu] < 0$. It follows that there exists an extremal cross-section $X$ of $N_{-k}$ with $\theta_\ell = 0$ (and $\theta_k = 0$). We can solve for the location of $X$:

$$0 = \theta_\ell[\mu] + \theta_k [v + \Box v + 2 \chi \cdot \nabla v],$$

where $v$ is a function of the transverse directions. There is a unique solution to this equation, with $v < 0$ (see [40][29]).

To complete our spacetime $M'$, we invoked CPT-conjugation across the extremal surface $X$. The junction conditions are satisfied at $X$ because $\theta_\ell = \theta_k = 0$ while $\chi_i$, $F_{ik}$, $A_i$ and $\phi$ are even under CPT; for more general matter fields, we expect that CPT-invariance ensures that this gluing is always possible.
