PHYSICAL REVIEW D **87,** 104018 (2013)

Black holes without firewalls

Klaus Larjo[*] and David A. Lowe[†]
Department of Physics, Brown University, Box 1843, Providence, Rhode Island 02912, USA

Larus Thorlacius[‡]
Nordita, KTH Royal Institute of Technology and Stockholm University, Roslagstullsbacken 23, SE-106 91 Stockholm, Sweden and University of Iceland, Science Institute, Dunhaga 3, IS-107 Reykjavik, Iceland
(Received 4 December 2012; published 16 May 2013)

The postulates of black hole complementarity do not imply a firewall for infalling observers at a black hole horizon. The dynamics of the stretched horizon, that scrambles and reemits information, determines whether infalling observers experience anything out of the ordinary when entering a large black hole. In particular, there is no firewall if the stretched horizon degrees of freedom retain information for a time of the order of the black hole scrambling time.

DOI: 10.1103/PhysRevD.87.104018 PACS numbers: 04.70.Dy, 04.62.+v

I. INTRODUCTION: COMPLEMENTARITY OR FIREWALL?

Black hole complementarity was introduced in Ref. [1] in terms of three postulates for black hole evolution:

(1) The process of formation and evaporation of a black hole, as viewed by a distant observer, can be described entirely within the context of standard quantum theory. In particular, there exists a unitary S-matrix which describes the evolution from infalling matter to outgoing Hawking-like radiation.
(2) Outside the stretched horizon of a massive black hole, physics can be described to good approximation by a set of semiclassical field equations.
(3) To a distant observer, a black hole appears to be a quantum system with discrete energy levels. The dimension of the subspace of states describing a black hole of mass M is the exponential of the Bekenstein entropy $S(M)$.

These postulates refer to observations made outside the black hole and provide a basis for a phenomenological description that is consistent with unitarity. A further key assumption, based on the equivalence principle, was made in Ref. [1] and can be expressed as a fourth postulate that applies to observers who enter the black hole:

(4) An observer in free fall experiences nothing out of the ordinary upon crossing the horizon of a large black hole.

The combination of this assumption and the three original postulates requires one to give up the notion of spacetime locality. In particular, the fate of observers entering a large black hole is very different depending on the frame of reference: In their own rest frame, they pass unharmed through the horizon and only come to harm as they approach the curvature singularity, while from the viewpoint of distant observers, they never pass through the horizon at all but are instead absorbed into the stretched horizon and thermalized before being reemitted along with the rest of the black hole in the form of Hawking radiation. It was argued in Ref. [2] that no low-energy observer can detect violations of known laws of physics even if information carried by infalling matter appears to be duplicated in the outgoing Hawking radiation.

The stretched horizon is a surface outside the global black hole horizon that remains timelike. Outside observers ascribe nontrivial microphysical dynamics to the stretched horizon that serves to absorb, thermalize, and eventually reemit the information contained in infalling matter. The usual thermodynamics of black holes is assumed to arise from a coarse graining of this (unspecified) microscopic dynamics. From the point of view of outside observers, no information ever enters the black hole in this description, and the stretched horizon is the end of the road for all infalling matter. In that sense it is indeed a firewall. According to the fourth postulate, the story is very different for an infalling observer. The spacetime curvature is weak at the horizon of a large black hole, and an infalling observer should not notice anything out of the ordinary upon crossing the horizon. In a recent paper, Almheiri *et al.* [3] claim, however, that the first two postulates imply that an infalling observer must also see a firewall—in other words, that the fourth postulate is inconsistent with the others.

The microscopic stretched horizon in Ref. [1] was placed at a proper distance of the order of the Planck length away from the global horizon. More generally in the present work, we require the stretched horizon to be placed at some large fixed redshift from asymptotic infinity [4]. According to the second postulate, physics outside the stretched horizon is described by semiclassical field equations of some low-energy local effective field theory. The definition of a low-energy theory includes specifying a cutoff. In the black hole context, this means that the spatial slices, on which the effective theory is defined, terminate at

[*]klaus_larjo@brown.edu
[†]lowe@brown.edu
[‡]larus@nordita.org

an *effective* stretched horizon located outside the microscopic stretched horizon. For concreteness, let us consider a quasistatic spherically symmetric black hole as shown in Fig. 1. The effective stretched horizon can then be taken as the surface Σ, where fiducial observers (who remain at rest with respect to the black hole) would measure a local temperature equal to a cutoff scale. Equivalently, Σ is the surface such that a radially outgoing massless particle is redshifted from the cutoff energy at Σ to the characteristic energy of Hawking radiation at infinity. In this approach, anything that is inside the effective stretched horizon is represented by degrees of freedom living on the effective stretched horizon. This includes the entire black hole region and the region between Σ and the global horizon, as indicated in Fig. 1. If the cutoff energy is taken very high, close to the Planck energy, then Σ approaches the microscopic stretched horizon of Ref. [1]. For lower values of the cutoff, the dynamics on the effective stretched horizon is, in principle, obtained from the dynamics on the underlying microscopic horizon by renormalization.

The argument of Ref. [3] proceeds as follows: At very late times, we have by supposition a pure state consisting only of outgoing Hawking radiation. Since we know the laws of physics up to the stretched horizon, we can evolve this state back mode by mode from late times to the stretched horizon. The authors of Ref. [3] then seem to introduce the hidden assumption that even beyond the surface Σ, we can still evolve the mode back all the way to the global horizon, ignoring the stretched horizon degrees of freedom. At that point the argument can be reduced to that of a single mode, since by locality such a mode very close to the global horizon does not have time to entangle with the stretched horizon, and thus cannot entangle with other outgoing Hawking modes emitted at later times. Thus, once one has evolved this single late-time mode back to a point very close to the global horizon, unitarity and local quantum field theory prevent any entanglement between the outgoing mode and the black hole state. This is sufficient to guarantee that an infalling observer will see high-energy modes. In fact, the states obtained in this manner are finite excitations of the so-called Boulware vacuum, which is well known to have a divergent stress-energy tensor on the horizon [5]. Vacuum states regular on the horizon include the Hartle-Hawking vacuum or Unruh vacuum [5], which requires entanglement between outgoing Hawking modes and negative energy modes inside the black hole.

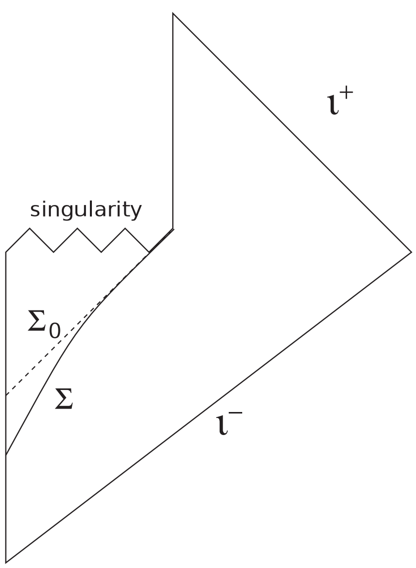

FIG. 1. Penrose diagram for black hole evaporation. Σ_0 is the global horizon, and Σ is a stretched horizon.

The firewall argument of Ref. [3] is flawed because the stretched horizon has been dispensed with. In the effective field theory description of postulate 2, Hawking radiation is emitted from the stretched horizon, which is a boundary of the spacetime. The fact that at late times the state of the stretched horizon is maximally entangled with the early Hawking radiation is no more of a problem than the corresponding statement about the remaining embers of a burning lump of coal that started out in a pure state. The firewall problem only arises if one attempts to extend the semiclassical description to the region inside the stretched horizon. If there is no entanglement between the outgoing modes and the state of the black hole, as is argued in Ref. [3], then the state of the field is given by an excitation of the Boulware vacuum, which has a stress-energy tensor that is power-law divergent as a function of proper distance as the global horizon is approached [5]. In that case infalling observers encounter drama already outside the stretched horizon, in violation of black hole complementarity. We will give a counterexample below, where the semiclassical description outside the stretched horizon is compatible with unitarity and locality and the expectation value of the stress-energy tensor remains finite in that region. This is achieved by making a different assumption about the semiclassical state of the system. Our semiclassical construction involves a firewall but only inside the stretched horizon, where the effective field theory of postulate 2 no longer applies. The presence of a firewall, both in our example and in Ref. [3], is at odds with postulate 4, but this is hardly surprising. Black hole complementarity was after all put forward to address problems arising from applying semiclassical theory in a region extending inside the stretched horizon.

Observations made by infalling observers are only well described in the local effective theory of postulate 2 as long as they remain outside the surface Σ in Fig. 1 and not after they pass through it. An alternative description should be possible, involving an effective quantum field theory on time slices where an infalling observer has low energy in the local frame of the slice [6,7]. However, to describe an infalling observer crossing the global horizon of the black hole requires time slices that extend past the location of the stretched horizon in the original effective field theory, making it difficult to map states and observables from one low-energy theory to the other. Moreover, it has been argued that, due to the large relative boosts involved, the effective low-energy description on time slices, such that both an infalling observer who has entered the black hole and the outgoing Hawking radiation are at low energy, cannot be a local field theory [6–9].

With some new assumptions about the stretched horizon dynamics, and taking care with the application of the semiclassical approach, we will argue in the following section that information may be recovered without introducing a firewall for infalling observers. Various alternatives or modifications of the firewall scenario have appeared in Refs. [10–17], but for the most part, these are also alternatives to black hole complementarity. The argument in the present paper, on the other hand, is consistent with black hole complementarity, as formulated in Ref. [1], with additional assumptions about the dynamics of the stretched horizon.

II. EMERGENCE OF INFORMATION

Let us begin by revisiting the setup of Hayden and Preskill [18], making explicit some of the relevant time scales. Alice throws her diary into an old black hole, where more than half the entropy has been emitted in Hawking radiation, and Bob measures the outgoing Hawking quanta, having first faithfully recorded all the quanta previously emitted by the black hole. That Bob can manipulate the state of the Hawking radiation in a relatively short time can be argued as follows. On average, a black hole of mass M emits a Hawking particle every M units of time, with an average energy of $\frac{1}{M}$. Hence, the total number of emitted quanta during the lifetime of a black hole will be M^2, and the total lifetime will be M^3. This timescale can be thought of as M^2 "boxes" of length M, and distributing the emission times of the M^2 Hawking particles into these boxes gives Bob access to roughly

$$N \sim (M^2)^{M^2}$$

different states, more than enough to differentiate between the e^{M^2} microstates making up the black hole.

As discussed in Refs. [18,19], the scrambling time of a black hole is given by

$$t_{\text{scramble}} \sim M \log M.$$

In this time an average of $\log M$ Hawking particles will be emitted, leading to a possible problem: since these Hawking particles can be entangled with the diary before the black hole scrambles, there may be a modification of the high-frequency quanta, and hence an infalling observer may see a firewall. On the other hand, as argued in Ref. [18], if the entanglement only appears after t_{scramble}, one finds compatibility with the postulates of black hole complementarity.

One can use information theory arguments to place a lower bound on the time scale of information retrieval. This is computed in Ref. [18] in Eq. (1). They find the probability for failure to decode a k-bit message in the diary satisfies

$$P_{\text{fail}} \leq 2^k 2^{-s}, \tag{1}$$

where s is the number of bits that Bob reads after the diary is thrown in. Now this seems to imply the information comes out faster than the scrambling time if $k \approx 1$, from which one might infer a firewall. There is, however, an error in this train of logic, since the authors of Ref. [18] assume scrambling has already happened when they make the estimate in Eq. (1). Prior to scrambling the emission rate of quantum information is not governed by Eq. (1) but rather depends on details of the stretched horizon dynamics [20].

Let us try to model these effects in more detail to determine their implications for the stretched horizon theory. Consider an old black hole prior to the diary being thrown in. It is fully entangled with the train of Hawking radiation that has already been emitted, and its state can be written as

$$|\Psi\rangle_{BBh} = \sum_i c_i |i\rangle_B \otimes |i\rangle_{Bh},$$

with $i \in [1, N]$ indexing the basis states of the black hole and $|i\rangle_B$ being the corresponding string of Hawking radiation recorded by Bob. Tracing over the Hawking radiation, the black hole density matrix is diagonal, with uniform entries due to the maximal entanglement,

$$\rho_B = \sum_i |c_i|^2 |i\rangle_{Bh}\langle i|_{Bh} = \frac{1}{N}\mathbb{I}_N, \quad \text{with} \quad N = \exp(M^2).$$

Into this state Alice throws her diary, which we initially take to consist of k bits in a pure state. Without loss of generality, we can take the state to be $|+_1, \ldots, +_k\rangle_A \equiv |(+)_k\rangle_A$. Immediately after the diary is inside the black hole, but not yet scrambled, the state and the black hole density matrix are given by

$$|\Psi\rangle = \sum_i c_i |i\rangle_B \otimes |i, (+)_k\rangle_{BhA},$$
$$\rho_{BhA} = \frac{1}{N}\begin{pmatrix} \mathbb{I}_N & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{2}$$

where the density matrix ρ is a $2^k N \times 2^k N$ matrix, written in term of $N \times N$ blocks above. The vanishing blocks of ρ_{BhA} correspond to states involving $|-\rangle_A$, over which Alice's diary does not have support yet. This is to be compared with the maximally entangled $2^k N \times 2^k N$ matrix

$$\rho_{\max} = \frac{1}{2^k N}\begin{pmatrix} \mathbb{I}_N & 0 & \cdots & 0 \\ 0 & \mathbb{I}_N & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \mathbb{I}_N \end{pmatrix}. \tag{3}$$

In modeling the stretched horizon dynamics, we assume that there is 1 degree of freedom per unit Planck area, consistent with the third postulate. The transverse wavelength of modes emitted from the stretched horizon then ranges from order M, for the low-angular momentum modes that make up the bulk of the Hawking radiation that reaches distant observers, to order 1 in Planck units, for high-angular momentum modes that never emerge far from the black hole and are reabsorbed by the stretched horizon.

A simple model for a long transverse wavelength, a Hawking particle emitted from the stretched horizon is an operator close to the identity operator in the $2^k N \times 2^k N$-dimensional Hilbert space, acting on all the stretched horizon states with approximately equal weight,

$$\mathcal{O}_{\text{long}} = \mathbb{I}_{2^k N} + \epsilon,$$

where ϵ is some small perturbation with vanishing trace. We see in each case (2) and (3) that

$$\mathrm{Tr}\rho \times \mathcal{O}_{\text{long}} = 1, \tag{4}$$

so these operators do a good job of making the emitted radiation look thermal, regardless of the stretched horizon state.

On the other hand, a model for the emission of a short transverse wavelength mode would be to pick an operator such as

$$\mathcal{O}_{\text{short}} = \begin{pmatrix} \mathbb{I}_N & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{5}$$

In this case, immediately after the diary was thrown in, we would find

$$\mathrm{Tr}\,\rho_{BhA} \times \mathcal{O}_{\text{short}} = 1 \qquad \mathrm{Tr}\rho_{\max} \times \mathcal{O}_{\text{short}} = 1/2^k,$$

so we see the answer is highly sensitive to the state of the stretched horizon. The danger is that operators of the form $\mathcal{O}_{\text{short}}$ will lead to strong modification of the high-frequency near-horizon Hawking modes, giving rise to a firewall. For this to happen, the short transverse wavelength modes coupling to such operators would have to themselves scramble and become entangled with the early Hawking radiation on a time scale that is short compared to the black hole scrambling time t_{scramble}. If, however, the stretched horizon dynamics is causal, then short transverse wavelength modes are unable to scramble (i.e., achieve approximate global thermalization with respect to the measure described in Ref. [18]) in a time $M \log M$. The fastest a localized signal can causally traverse the stretched horizon is in time M using time measured at the stretched horizon. This then redshifts to M^2 when measured using Schwarzschild time. We therefore introduce another new assumption about the dynamics of the stretched horizon theory—that it be local and causal. Without this assumption, the stretched horizon dynamics can, in principle, contaminate the causal physics outside, violating postulate 2.

The bounds on information retrieval time placed in Ref. [18] are lower bounds. The diary will scramble most efficiently if it is coded into modes with transverse wavelength of order M, as exemplified by the operator (4). Because of the long transverse wavelength, such modes couple globally to the stretched horizon degrees of freedom, and there is no causal bound preventing an $M \log M$ scrambling time. If, on the other hand, the information in the diary is present in short transverse wavelength modes, such as Eq. (5), then it may be emitted more slowly, on a timescale of order M^2 or longer.

It is, however, necessary to assume there is a genuine information retention time during which no "prompt" information is emitted from the stretched horizon while these long transverse wavelength modes scramble. This distinguishes the dynamics of the stretched horizon from, for instance, an accelerating mirror which would indeed look like a firewall from the point of view of a freely falling observer. The dynamics of the stretched horizon must be such that the reflection coefficient vanishes; the information is retained for a time t_{scramble}, at which point it is then primarily emitted in long transverse wavelength modes.

It is important in the argument of Ref. [18] that the diary be much smaller than the black hole. This can also be seen from the following estimate of the maximum number of degrees of freedom that can scramble fast enough. We ask that a causal signal from a cell of size $\lambda_{\min}$ on the stretched horizon overlaps with a neighboring cell after t_{scramble} and assume that this is sufficient for scrambling, with respect to the measure of Ref. [18], to take place. This is rather strong assumption about the efficiency of the scrambling dynamics so the resulting number of fast scramblers is likely to be an overestimate. The above condition implies $\lambda_{\min} \sim \log M$, in which case the number of independent fast scrambling degrees of freedom is of order $(M/\log M)^2$. Sending in a larger diary than this will compromise the rapid rate of information retrieval, as more generic short transverse wavelength modes scramble on a slower timescale of order M^2.

There is a finite time delay during which information scrambles on the stretched horizon after the infalling diary is absorbed. An early infalling observer (see Fig. 2) sees no substantial difference from the Unruh or Hartle-Hawking vacua in this time interval upon crossing the global horizon. However, an infalling observer crossing the outgoing

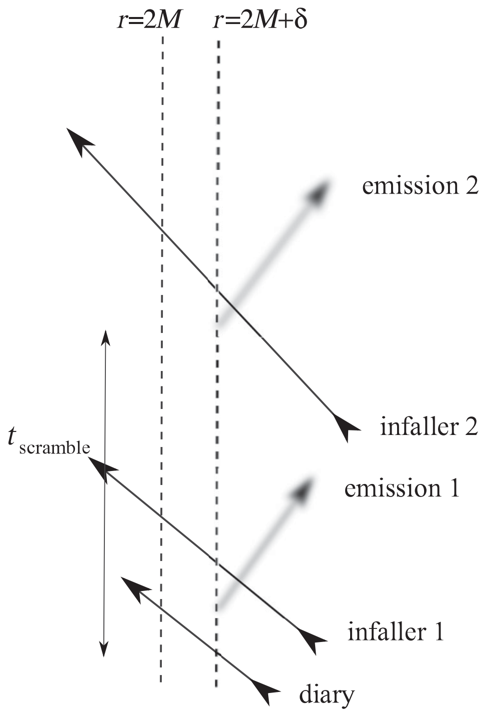

FIG. 2. Different infalling observers encountering outgoing Hawking modes. The stretched horizon is shown as the right dashed line and the global horizon as the left dashed line. Before the infalling diary scrambles on the stretched horizon, the outgoing mode is unentangled with it. Only after scrambling will an infalling observer notice entanglement with the diary. Proper time along the stretched horizon provides a distinguished set of clocks which demarcate this interval.

mode after this time interval sees a mode that has had time to spread a distance at least of order M from the stretched horizon. This mode is now entangled with the diary [21].

Next let us consider the argument of Ref. [3], which essentially replaces the infalling diary by a set of vacuum Hawking modes. The difference is illustrated in Fig. 3. An observer outside the stretched horizon must see entanglement of the outgoing mode with the early Hawking radiation according to Ref. [18]. At the same time, an observer crossing the global horizon sees the mode entangled with interior modes, not with the early Hawking radiation. Because the time scale separating infaller 1 and infaller 2 can be much shorter than the scrambling time, this creates an apparent paradox. In the work of Ref. [3], it is argued that the radiation on the outside must be some purely outgoing mode at infinity. As noted above, this can be viewed as a finite excitation of the Boulware vacuum. The Boulware vacuum has a continuous divergence outside the global horizon. A freely falling observer will see temperatures of the order of the ultraviolet cutoff scale upon crossing the stretched horizon. This then leads to a violation of the postulates of black hole complementarity, since the physics outside but close to the stretched horizon is no longer described by a conventional theory.

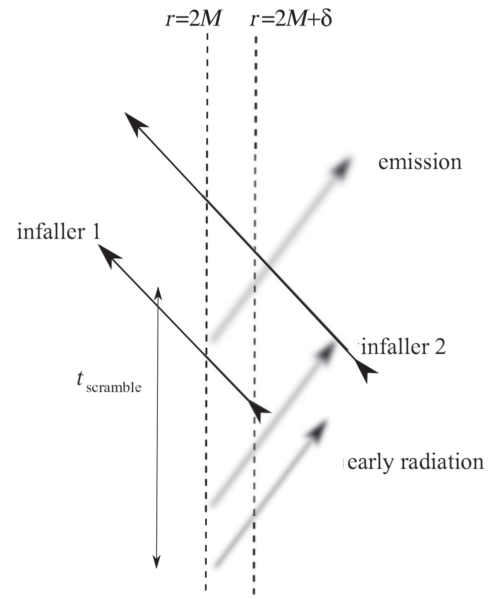

FIG. 3. The analog of Fig. 2 with the diary replaced the ordinary Hawking modes. An infaller measuring a mode outside the stretched horizon (infaller 2) will see it maximally entangled with the early Hawking radiation. However, an earlier infalling observer (infaller 1) must see vanishing entanglement with the early Hawking radiation if postulate 4 holds.

However, it suffices to show the firewall needs to only ever appear behind the stretched horizon to exhibit the flaw in the reasoning of Ref. [3]. To do this it is helpful to work with the Hilbert space separated as shown in Fig. 4. The Hartle-Hawking vacuum involves an entanglement of the modes on each side of the horizon [22]. However, the left modes never propagate into the external region on the right. Both sets of modes propagate into the interior of the black hole, and their entanglement is essential for the absence of drama for an infalling observer.

In this picture, the exterior modes are a superposition of infalling and outgoing modes. Consistency with Fig. 3 then demands that modes representing Hawking particles

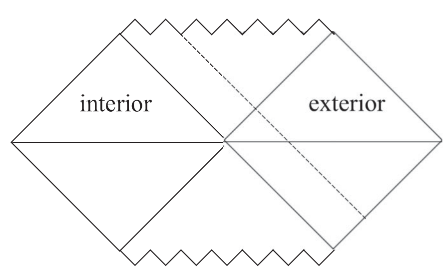

FIG. 4. The Penrose diagram for the maximally extended Schwarzschild black hole. The area to the right of the dashed line provides a classical model for black hole formation. The Hilbert space of states may be factored into states on the interior and the exterior along the time slice indicated by the horizontal line. Both sets of modes propagate at later times into the upper quadrant.

emitted after a time of the order of the Page time M^3 be maximally entangled with the earlier radiation. From the exterior viewpoint, there is no contradiction with unitarity and locality. The problem arises when one considers an infalling observer. Following the argument of Ref. [3], the exterior mode cannot be simultaneously entangled with the early Hawking radiation and the interior mode.

We can model a state where the exterior modes have no entanglement with the interior modes by simply placing the left interior modes in their vacuum state. The argument is cleanest if the exterior modes are placed in a thermal density matrix rather than a pure state. Unlike the Boulware vacuum, this does not change the expectation value of the stress-energy tensor in the exterior region [23]. It does, however, produce an infinite firewall on the global horizon. However, this is a crucial difference, because now we have a counterexample where the firewall only needs to appear behind the stretched horizon, and the expectation value of the stress-energy tensor needs not depart by a substantial amount from that obtained in the Hartle-Hawking vacuum, even if the stretched horizon is Planck scale. Therefore, we can conclude that outside a Planck distance from the global horizon, there is no sign that the postulates of black hole complementarity break down. At or inside the global horizon, all bets are off for a conventional description of the quantum theory, as emphasized in Refs. [6–8].

It is also interesting to estimate whether the outgoing Hawking radiation leads to an observable deviation in a local quantity outside the stretched horizon, such as the expectation value of the stress-energy tensor. Along the path of the early infalling observer in Fig. 2, the result will match that of Ref. [5], which found, for example, a $1/M^4$ contribution to the trace of the stress-energy tensor near the horizon due to Hawking radiation in the Hartle-Hawking vacuum. The later infalling observer will see the same phenomena, with small differences due to the interference with the earlier outgoing Hawking radiation. Since the outgoing radiation at this point is maximally entangled, any fluctuations away from the thermal expectation value for the stress-energy tensor are expected to be down by an extra factor of $1/M$ or more.

III. ENTROPY SUBADDITIVITY BOUNDS

Finally, we note that one of the arguments for the firewall of Ref. [3] is based on entropy subadditivity bounds [24,25]. They divide the system into A, the early Hawking modes; B, a late outgoing Hawking mode; and C, the interior partner mode of B. They claim the entropy subadditivity bound

$$S_{AB} + S_{BC} \geq S_B + S_{ABC} \tag{6}$$

is violated. Let us analyze this bound, first in the stretched horizon theory of postulate 2 and then in a model with time slices that extend inside the stretched horizon but terminate on the global horizon.

In the effective field theory of postulate 2, it is incorrect to view Hawking radiation as the formation of a maximally entangled pair B and C outside the stretched horizon, with the negative energy C mode subsequently absorbed by the stretched horizon. Introducing C degrees of freedom outside the stretched horizon, and insisting that they are maximally entangled with the outgoing B modes, indeed leads to a violation of entropy subadditivity as we will see momentarily. It amounts to cloning of quantum information, and by assumption the effective field theory of postulate 2 is a local quantum field theory where such cloning cannot occur. Rather the effective field theory of postulate 2 only describes the A and the B modes. In this theory the stretched horizon is a boundary of spacetime. It is a hot surface from which the Hawking radiation is emitted. At late times the state of the stretched horizon of the remaining black hole is maximally entangled with A, the early Hawking modes. When a late Hawking mode B is emitted, the size of the stretched horizon Hilbert space gets reduced accordingly, and both B and the new stretched horizon state are separately maximally entangled with A. There is, however, no entanglement between B and the new stretched horizon state. This is entirely in line with what happens at late times for a burning lump of coal that starts out in a pure state. Entropy subadditivity reduces to the statement [24]

$$|S_A - S_B| \leq S_{AB} \leq S_A + S_B, \tag{7}$$

which is close to saturated at late times $S_{AB} = S_A - S_B$.

We now turn our attention to a description where time slices extend inside the stretched horizon and C modes are included. If we take that description to be a conventional local quantum field theory, then we will run into problems with entropy subadditivity as pointed out in Ref. [3], and these problems can indeed be avoided by introducing a firewall for infalling observers. It is important to note, however, that in this case we are no longer considering the effective field theory of postulate 2 but have made the further assumption that the local effective field theory can be extended to the region inside the stretched horizon.

For the sake of argument, let us instead consider a model where C modes are included and the physics inside the black hole region is described by some quantum dynamics on the global horizon (rather than on the stretched horizon as in postulate 2). A BC pair appears in a pure state due to a quantum fluctuation, and we assume that the C mode then scrambles with the state on the global horizon in a time of order $M \log M$. There are two limits where the entropy subadditivity bound can be easily analyzed: before scrambling has had a chance to occur and after the scrambling time. Prior to scrambling, BC remains in a pure state independent of A, so $S_{BC} = 0$, and $S_{ABC} = S_A + S_{BC} = S_A$. Substituting into Eq. (6) yields

$$S_{AB} \geq S_B + S_A.$$

At first sight this seems similar to the analysis of Ref. [3]. However, before C is scrambled, we expect the entropy of the outgoing radiation to increase,

$$S_{AB} > S_A, \tag{8}$$

rather than decrease as stated in Ref. [3] because one simply has one additional thermal Hawking particle. We conclude $S_{AB} = S_A + S_B$ because A and B are independent prior to scrambling.

It is helpful to break the interaction of C with the black hole, described by some Hilbert subspace D, into two steps. First C interacts with the black hole, reducing the number of degrees of freedom there. This process requires working in some infinite dimensional Fock space, as is usual in second quantized field theory. However, immediately after this interaction, there will be a dimension $\dim(C)$ subspace of the global horizon Hilbert space that is entangled both with B and with A.

After scrambling things become simpler. Scrambling mixes C with all the other horizon degrees of freedom. B will become maximally entangled with A, so $S_{AB} = S_A - S_B$ and the entropy of the external radiation decreases, $S_{AB} < S_A$. After scrambling it is no longer true that $S_{BD} = 0$. Rather if the dimension of A is much larger than BD, we expect B and D to become independent, with $S_{BD} \approx S_B + S_D = 2S_B$. Likewise the BD system will be close to maximally entangled with A after scrambling, so $S_{ABD} \approx S_A - S_{BD}$. Substituting into Eq. (6) yields

$$S_A - S_B + S_{BD} \geq S_B + S_A - S_{BD} \Longleftrightarrow S_{BD} \geq S_B,$$

which is satisfied.

It should be noted that the reduced density matrix

$$\rho_{AB} = \mathrm{Tr}_{CD}\rho_{ABCD}$$

is independent of unitary transformations that act within the $C \times D$ subspace. Therefore, the only way to accomplish such a change of entanglement described above is via a nonunitary transformation. However, if we follow the picture described in the previous section, such a nonunitary horizon theory is only needed on the global horizon rather than the stretched horizon and so remains consistent with the postulates of black hole complementarity.

We find no violation of entropy subadditivity implied by the postulates of black hole complementarity. The correct description of the stretched horizon theory does not allow for a description of interior C modes that is independent of the outgoing B modes. Reference [3] assumes that BC pairs form outside the stretched horizon, which amounts to cloning quantum information in the exterior region. This leads to the claim that simultaneously $S_{BC} = 0$ (as in our global horizon model prior to scrambling) and $S_{AB} < S_A$ (as in our model only after scrambling).

IV. DISCUSSION

The main point of this paper is that a firewall for infalling observers is not an unavoidable consequence of postulates 1 and 2, as is claimed by Ref. [3]. We do not claim that such a firewall is impossible. In fact, we have considered several examples where firewalls do occur. In each case, including that of Ref. [3], the firewall follows from making assumptions about physics in the region inside the stretched horizon that do not follow from postulate 2.

It is interesting to further consider the history of the infalling observers in Fig. 2. As a model for the dynamics inside the horizon, let us imagine we use the simple nonunitary model described in the previous section. The early infalling observer smoothly passes through the stretched horizon according to postulate 4. However, once an interval of the order of the scrambling time passes, the entanglement between the B and the C modes will change, and this observer will no longer experience a vacuum state. By this time they will have passed a distance at least of order M inside the stretched horizon. This, however, coincides with the location of the curvature singularity. This picture, where information cloning was prevented by a firewall located near the classical curvature singularity, was advocated in Ref. [26] and supported by AdS/CFT computations in Ref. [27]. These works focus on resolving cross-horizon complementarity issues rather than the outside-the-horizon issues that are the main focus of the present work. Scrambling therefore allows the early infalling information to be safely annihilated before the later infalling observer, who has access to the information from the exterior Hawking radiation, is able to enter the horizon and potentially see a contradiction.

ACKNOWLEDGMENTS

We thank J. Polchinski and the authors of Ref. [13] for helpful comments. The research of K. L. and D. A. L. is supported in part by DOE Grant No. DE-FG02-91ER40688-Task A and an FQXi grant. The research of L. T. is supported in part by grants from the Icelandic Research Fund and the University of Iceland Research Fund.

[1] L. Susskind, L. Thorlacius, and J. Uglum, Phys. Rev. D **48**, 3743 (1993).
[2] L. Susskind and L. Thorlacius, Phys. Rev. D **49**, 966 (1994).
[3] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, J. High Energy Phys. 02 (2013) 062.
[4] *Black Holes: The Membrane Paradigm*, edited by K. S. Thorne, R. H. Price, and D. A. MacDonald (Yale University, New Haven, CT, 1986).
[5] P. Candelas, Phys. Rev. D **21**, 2185 (1980).
[6] D. A. Lowe, J. Polchinski, L. Susskind, L. Thorlacius, and J. Uglum, Phys. Rev. D **52**, 6997 (1995).
[7] D. A. Lowe, Nucl. Phys. **B456**, 257 (1995).
[8] D. A. Lowe and L. Thorlacius, Phys. Rev. D **60**, 104012 (1999).
[9] S. B. Giddings and M. Lippert, Phys. Rev. D **69**, 124019 (2004).
[10] Y. Nomura, J. Varela, and S. J. Weinberg, J. High Energy Phys. 03 (2013) 059.
[11] S. D. Mathur and D. Turton, arXiv:1208.2005.
[12] B. D. Chowdhury and A. Puhm, arXiv:1208.2026.
[13] S. G. Avery, B. D. Chowdhury, and A. Puhm, arXiv:1210.6996.
[14] T. Banks and W. Fischler, arXiv:1208.4757.
[15] S. B. Giddings, Y. Shi, and Y. Shi, Phys. Rev. D **87**, 064031 (2013).
[16] D.-I. Hwang, B.-H. Lee, and D.-H. Yeom, J. Cosmol. Astropart. Phys. 01 (2013) 005.
[17] S. Hossenfelder, arXiv:1210.5317.
[18] P. Hayden and J. Preskill, J. High Energy Phys. 09 (2007) 120.
[19] Y. Sekino and L. Susskind, J. High Energy Phys. 10 (2008) 065.
[20] If, for example, the message was unentangled with the bits received by Bob, he would still have a probability of success of 2^{-k} of decoding the message by measuring the first k bits. However, this does not improve as additional bits are measured, so it does not correspond to a useful measurement of information in the diary.
[21] Note that entanglement depends on the time slicing in this context and is a quantity not accessible to local probes. Nevertheless, for a given time slicing, the state undergoes unitary time evolution, and all external observers agree on the state.
[22] W. Unruh, Phys. Rev. D **14**, 870 (1976).
[23] One may also place the exterior modes in a pure state where the expectation value of the stress-energy tensor remains very close to that of the Hartle-Hawking vacuum. Note such a state involves an infinite number of infalling and outgoing exterior modes, so the state may be picked so fluctuations in such a local observable are very small.
[24] H. Araki and E. Lieb, Commun. Math. Phys. **18**, 160 (1970).
[25] E. Lieb and M. Ruskai, J. Math. Phys. (N.Y.) **14**, 1938 (1973).
[26] D. A. Lowe and L. Thorlacius, Phys. Rev. D **73**, 104027 (2006).
[27] D. A. Lowe, Phys. Rev. D **79**, 106008 (2009).