Post-firewall paradoxes

Samuel L. Braunstein and Stefano Pirandola

Computer Science, University of York, York YO10 5GH, United Kingdom

(Dated: November 27, 2014)

We derive two variations of the firewall paradox for evaporating black holes. Our approach sweeps away many unnecessary assumptions. There is no need: to assume ‘nice slices’ through the black hole spacetime; to decode Hawking radiation; to make any assumptions about the physics of Planck scale black holes; and other simplifications. Theorem 1: A contradiction exists between: 1.a) completely unitarily evaporating black holes, 1.b) a freely falling observer notices nothing special until they pass well within a large black hole’s horizon, and 1.c) the black hole interior Hilbert space dimensionality may be well approximated as the exponential of the Bekenstein-Hawking entropy. Theorem 2: A contradiction exists between: 2.a) completely unitarily evaporating black holes, 2.b) large black holes are described by local physics, and 2.c) externally, a large black hole should resemble its classical theoretical counterpart (aside from its slow evaporation). At least one of these conventionally accepted verities from each theorem must be false.

PACS numbers: 04.70.Dy, 03.65.Xp, 03.67.a, 03.70.+k

The fundamental physics of quantum mechanical black holes has been an enduring mystery [1] due largely to the virtual impossibility of performing experiments on idealized black holes which are isolated even from the cosmic microwave background (though there is significant progress on observations of astrophysical black holes in complicated environments [2]). For this reason, arguments about the actual behavior of idealized black holes have relied on a patchwork of analytic results from quantum field theory on curved spacetime, consistency with thermodynamics, correspondences to classical black holes, and general physical principles, to name but a few.

A paradigm for such patchwork reasoning is the information paradox for evaporating black holes which suggests that a contradiction exists between quantum mechanics and General Relativity [1]. On the one hand, quantum mechanics provides a unitary description of all physical processes which guarantees logical and physical reversibility, and on the other, General Relativity’s black holes have as their defining feature an event horizon which acts as a semi-permeable membrane: anything may enter, nothing may leave. As decades of research has shown, however, pinning down all the assumptions behind the information paradox has proven highly non-trivial [3]. For example, the information content of the matter that initially collapses to form a black hole is only a vanishingly small fraction of the entropy that eventually appears in the Hawking radiation [4], making any assumptions about the physics of black hole spacetime; to decode Hawking radiation; to make any assumptions about the physics of Planck scale black holes; and other simplifications. Theorem 1: A contradiction exists between: 1.a) completely unitarily evaporating black holes, 1.b) a freely falling observer notices nothing special until they pass well within a large black hole’s horizon, and 1.c) the black hole interior Hilbert space dimensionality may be well approximated as the exponential of the Bekenstein-Hawking entropy. Theorem 2: A contradiction exists between: 2.a) completely unitarily evaporating black holes, 2.b) large black holes are described by local physics, and 2.c) externally, a large black hole should resemble its classical theoretical counterpart (aside from its slow evaporation). At least one of these conventionally accepted verities from each theorem must be false.

Before proceeding, let us review an important tool: The quantum mutual information, \( S(X : Y) \equiv S(X) + S(Y) - S(X,Y) \), provides a measure of correlations between a pair of systems \( X \) and \( Y \). Here the von Neumann entropy, denoted \( S(X) \), gives the thermodynamic entropy for an isolated system [8], \( X \). As with some earlier works studying black hole evaporation [5][9] we shall rely on the property of strong subadditivity [10] which may be simply expressed, for any extra system \( W \), as

\[
S(X : Y) \leq S(W, X : Y).
\] (1)

We will leverage this tool to identify contributions to thermodynamic entropy in even very large quantum systems. Suppose \( X \) and \( Y \) are remotely separated and no longer interact. If the correlations between these systems correspond to maximally entangled subspaces of \( X \) and \( Y \), then they make a distinct contribution [11] of \( \frac{1}{2}S(X : Y) \) to the thermodynamic entropy in \( X \) (and in \( Y \)); an entropy that is observable by even highly delocalized detectors. This is in contrast to local correlations.
due, for example, to entropy of entanglement in the vacuum state across a boundary \[12\]. In that case, correlations are localized to a narrow layer at the boundary and unless one’s detectors are localized on scales comparable to the cutoff (presumably Planckian) such entropy is unobservable.

Non-exotic atmosphere: To reconcile gravity with quantum mechanics, it is generally assumed that there exists a correspondence between the physical characteristics of a real (i.e., quantum mechanical) black hole and its theoretical classical counterpart. The tightest correspondence assumes that black holes evaporate into vacuum (as seen by an infalling observer) \[1\]. Here, our two theorems shall make much weaker assumptions (1.b and 1.c) than that the quantum field into which the black hole evaporates is in (or near to) a specific quantum state.

In part, we achieve this by focusing on the gross thermodynamic properties of the neighborhood, \(N\), external to and surrounding a black hole that reaches out far enough to encompass any process by which radiation is produced. Now, recall ‘t Hooft’s entropic bound \[4, 13\], which shows that if one excludes configurations of ordinary matter that will inevitably undergo gravitational collapse, one finds \(S_{\text{matter}} \leq A^{3/4}\), where \(S_{\text{matter}}\) is the thermodynamic entropy of the matter (with Boltzmann’s constant set to unity) and \(A\) is the surface area of the region containing that matter (in Planck units).

Suppose that the external neighborhood, at some specific epoch, has a surface area \(\mu\) times that of the black hole’s initial event horizon, so \(A = \mu S_{\text{BH}}/4\), where throughout \(S_{\text{BH}}\) will denote the black hole’s initial Bekenstein-Hawking entropy. Then to satisfy ‘t Hooft’s bound, the thermodynamic entropy of the external neighborhood, \(S_{\text{therm}}^N\), must be bounded by

\[
S_{\text{therm}}^N \leq \left(\frac{\mu S_{\text{BH}}}{4}\right)^{3/4} \leq \left[\frac{1}{2\sqrt{2}}\left(\frac{\mu^3}{S_{\text{BH}}^4}\right)^{1/4}\right] S_{\text{BH}}. \tag{2}
\]

For large black holes, the prefactor in square brackets is much much less than unity even for extremely large neighborhoods, e.g., \(\mu = 10^4\). If this bound fails, the external neighborhood, \(N\), must consist of some exotic matter, such as an atmosphere of microscopic black holes.

Proof of Theorem 1: Assumption 1.a (and identically 2.a) has two ingredients: That black hole evaporation is a unitary process and that it is complete leaving behind no remnant. We start by unraveling the implications of these two ingredients separately:

Unitary evaporation: Consider the unitary generation of radiation from a black hole by an arbitrary process (Fig. 1). We associate this process with some specific black hole and presume that to an excellent approximation radiation is not produced further out beyond \(N\).

Applying strong subadditivity to Fig. 1 trivially yields \(S(R' : R) \leq S(B', N', R' : R)\). Our work significantly departs from prior analyses \[9\] by both allowing for arbitrary evaporation processes and for using the invariance of entropy to unitary transformations, so \(S(B, N : R) = S(B', N', R' : R)\), and hence

\[
S(R' : R) \leq S(B, N : R). \tag{3}
\]

In particular, this invariance allows us eliminate the \(B'\) and \(N'\) entirely from Eq. (3). Thus, even when evaporation leaves only a microscopic black hole behind, we can work instead with quantities \(B\) and \(N\) which may be associated with a still large black hole. Thus, Eq. (3) can be used to study even very long stretches of a black hole’s lifetime for arbitrary unitary processes greatly extending previous results \[5, 9\].

Complete evaporation: Consider a black hole created by collapsing matter initially in a pure quantum state (the Appendix considers more general scenarios). After the black hole has completely evaporated away the radiation should also be in a globally pure quantum state to preserve unitarity. Thus one might expect that correlations would exist between the early and late epoch radiation, \(R\) and \(R'\), respectively.

In fact, the study of random unitary operations allows us to say much more: Since the Hilbert space dimensionalities involved are so huge Levy’s lemma guarantees a generic behavior for entropy in all but a set of measure zero of evaporation mechanisms \[6\]. In particular \[6, 14\], the entropy of the radiation grows monotonically (at almost exactly the maximal rate of one bit’s worth of entropy per qubit of radiation emitted \[15\]) until the Page time (when the black hole’s area has halved). From the Page time onwards the entropy in the net radiation (i.e., including any radiation from earlier epochs) monotonically decreases at the same rate, reaching zero when the black hole has evaporated away \[6, 14\].

Consequently, the pre-Page time radiation and post-Page time radiation are nearly maximally entangled with each other; each carrying entropy of almost exactly \(\frac{1}{2}S_{\text{BH}}\) for an initial black hole with Bekenstein-Hawking entropy.
of $S_{BH}$. Note, that Levy’s lemma does not exclude evaporation mechanisms that yield a behavior different from the generic (e.g., as suggested in Ref. [16]), but it shows that such mechanisms will be exponentially unstable to any but a set of measure zero of perturbations.

Consider now the scenario where we follow a black hole to a relatively late stage of its complete evaporation. In particular, when its area has shrunk to some small fraction of its original size, but is still much larger than the Planck scale so that the physics of Planck scale black holes plays no part. We denote all pre-Page time radiation as $R$ and the post-Page time radiation as $R'$ (produced up until the black hole has reached a specified fraction, say roughly $\epsilon/2$, of its original area). It follows therefore from the generic behavior of entropy during evaporation [6] [14] that

$$S(R' : R) = (1 - \epsilon)S_{BH}, \quad \epsilon \ll 1. \quad (4)$$

Combining this with Eq. (3) we find

$$S(B, N : R) \geq (1 - \epsilon)S_{BH}, \quad \epsilon \ll 1. \quad (5)$$

Equation (5) tells us that the radiation is almost perfectly maximally entangled with a subspace of the joint system $(B, N)$ and as they quickly become remotely separated we may conclude that $\frac{1}{2}(1 - \epsilon)S_{BH}$ represents a lower bound to the thermodynamic entropy of this joint system. This follows straightforwardly and differs from earlier arguments [5] which invoke a likely impossible capacity for decoding Hawking radiation [17].

Free-fall equanimity: Consider now a freely-falling observer who is believed to see nothing special until they pass well within a large black hole’s horizon (assumption 1.b). For black holes formed by matter in a pure quantum state, the (global) state of $(B, N, R)$ can also be treated as pure implying $S(B, N : R) = S(B : R) + S(N : R)$. To ensure that our infalling observer is not passing through an atmosphere of exotic matter even before they reach the horizon, Eq. (2) implies that Eq. (5) reduces to

$$S(B : R) \gtrsim S_{BH}. \quad (6)$$

Now a trivial bound to the quantum mutual information is that $2\log_\epsilon(\text{dim}(B)) \geq S(B : R)$. If this bound were saturated, the huge thermodynamic entropy that Eq. (6) tells us is in $B$ would imply that an infalling observer would immediately encounter an incredibly mixed state (e.g., a near uniform mixture of roughly $10^{10^{77}}$ orthogonal quantum states for an initially stellar mass black hole) with correspondingly huge energies as soon as they passed the horizon. They would immediately encounter an ‘energetic curtain’ [6] or firewall [5] upon entering the black hole. To ensure, therefore that assumption 1.b holds, the above bound must be far from saturation, i.e.,

$$\log_\epsilon(\text{dim}(B)) \gg \frac{1}{2}S_{BH}, \quad (7)$$

where $B$ is the black hole interior at the Page time.

Finite interior Hilbert space: We may now arrive at a direct contradiction along the lines of the original firewall paradox if we add the final assumption (1.c) that the black hole interior has Hilbert space dimensionality that is well approximated by the exponential of the Bekenstein-Hawking entropy. In particular, at the Page time, when a black hole’s surface area has shrunk to one half of its original value this implies

$$\log_\epsilon(\text{dim}(B)) \approx \frac{1}{2}S_{BH}, \quad (8)$$

which directly contradicts Eq. (7).

The AdS/CFT correspondence crisis: The strongest contender for a fully unitary theory to describe black hole evaporation involves the AdS/CFT correspondence, which formalizes the holographic principle. Specifically within the context of anti-de Sitter space, this correspondence maps black holes to a hot gas in a conformal field theory without gravity where unitary behavior seems unquestionable [19]. Unfortunately, the AdS/CFT correspondence gives absolutely no inkling of the existence of a firewall, nor that anything unusual might happen from the Page time onwards.

This disconnect with the firewall paradox has precipitated such a crisis for the AdS/CFT and holographic approaches that their creators have resorted to a radical [19] (many in the community call it an absurd [20]) solution. In particular, it is proposed that black holes occur as pairs with maximal entanglement between their degrees-of-freedom [19]; it is claimed that this pre-entanglement provides a way out of the firewall paradox [19].

Theorem 1, however, appears to immediately dispose of this proposed solution: i) identify our subsystems $B, N$ and $R$ etc. with those of the joint Hilbert spaces for the paired black hole interiors, their joint neighborhoods and their joint radiation subsystems, respectively; and ii) replace axiom 1.c by “the interior Hilbert space dimensionality of paired black holes may be well approximated as the exponential of the sum of their individual Bekenstein-Hawking entropies”. Everything goes through without change as before; complexity or an ability to decode Hawking radiation plays no role; all we rely upon is the rise and fall of the entropy in the radiation as dictated by Levy’s lemma. Thus, our version of the firewall paradox is insensitive to the presence of any amount of hypothetically pre-entanglement, and the proposed pairing of entangled black holes [19] does not alleviate the clash between the firewall paradox and the AdS/CFT correspondence.

Notwithstanding the above, the firewall paradox implies that there is at least something important missing in the AdS/CFT description of individual black holes [19]. As such, we can no longer rely on it as a proposed resolution to the black hole information paradox. More generally, this in turn implies that we cannot trust the intuitions coming from the holographic principle [4, 19, 21].
in terms of a non-local representation of a black hole, nor can we trust its proposal that some non-local physics allows information to leak out of the black hole during evaporation. It is therefore worthwhile to re-evaluate the role of locality during black hole evaporation. We do this here in Theorem 2. Indeed, we shall see that any potential clash between unitarity and locality actually requires a third ingredient left out from holographic considerations.

**Proof of Theorem 2:** Let us now replace the assumptions 1.b and 1.c while retaining assumption 1.a, though with the new label 2.a.

Local physics: In addition to assuming the complete unitary evaporation of a black hole (2.a), we shall suppose that whatever process generates the radiation it is constrained to be local for large black holes (assumption 2.b). In particular, we shall focus on the fact that local physics forbids communication across light cones \[22\], so that there can be no communication from within a large black hole’s event horizon to the exterior.

In order to make use of this non-communication property we recall the no-communication decomposition theorem \[23\] (see Fig. 2) which tells us that any unitary process \(U\) which does not allow communication from a set of inputs to a set of outputs may be decomposed into a pair of unitary subprocesses \(V\) and \(W\) with at most some backwards communication within a subsystem \(C\). (The dots denote ancillary degrees-of-freedom.)

This theorem requires that the inputs and outputs form distinct components of a tensor product decomposition of the overall Hilbert space; a requirement which is automatic for finite dimensions. For any local quantum field theory we may rely on the fact that operators with support only outside each others light cones must commute. Thus, locality dictates the existence of the required tensor product structure across a black hole’s event horizon (defined as a null, light cone, surface) \[11, 19\]. We may therefore apply the circuit equivalence in Fig. 2 to the black hole evaporation of Fig. 1 to give the structure of an arbitrary unitary black hole evaporation process consistent with local physics (see Fig. 3).

Note that Fig. 3 is not to be interpreted as a space-time diagram. In particular, we do not require that there is any space-like hypersurface which simultaneously cuts through the subsystems there displayed. For example, we do not require that subsystem \(C\) all arrives in one block for unitary processing inside the black hole. From this perspective, a quantum circuit is a very powerful construct (see the Appendix for a generalization that includes arbitrary infallen matter).

Applying strong subadditivity to Fig. 3 yields a new relation \(S(R' : R) \leq S(C, N', R' : R)\). Again, using the unitary invariance of entropy we have \(S(C, N', R' : R) = S(N : R)\) leading to

\[
S(R' : R) \leq S(N : R). \tag{9}
\]

Note that this inequality involves only correlations between external degrees-of-freedom and hence relates quantities which are, in principle, directly observable and reportable. Combining this with the assumption of complete evaporation, Eq. (1), we easily find

\[
S(N : R) \geq (1 - \varepsilon)S_{BH}, \quad \varepsilon \ll 1. \tag{10}
\]

Locality (assumption 2.b) has allowed us to eliminate \(B\) from Eq. (6), which in turn allows us to do without any specific bound to the size of the interior Hilbert space. More surprisingly, locality implies a very different picture: one where huge thermodynamic entropies must reside outside the black hole instead of inside it.

At first sight, this appears reminiscent of arguments based on time-reversing Hawking radiation. Ordinary Hawking radiation evolves out of vacuum modes, but any (information bearing) deviations were argued to have started out as high-energy excitations near the horizon \[24\]. By contrast, the huge entropies in Eq. (10) are associated with degrees-of-freedom that are maximally entangled with the outgoing radiation and therefore correspond to an effect of the “infalling partners” to the radiation. Thus, Eq. (10) represents a distinct (and much stronger) phenomenon imposed by locality.
**Non-exotic atmosphere:** Assumption 1.c is much weaker than 1.b, only requiring that externally, black holes should resemble their classical counterparts (aside from their slow evaporation). In turn, we apply this in a very weak manner to only suppose that the black hole does not contain an atmosphere of super-entropic exotic matter. Therefore, from Eq. (2), we obtain

$$S(N : R) \leq \eta S_{\text{BH}}, \quad \text{with} \quad \eta \ll 1. \quad (11)$$

Combining Eqs. (10) and (11) yields the contradiction

$$1 \leq \varepsilon + \eta \ll 1, \quad (12)$$

whatever the details of the radiation process.

**Discussion:** Let us just recall that the consequences for both theorems “kick in” at the Page time, therefore, all our discussion below will be about the behavior of large black holes. Theorem 1 ignores the local structure of a black hole’s horizon and suggests that huge thermodynamic entropies reside within the black hole. Theorem 2 incorporates this structure and instead suggests that this entropy resides external to the event horizon.

The cheapest resolution, cut along the lines of Occam’s razor, would be to reject assumption 1.a (2.a). However, for unitarity to be preserved, black hole evaporation cannot stop when some ‘stable remnant’ is reached [1,25,26], nor can the black hole interior ‘bud off’ as a baby universe [4,27]. Any such loss of unitarity would infect almost every other quantum mechanical process [4,25]. Unfortunately, as already noted, the firewall paradox has precipitated a crisis for the AdS/CFT correspondence at least for individual black holes [19]. So this route to ensuring unitarity during black hole evaporation can no longer be relied upon.

If we accept 1.a (2.a), either theorem leaves us with a striking dichotomy. Let us start with the consequences of Theorem 1. We must reject at least one assumption of 1.c or 1.b. Although we do not yet have a proper theory of quantum gravity, whatever form such a theory will eventually take it should still be possible to argue that the Hamiltonian constraint of describing an initially compact object with finite mass must be effectively limited to a finite-dimensional Hilbert space. If the dimensionality of this effective space is not well described by assumption 1.c it is hard to see how the theory of black hole thermodynamics could survive.

By contrast, assumption 1.b, although usually considered a consequence of the Equivalence Principle of General Relativity is in fact nothing more than a boundary condition on the quantum fields at the event horizon; it is well known that different choices of ‘vacuum state’ lead to wildly different behaviors for the energy-momentum tensor. From this perspective, the least shocking result may well be to accept that there is something like a firewall (or energetic curtain) no deeper than the horizon. Indeed, weakening assumption 1.b to “a freely falling observer notices nothing special until they reach a large black hole’s horizon” removes the contradiction found in Theorem 1 and implies the firewall phenomenon is real.

Finally, consider the options left by Theorem 2: to reject at least one of the assumptions 2.c or 2.b. Now the failure of 2.c would imply that the exterior must consist of super-entropic exotic matter (such as an atmosphere of microscopic black holes), and so would almost certainly have some observational consequences [29]. The only other minimal option (rejecting assumption 2.b) would be to assume that communication from the black hole interior to exterior across the horizon was possible. In particular, one might note that a “tunneling” mechanism has been long anticipated to provide a more powerful explanation for black hole radiation [6,30,31]. Unfortunately, tunneling across the horizon alone [30] is insufficient, as it still leaves unanswered how the degrees-of-freedom carrying information from deep within a black hole manage to (non-locally) reach up to just inside the horizon where they can participate in such tunneling.

We end with a purely speculative description of how black holes might evolve during evaporation, as suggested by the discussion above. During the initial stages of evaporation, prior to the Page time, entanglement grows between the interior and distant radiation [14]. Now entanglement cannot be compressed into fewer qubits than given by its entropy (a principle which in no way assumes that the associated matter has been compressed to Planck densities). Therefore, accepting an effectively finite size Hilbert space to the black hole interior, this slowly fills up with incompressible entanglement. If something like the conventional spacetime structure nonetheless exists within the black hole this incompressible ‘stuff’ would grow outwards from what would classically be the black hole singularity, while simultaneously, the black hole’s horizon is shrinking as radiation is emitted.

We might conjecture therefore that there is some well-defined internal entanglement surface, or bubble, that contains the entanglement growing outward. At the Page time, the entanglement surface and black hole horizon meet. At that stage the horizon may survive, or may be replaced by the entanglement surface. In the former case, evaporation would continue by something very much like quantum tunneling [6,30] from degrees-of-freedom on the entanglement surface just inside the horizon. In the latter case, evaporation may continue via direct thermal ejection from the entanglar (entangled star) though its detailed spectrum and lifetime may well differ from a true black hole with otherwise identical mass, charge and angular momentum. Naively, an entanglar (of even a modest size) would take far longer than the age of the universe to evolve from a black hole, so none can be expected to currently exist. Conversely, the unambiguous observation of such an entanglar would yield prima facie evidence for an object that far predates the Big Bang.
Appendix

Generics of black hole radiation

In the manuscript we considered a black hole with thermodynamic entropy $S_{\text{BH}}$ which can completely evaporate into a net pure state of radiation. As discussed, the generic evaporative dynamics of such a black hole may be captured by Levy’s lemma for the random sampling of subsystems from an initially pure state consisting of $S_{\text{BH}}$ qunats [6]. This either assumes the infallen matter is pure (as in the manuscript) or ignores it entirely. Throughout, we set Boltzmann’s constant to unity and work with natural logarithms leading to the measure of qunats (i.e., $\log_2$ times the number of qubits [15]).

In order to extend our analysis to include infallen matter carrying some (von Neumann) entropy $S_{\text{matter}}$, we need only take the initially pure state used above and replace it with a bipartite pure state consisting of two subsystems: $S_{\text{BH}}$ qunats to represent the degrees-of-freedom that evaporate away as radiation; and a reference subsystem. Without loss of generality, the matter’s entropy may be treated as entanglement between these two subsystems, however, here we shall simplify our analysis by assuming uniform entanglement between the black hole subsystem and $S_{\text{matter}}$ reference qunats. The generic properties of the radiation may then again be studied by random sampling the former subsystem to simulate the production of radiation [3].

The behavior is generic and for our purposes may be summarized in terms of the radiation’s von Neumann entropy, $S(R)$, as a function of the number of qunats in this radiation subsystem. One finds [6] that $S(R)$ initially increases by one qumat for every extra qumat in $R$, until it contains $\frac{1}{2}(S_{\text{BH}} + S_{\text{matter}})$ qunats. From that stage on it decreases by one qumat for every extra qumat in $R$ until it drops to $S_{\text{matter}}$ when $R$ contains $S_{\text{BH}}$ qunats and the black hole has completely evaporated.

Because the von Neumann entropy of a randomly selected subsystem only depends on the size of that subsystem, the same behavior is found whether $R$ above represents the early or late epoch radiation with respect to any arbitrary split. Further, in the simplest case where we choose the joint radiation $(R, R')$ to correspond to the net radiation from a completely evaporated black hole we may immediately write down the generic behavior for the quantum mutual information $S(R' : R)$.

In particular, $S(R' : R)$ starts from zero when $R$ consists of zero qunats. From then on, it increases by two qunats for every extra qumat in $R$ until $S(R' : R)$ reaches $S_{\text{BH}} - S_{\text{matter}}$ when $R$ contains $\frac{1}{2}(S_{\text{BH}} - S_{\text{matter}})$ qunats. From that stage on until $R$ contains $\frac{1}{2}(S_{\text{BH}} + S_{\text{matter}})$ qunats $S(R' : R)$ remains constant, after which $S(R' : R)$ decreases by two qunats for every extra qumat in $R$ until it drops to zero once the $R$ contains the full $S_{\text{BH}}$ qunats of the completely evaporated black hole [6]. Interestingly, it is during the region where $S(R' : R)$ is constant that the information about the infallen matter becomes encoded into $R$ for the first time [6]. Finally, setting $S_{\text{matter}}$ to zero gives the ‘standard’ behavior for $S(R)$ and $S(R' : R)$ upon which the results in the manuscript are derived.

From the above, we are motivated to generalize the Page time: we define any time where $S(R' : R)$ is maximal a (generalized) Page time; the earliest such time the ‘initial Page time’; and the latest the ‘final Page time’. Prior to the initial Page time, the quantum information about the initial infalled matter is encoded entirely within the black hole interior [6]. After the final Page time this information is encoded entirely within the radiation [6].

Including infallen matter

Let us start with a consideration of how the reasoning in Theorem 2 becomes modified by the presence of infallen matter carrying entropy.

Theorem 2 generalized: In the main body of the manuscript we did not explicitly include entropy associated with infallen matter. Fig. 4 shows the most general scenario. Subsystem $I$ denotes the matter that falls into the region surrounding the black hole where radiation is produced. Thus, we suppose that late epoch radiation can in principle come from the joint subsystem $(N, I)$. In this figure we also include subsystem $I_{\text{early}}$ denoting matter that has fallen into the region surrounding the black hole at an earlier epoch or indeed matter that may have collapsed to form the black hole in the first place.

As in the manuscript we apply strong subadditivity:

$$S(R' : R) \leq S(C, N', R' : R)$$

$$= S(N, I : R) = S(N : R).$$

Here, we used the fact that joint subsystems $(C, N', R')$ and $(N, I)$ are unitarily related. Finally, the most natural assumption is that the infallen matter $I$ is independent of the quantum state of the black hole, $(B, N)$, or its early
epoch radiation $R$. The original inequality of Eq. (9) from the manuscript is thus found to still hold in the presence of infallen matter.

From the summary above of generic radiation production including infallen matter we have enough to generalize Theorem 2. As in the manuscript, we take $R$ to be all the early epoch radiation until the Page time (for this theorem we may take any generalized Page time), and we let $R'$ denote all the radiation generated from the Page time onwards until the black hole has shrunk to a size much smaller than the original black hole (say roughly $\varepsilon/2$ of its original area), but still much larger than the Planck scale. In this case, instead of Eq. (4) from the manuscript, we have

$$S(R' : R) = (1 - \varepsilon) S_{BH} - S_{\text{matter}} \
(\varepsilon \ll 1), \quad (14)$$

where $S_{\text{matter}} = S(I_{\text{early}}, I)$ is the net entropy in all the infallen matter. Combining this with Eq. (9) gives

$$1 - \frac{S_{\text{matter}}}{S_{BH}} \leq \varepsilon + \eta \ll 1. \quad (15)$$

Once again we obtain a contradiction except in the extreme case of a black hole whose net infallen matter contains virtually as much entropy as the entire black hole’s original entropy $S_{BH}$.

**Theorem 1 generalized:** It is simple enough to repeat the above reasoning for Theorem 1, where we no longer make use of locality. In this case, we may still use Fig. 4 provided we ignore the no-communication decomposition structure. In particular, strong subadditivity yields

$$S(R' : R) \leq S(B', N', R' : R) = S(B, N : R), \quad (16)$$

which is identical to Eq. (3) of the manuscript. Here, we use the fact that joint subsystems $(B', N', R')$ and $(B, N, I)$ are unitarily related. Again, the most natural assumption is that the infallen matter $I$ is independent of the quantum state of the black hole, $(B, N)$, or its early epoch radiation $R$.

Applying Eq. (14) to any Page time then tells us that for a unitarily and completely evaporating black hole

$$S(B, N : R) \geq (1 - \varepsilon) S_{BH} - S_{\text{matter}}, \quad (\varepsilon \ll 1). \quad (17)$$

To simplify our argument, we shall suppose that the infallen matter $(I_{\text{early}}, I)$ has actually entered the black hole. In that case, for any times prior to the initial Page time, the infallen matter’s external reference qunats are maximally entangled with some subsystem of the black hole interior $[8]$. We shall label the orthogonal complement of this subsystem within $B$ as $B^\perp$. It is clear that: i) $(B^\perp, N, R)$ can be treated as a pure quantum state; and ii) $S(B, N : R) = S(B^\perp, N : R)$. So that

$$S(B^\perp : R) + S(N : R) \geq (1 - \varepsilon) S_{BH} - S_{\text{matter}}, \quad (\varepsilon \ll 1). \quad (18)$$

To ensure that our infalling observer is not passing through an atmosphere of exotic matter before they reach the horizon, Eq. (2) from the manuscript for a large black hole implies that Eq. (18) reduces to

$$S(B^\perp : R) \gtrsim S_{BH} - S_{\text{matter}}. \quad (19)$$

Since $\log_e(\dim(B^\perp)) = \log_e(\dim(B)) - S_{\text{matter}}$ by construction, we find the trivial bound

$$\log_e(\dim(B)) \gtrsim \frac{1}{2} (S_{BH} + S_{\text{matter}}). \quad (20)$$

If this bound were saturated, then the huge thermodynamic entropy in $B$ would imply that an infalling observer would immediately encounter an incredibly mixed state with correspondingly huge energies as soon as they passed the horizon. The would immediately encounter an ‘energetic curtain’ or firewall upon entering the black hole. To ensure, therefore that assumption 1.b holds, the above bound must be far from saturation, i.e.,

$$\log_e(\dim(B)) \gg \frac{1}{2} (S_{BH} + S_{\text{matter}}). \quad (21)$$

where $B$ is the black hole at the *initial* Page time.

However, assumption 1.c would require that the left- and right-hand-sides of Eq. (21) should be nearly equal. As with the generalization of theorem 2, we again obtain a contradiction, in this case, however, apparently independent of the amount infallen matter.

* sam.braunstein@york.ac.uk

http://www.cs.york.ac.uk/~schmuel

[1] S. W. Hawking, Phys. Rev. D 14, 2460 (1976).
[2] M. Zamaninasab, et al., Nature 510, 126(2014).
[3] V. Balasubramanian and B. Czech, Class. Quant. Grav. 28, 163001 (2011).
[4] G. ’t Hooft, in *Salamfestschrift: A Collection of Talks*, edited by A. Ali, J. Ellis, and S. Randjbar-Daemi (World Scientific, Singapore, 1993), Vol. 4, p. 284.
[5] A. Almheiri, et al., J. High Energy Phys. 02 (2013) 062.
[6] S. L. Braunstein, [arXiv:0907.1190v1 (2009); published as S. L. Braunstein, S. Pirandola and K. Zyczkowski, Phys. Rev. Lett. 110, 101301 (2013);]
[7] D. A. Lowe, et al., Phys. Rev. D 52, 6997 (1995).
[8] A. Wehrl, Rev. Mod. Phys. 50, 221 (1978).
[9] S. D. Mathur, Classical Quantum Gravity 26, 224001 (2009).
[10] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, 2013) p. 270.
[11] This subspace of $X$ must have Schmidt coefficients consistent with the entanglement and hence carrying entropy $\frac{1}{2} S(X : Y)$. By mimicking this subspace with a locally created probabilistic state given by $\text{tr}_Y(p_{X | Y})$ every local property, including thermodynamic, can be seen to be identical in the original and mimicked states.
[12] J. Eisert, M. Cramer and M. B. Plenio, Rev. Mod. Phys. 82, 277 (2010).
The terms “bits” and “qubits” are used as units for information content and do not imply a set of two-level systems.

R. Brustein and A. J. M. Medved, Phys. Rev. D 90, 024040 (2014).

J. M. Maldacena and L. Susskind, arXiv:1306.0533

“I was repeatedly asked why I believe something as absurd as ER=EPR [imposing pre-entanglement between black hole pairs]. My answer was: “It was proved by AMPS [the firewall paradox].””, L. Susskind in arXiv:1411.0690v1.

L. Susskind, J. Math. Phys. 36, 6377 (1995).