Remarks on Black Hole Complexity Puzzle

Beni Yoshida

Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada

Abstract

Recently a certain conceptual puzzle in the AdS/CFT correspondence, concerning the growth of quantum circuit complexity and the wormhole volume, has been identified by Bouland-Fefferman-Vazirani and Susskind. In this note, we propose a resolution of the puzzle and save the quantum Extended Church-Turing thesis by arguing that there is no computational shortcut in measuring the volume due to gravitational backreaction from bulk observers. A certain strengthening of the firewall puzzle from the computational complexity perspective, as well as its potential resolution, is also presented.

1 Introduction

Recently a certain conceptual puzzle in the AdS/CFT correspondence has been identified by Bouland-Fefferman-Vazirani (BFV) [1] and Susskind [2]. The puzzle is concerned with the growth of quantum circuit complexity in boundary CFTs and its dual on the bulk; the growth of the wormhole volume. It is widely believed that finding the quantum circuit complexity of a quantum state is generically a difficult computational problem. Thus, there should be no efficient way of computing the complexity of the boundary CFT wavefunction. But the wormhole volume is a macroscopic quantity which appears to be easily measurable by bulk observers. This would suggest that a certain physical phenomena in quantum gravity may not be efficiently simulated on a quantum computer, violating the quantum Extended Church-Turing (qECT) thesis [3].

In this note, we propose a resolution of this puzzle. Any attempt to see the black hole interior by an observer always introduces significant gravitational backreaction to the underlying geometry, and naive predictions on the black hole interior based on effective bulk descriptions break down. This invalidates protocols that would appear to measure the volume/complexity efficiently on the bulk, and thus the qECT thesis remains valid. To demonstrate this point explicitly, we discuss how to construct interior entangled partner operators with gravitational backreaction taken into account. Namely, we show that construction of interior operators changes dynamically due to backreaction from an infalling observer. This conclusion follows from a certain quantum information theoretic theorem which relates quantum information scrambling, as quantified by OTOCs, to the decoupling phenomena. The theorem suggests that the infalling observer's backreaction disentangles the outgoing Hawking mode from the other side

of the black hole. We also critically comment on the black hole complementarity approach to the complexity puzzle advocated by Susskind [2].

This note is written in a non-technical manner with the hope to convey main messages effectively. This note is organized as follows. In Section 2 and 3, we provide a brief review of the black hole complexity puzzle. In Section 4, we discuss the backreaction from the infalling observer. In Section 5, we present the resolution of the puzzle. We also present a refined interpretation of a certain thought experiment due to Hayden and Preskill. In Section 6, we present brief discussions. In Section 6.1, we present a certain strengthening of the firewall puzzle from the perspective of the qECT thesis. In Appendix A, we present answers to some of frequently asked questions.

2 Complexity

The quantum Extended Church-Turing (qECT) thesis is the following belief/physical principle [3]:

– *All the physical processes that obey fundamental laws of physics, including quantum gravity, are efficiently simulable on a quantum computer.*

A thesis is a proposal for the sake of argument. It is not formulated as a precise mathematical conjecture and its correctness remains to be verified. If the qECT thesis is wrong, one could potentially speed up quantum computation by using quantum gravity effects [1]. The AdS/CFT correspondence is a prime setup to test the qECT thesis as it posits that all the physical phenomena in the bulk quantum gravity are encoded in boundary quantum mechanical systems and may be efficiently simulated on a quantum computer.

To derive the puzzle, we will need a contraposition of the qECT thesis within the AdS/CFT correspondence.

– *Any physical processes that cannot be efficiently simulated in boundary CFTs cannot be efficiently simulated in bulk AdS quantum gravity.*

To make the logic of this note clear, it is worth representing the statement of the qECT thesis schematically:

bulk		boundary
easy	$\rightarrow$	easy
difficult	$\leftarrow$	difficult

Since the bulk consists both of quantum mechanics and gravity, it is reasonable to think that the bulk is stronger than (or equal to) the boundary in terms of the computational power [2]. The qECT thesis says that the AdS gravity does not provide any quantum computational speedup.

[1] It is implicitly assumed that a physical system can be described in a finite-dimensional Hilbert space.

[2] If some problem is easy on the boundary, one can simply bring a quantum computer to the bulk and solve it without using gravity. Hence, "easy $\leftrightarrow$ easy" and "difficult $\leftrightarrow$ difficult".

The central object of interest is the complexity of boundary wavefunctions in the AdS/CFT correspondence. The quantum circuit complexity $\mathcal{C}$ of a quantum state $|\psi\rangle$ is the minimal number of few-body quantum gates that are required to create $|\psi\rangle$ from some simple reference state such as $|0\rangle^{\otimes n}$. While the precise definition of $\mathcal{C}$ depends on choices of elementary gate sets and other details, these subtleties will not be essential in demonstrating and resolving the complexity puzzle.

Recall that the two-sided eternal AdS black hole is conjectured to be dual to the thermofield double (TFD) state $|\text{TFD}\rangle \propto \sum_i e^{-\beta E_i}|\psi_i\rangle_L \otimes |\psi_i\rangle_R$. Here we will focus on black hole geometries arising from time-evolutions of $|\text{TFD}\rangle$ under weak perturbations that can be treated as gravitational shockwaves. We will assume that thermodynamic properties of the black hole, such as the temperature, do not change by such perturbations. The following conjecture was proposed [4]:

– *The quantum circuit complexity $\mathcal{C}$ of boundary CFT wavefunction is dual to the wormhole volume $\mathcal{V}$ in the maximal volume slice.*

In other words, $\mathcal{C} \approx \mathcal{V}$ in some appropriate normalization.

The $\mathcal{C} \approx \mathcal{V}$ conjecture has passed several non-trivial tests in a qualitative sense. To gain some insight, consider a time-evolved state $|\text{TFD}(t)\rangle \equiv (e^{-iHt} \otimes I)|\text{TFD}\rangle$. It is naturally expected that the complexity of $|\text{TFD}(t)\rangle$ linearly increases in t when H is chaotic. While this statement is unproven, there is a large body of supporting evidence from various perspectives, such as quantum complexity theory [5] and pseudorandomness [6, 7]. On the bulk, consider the time-evolved TFD state $|\text{TFD}(t)\rangle = |\text{TFD}(\frac{t}{2}, -\frac{t}{2})\rangle$ and its dual wormhole in the maximal volume slice. Here we used the fact that $|\text{TFD}(t)\rangle = |\text{TFD}(t_L, t_R)\rangle \equiv (e^{-iHt_L} \otimes e^{+iHt_R})|\text{TFD}\rangle$, where $t = t_L - t_R$. The volume $\mathcal{V}$ indeed grows roughly linearly in time t for an exponentially long time, as in Fig. 1(a). Hence, after some proper normalization, we expect to have $\mathcal{C} \approx \mathcal{V}$. Another justification of $\mathcal{C} \approx \mathcal{V}$ is obtained by invoking the tensor network representation of the spacetime [8].

In demonstrating the black hole complexity puzzle, we will be interested in the computational difficulty of determining $\mathcal{C}$ from $|\psi\rangle$, not $\mathcal{C}$ itself. Suppose that some unknown wavefunction $|\psi\rangle$ is given. Without prior knowledge on how $|\psi\rangle$ was prepared, determining its quantum circuit complexity $\mathcal{C}$ is a difficult computational problem. In the most generic setting of the problem, we could at best try to apply quantum gates to see if $|\psi\rangle$ returns to some simple state. Since there are $\exp\big(O(\mathcal{C})\big)$ nearly-orthogonal states with complexity $\mathcal{C}$, the complexity of estimating $\mathcal{C}$ is expected to be $\sim \exp(\mathcal{C})$.

– *The quantum computational complexity of estimating $\mathcal{C}$ is $\sim \exp(\mathcal{C})$ in general, and thus is a difficult computational problem.*

While unproven, there are quantum computational complexity theoretic arguments supporting this conjecture, see [1] for instance.

Let us pause for a technical comment. If one knows the Hamiltonian H a priori, it is not difficult

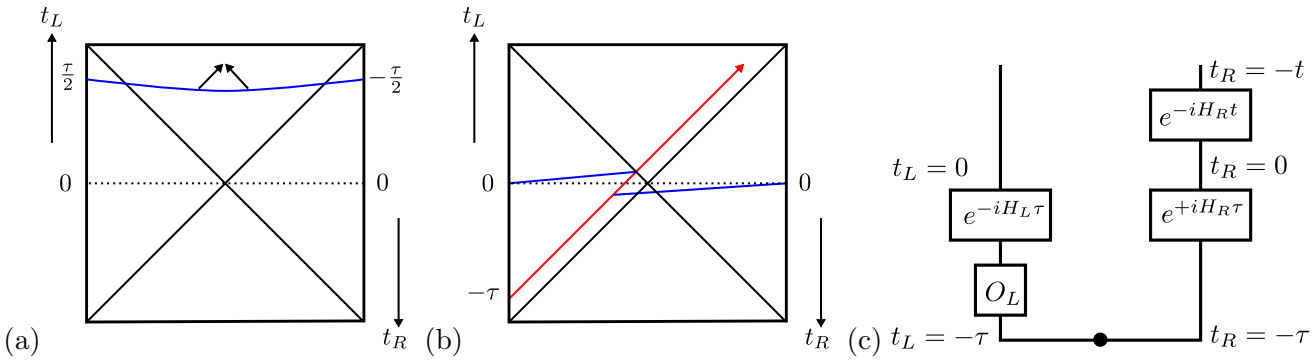


Figure 1: (a) A maximal volume slice. A pair of bulk observers in the interior are shown. (b) A slice with a shock from the past $t_L = -\tau$. (c) A circuit representation of a shockwave geometry. The horizontal line represents the TFD state.

to estimate the complexity of $|\text{TFD}(t)\rangle$ as one can undo the time-evolution and see if the state returns to $|\text{TFD}(0)\rangle$. To circumvent this counterargument, BFV considered an ensemble of wavefunctions generated by insertions of multiple randomly-chosen shockwaves $\{O_1, \ldots, O_m\}$:

$$(I \otimes e^{-iH_R t_m} O_m e^{-iH_R t_{m-1}} \cdots e^{-iH_R t_1} O_1)|\text{TFD}\rangle. \tag{1}$$

They provided an argument suggesting that it is indeed difficult to estimate $\mathcal{C}$ even if one knows H since it is difficult to distinguish the above ensemble from some random ensemble of much higher complexity. Here it is important to assume that the choice of $\{O_1, \ldots, O_m\}$ remains unknown to the boundary observer. Unless otherwise stated, we shall focus on $|\text{TFD}(t)\rangle$ without insertion of multiple shockwaves for simplicity of discussion.

Combining the above conjecture with another one $\mathcal{C} \approx \mathcal{V}$, we arrive at the following conjecture:

– *Estimating $\mathcal{V}$ from generic boundary CFT wavefunctions is a difficult computational problem.*

Below is a side comment. There is a certain heuristic relation between two-point correlation functions and the complexity of estimating $\mathcal{C}$. Two-point correlation functions between two sides of a black hole decay exponentially as the separation increases. This suggests that measurement of two-point functions by bulk observers, which would go as $\approx e^{-\gamma\mathcal{C}}$, can give an estimate of the wormhole length where γ is some positive constant. When the complexity $\mathcal{C}$ is high, the measurement outcome is too weak, and one would need $\approx e^{+\gamma\mathcal{C}}$ samples to make a reliable measurement of two-point functions. This is consistent with the fact that the complexity of estimating $\mathcal{C}$ is conjectured to be $\approx e^{\gamma'\mathcal{C}}$ for some constant γ'. Turing the argument around, the exponential growth of the complexity of estimating $\mathcal{C}$ would suggest the exponential (not polynomial) decay of two-point functions with $\gamma \gtrapprox \gamma'$. It is interesting that bulk field theories can put some restrictions on the complexity of estimating $\mathcal{C}$ [3]. If the bulk theory is somehow gapless with polynomially decaying correlation functions, the complexity of estimating $\mathcal{C}$ would be at

[3]It is also interesting to note that two-point correlation functions at late times measure spectral form factors [9]. These contain linearly growing contribution, much like the wormhole volume [10].

most poly($\mathcal{C}$).

3 Puzzle

In the light of the qECT thesis, estimating the wormhole volume $\mathcal{V}$ should also be computationally difficult since estimating $\mathcal{C}$ is difficult. This is, however, strange. The wormhole volume is roughly proportional to the length of the wormhole which is a simple macroscopic quantity that appears to be easily computed or measured. This would suggest that estimating $\mathcal{V}$ may actually be a computationally tractable problem for bulk observers, violating the qECT thesis.

BFV sharpened this observation by proposing a certain protocol to measure the wormhole volume by using multiple observers who live on multiple copies of the system. Imagine that one somehow populates the black hole interior with many observers along the wormhole, and see if pairs of observers can send signals to one another before reaching the singularity (Fig. 1(a)). The number of successful signal transmissions will be a coarse estimate of the wormhole volume.

One potential problem in the BFV protocol is the use of multiple observers who are causally disconnected. Another issue is that, to place observers in the interior, one would need to know the shape of the wormhole a priori, which implicitly assumes prior knowledge of $\mathcal{V}$. At the same time, however, the description of the protocol in [1] is rather brief, and we are unsure if we have addressed its full intent [4].

Susskind presented another version of the puzzle which is free from these problems. Suppose that some boundary wavefunction $|\psi(t)\rangle$ undergoes time-evolution. Since it is difficult to estimate the complexity of $|\psi(t)\rangle$, it should be also difficult to judge if $\mathcal{C}(t)$ is increasing or decreasing at given time. Combining this observation with the $\mathcal{C} \approx \mathcal{V}$ conjecture, the following conjecture is obtained.

– *Judging if $\mathcal{V}(t)$ is increasing or decreasing from boundary CFT is computationally difficult.*

To demonstrate the puzzle based on this observation, let us look at two classes of boundary wavefunctions. First, consider the time-evolution of the TFD state $|\Psi(t)\rangle \equiv |\mathrm{TFD}(0, -t)\rangle$ at $t_L = 0$ and $t_R = -t$. As we have seen, $\mathcal{C}(t)$ of $|\Psi(t)\rangle$ increases since $\mathcal{V}(t)$ increases in time (Fig. 1(a)).

Next, consider another wavefunction at $t_L = 0$ and $t_R = -t$ where, at some negative time $t_L = -\tau$, a small perturbation is applied (Fig. 1(b)):

$$|\Phi(t)\rangle \equiv (e^{-iH\tau} \otimes I_R)(O_L \otimes I_R)|\mathrm{TFD}(-\tau, -t)\rangle. \tag{2}$$

It has been pointed out that $\mathcal{C}(t)$ of $|\Phi(t)\rangle$ actually decreases for $\tau \gtrapprox t$. On the bulk, this can be verified by directly computing the volume $\mathcal{V}$, see [4] for details. On the boundary, this decrease can be easily seen by drawing a quantum circuit diagram as in Fig. 1(c). Here we have prepared the initial state $|\Phi(0)\rangle$ by inserting a perturbation on the thermofield double state $|\mathrm{TFD}(0)\rangle = |\mathrm{TFD}(-\tau, -\tau)\rangle$ at

[4] Perhaps a simpler protocol would be to prepare a pair of observers from two sides, without telling them what H is, and let them meet inside the black hole and compare their watches. Unfortunately this protocol does not work as we will discuss later.

$t_L = t_R = -\tau$, and time-evolving both sides to $t_L = t_R = 0$. We see that time-evolution of $e^{-iH_R t}|\Phi(0)\rangle$ on the right hand side cancels the $e^{iH_R\tau}$, and hence the complexity $\mathcal{C}(t)$ actually decreases until $t = \tau$.

The take-away from the above analysis is that complexity of wavefunctions can decrease or increase depending on whether a perturbation is applied on the left side or not. This is, however, strange from the bulk observer's perspective. An observer simply needs to jump into the black hole and see if she will be hit by a gravitational shockwave which lets her allow to judge the sign of $\frac{d\mathcal{C}}{dt}$. Once again, the qECT thesis seems to be violated.

Several flaws in the above arguments can be immediately identified. Introducing a bulk observer behind the horizon corresponds to high complexity quantum operation on the boundary. Namely, when approaching near the horizon, an infalling observer herself induces significant gravitational backreaction to the underlying geometry. These effects often invalidate bulk effective descriptions, and hence should be examined carefully, which is exactly what we will do in the next section. Also it is unclear why an infalling observer from the right hand side is able to see a shockwave from the left hand side while two sides are not coupled. Later we shall indeed see that an observer will not be able to see the shockwave due to the backreaction caused by the observer herself.

In the reminder of the note, we argue that measuring $\mathcal{V}$ or $\frac{d\mathcal{V}}{dt}$ on the bulk is not necessarily a computationally easy task by examining the effect of backreaction by infalling observers.

For convenience of readers, we summarize the argument of the complexity puzzle [5].

1). On the boundary, quantum circuit complexity $\mathcal{C}$ (or $\frac{d\mathcal{C}}{dt}$) is believed to be not efficiently computable.

2). On the bulk, the wormhole volume $\mathcal{V}$ (or $\frac{d\mathcal{V}}{dt}$) seems to be efficiently measurable by infalling observers.

3). The wormhole volume $\mathcal{V}$ and the quantum circuit complexity $\mathcal{C}$ are roughly proportional to each other.

4). The qECT says that, as $\mathcal{C}$ (or $\frac{d\mathcal{C}}{dt}$) is not efficiently computable, $\mathcal{V}$ (or $\frac{d\mathcal{V}}{dt}$) should not be efficiently measurable either.

We will argue that 2) is wrong. Here 4) is the statement of the qECT thesis while 1) and 3) are widely-accepted conjectures.

4 Backreaction

The main goal of this section is to discuss the backreaction from an infalling observer on the black hole interior geometries. Our strategy is to find interior partner modes that are entangled with outgoing modes in boundary CFTs. By studying how entanglement structure changes by the addition of the

[5] BFV's argument is a bit more involved and considers the complexity of the holographic dictionary. We will comment on it in Section 6.

infalling observer, we can deduce bulk geometries with appropriate gravitational backreaction. This section is a short summary of [11, 12].

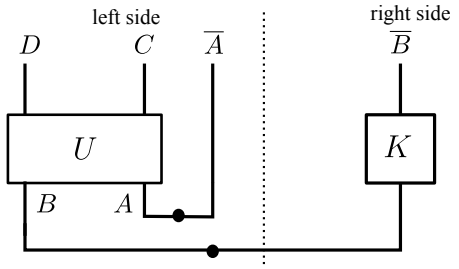
For simplicity of discussion, we represent the black hole as a quantum system of n qubits. As the initial state of the black hole, consider a generic maximally entangled state between the left side B and the the right side $\overline{B}$:

$$(I \otimes K)|\mathrm{EPR}\rangle_{B\overline{B}} \qquad |\mathrm{EPR}\rangle_{B\overline{B}} \equiv \frac{1}{\sqrt{d_B}} \sum_j |j\rangle_B \otimes |j\rangle_{\overline{B}} \tag{3}$$

where K is an arbitrary unitary operator. Imagine that some measurement probe (or an infalling observer) A is dropped into the black hole at time $t = 0$ as in Fig. 2(a). Rather than keeping track of the outcomes for all the possible input states on A, it is convenient to append a reference system $\overline{A}$ which is entangled with the probe A and forming EPR pairs $|\mathrm{EPR}\rangle_{A\overline{A}} = \frac{1}{\sqrt{d_A}} \sum_j |j\rangle_A \otimes |j\rangle_{\overline{A}}$.

After the time-evolution by some unitary dynamics U, the system evolves to

$$|\Psi\rangle \equiv (U_{BA} \otimes I_{\overline{A}} \otimes K_{\overline{B}})(|\mathrm{EPR}\rangle_{B\overline{B}}|\mathrm{EPR}\rangle_{A\overline{A}}) = \tag{4}$$



where D is the outgoing mode and C is the remaining black hole. Horizontal lines represent EPR pairs. Here black dots represent factor of $1/\sqrt{d_R}$ in a subsystem R for proper normalization of $|\mathrm{EPR}\rangle_{R\overline{R}}$. A dotted line divides the left and right sides of the black hole.

In the absence of the infalling observer A, the outgoing mode D is entangled with some mode R_D on the right hand side $\overline{B}$. Here we are interested in finding what degrees of freedom D is entangled with after adding the infalling observer A to the system.

Let us begin by briefly recalling the concept of quantum information scrambling. A quantum black hole delocalizes quantum information rapidly and its effect on boundary CFTs can be characterized by out-of-time order correlations (OTOCs). Let V_A and W_D be arbitrary traceless basis operators, such as Pauli or Majorana operators, supported on A and D respectively. Initially at $t = 0$, we have $\langle V_A(0)W_D(0)V_A^\dagger(0)W_D^\dagger(0)\rangle = 1$ when V_A and W_D do not overlap. Here we took the quantum state to be the maximally mixed state $\rho = \frac{1}{d}I$. After the scrambling time, on the other hand, OTOCs decay to small values:

$$\langle V_A(0)W_D(t)V_A^\dagger(0)W_D^\dagger(t)\rangle \approx 0 \qquad t \gtrapprox t_{\mathrm{scr}} \tag{5}$$

For more details on the definition of scrambling via OTOCs, see [13].

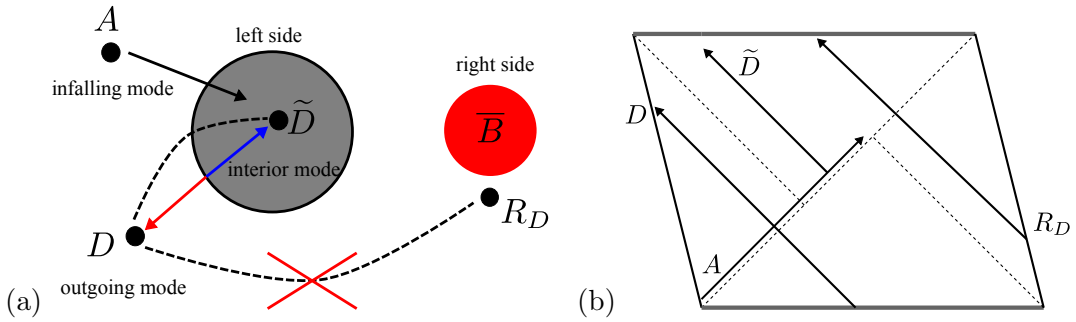


Figure 2: (a) Backreaction by an infalling observer. The outgoing mode D is entangled with a new mode $\widetilde{D}$ which is dynamically created by A and has nothing to do with RHS. (b) Bulk interpretation with a shockwave.

The inclusion of the infalling observer has drastic effect on entanglement structure due to the scrambling dynamics of the black hole. The following statement was proven in [14]. Suppose that the system is scrambling in the above sense and $d_A \gg d_D$ [6] . Then, the subsystems D and $\overline{B}$ in $|\Psi\rangle$ are decoupled (not entangled):

$$\rho_{D\overline{B}} \approx \rho_D \otimes \rho_{\overline{B}}. \tag{6}$$

where the error is $O\left(\frac{d_D^2}{d_A^2}\right)$. Hence, we arrive at the following conclusion:

– *The outgoing mode D is no longer entangled with the right hand side $\overline{B}$. Instead D is entangled with $C\overline{A}$. Thus the entangled partner mode is found exclusively on the left hand side.*

An immediate, yet important implication of this conclusion is that the construction of the interior partner mode is independent of the initial state of the black hole (or the unitary K in Eq. (3)). In fact, the same expression of the interior partner mode works for any initial state of the black hole, be it two-sided or one-sided since it is supported exclusively on the left hand side without using degrees of freedom from $\overline{B}$. Hence, infalling observer's backreaction enables us to obtain *state-independent* construction of interior partner operators [7].

The bulk interpretation of the above disentangling phenomena can be obtained by treating the backreaction from an infalling observer as a gravitational shockwave (Fig. 2(b)). *In the absence of the infalling observer,* given the outgoing mode D, interior partner operators can be constructed by time-evolving a corresponding mode on the right hand side. This mode, constructed exclusively on the degrees of freedom on the right hand side $\overline{B}$, is denoted by R_D. We now include the effect of the infalling observer A as a gravitational shockwave and draw the backreacted geometry where the horizon is shifted as depicted in Fig. 2(b). If the time separation between the outgoing mode D and the infalling

[6] See [11] for discussions on cases where the time separation is shorter than the scrambling time as well as physical interpretation of the condition $d_A \gg d_D$.

[7] It is worth emphasizing, however, that the construction depends on the initial state of the infalling observer, and thus is *observer-dependent*. See [12] for details.

observer A is longer than or equal to the scrambling time, the interior mode $\widetilde{D}$, which is *distinct* from R_D, can be found across the horizon [8].

One successful application of the disentangling phenomena by an infalling observer is the resolution of the firewall puzzle. Recall that the outgoing mode D was initially entangled with some degrees of freedom R_D in $\overline{B}$. Assuming the smooth horizon, an infalling observer Alice would see an interior mode $\widetilde{D}$ which is entangled with the outgoing mode D. This, however, leads to a contradiction because D is also entangled with R_D. If we would think that D remains entangled with R_D, then D and $\widetilde{D}$ would not be entangled, leading to possible high energy density at the horizon.

Then the resolution of the firewall puzzle is immediate. When an infalling observer jumps into a black hole, the outgoing mode D is disentangled from R_D due to her own backreaction. She will be able to observe the interior mode $\widetilde{D}$ which is distinct from the original partner mode R_D. We will revisit this resolution by strengthening the firewall puzzle from the perspective of the qECT thesis in Section 6.

5 Resolution

We begin by arguing that Susskind's protocol for measuring $\frac{d\mathcal{C}}{dt}$ cannot be performed. In his protocol, an infalling observer from one side jumps into the black hole and see if she will be hit by a shockwave from the other side or not. As is already mentioned, it is strange to expect that the infalling observer could see the shockwave when two sides of the black hole are not coupled [9]. The key to resolve this misunderstanding is to consider the effect of backreaction by an infalling observer as in Fig. 2(b). Assume that the shockwave was created by exciting the boundary mode R_D. Since this mode R_D was initially entangled with D, the infalling observer would expect to be hit by the entangled partner mode of D behind the horizon. Due to her own backreaction, however, D will no longer be entangled with R_D once the infalling observer jumps into the black hole. Inside the black hole, she will just see the interior mode $\widetilde{D}$ which is distinct from R_D. Hence the infalling observer will not be able to see the shockwave from the other side .

Here we assumed that two sides of the black hole are not coupled. In order for two signals/observers from opposite sides to meet inside the black hole, two boundaries need to be coupled in an appropriate manner. By using the traversable wormhole phenomena [16] or the Hayden-Preskill decoding protocol [13], two observers may be able to see each other and then travel to the other sides of the black hole. To implement these protocols, however, one needs to reduce the boundary CFT wavefunctions back to low complexity states (such as the TFD state), so there is no shortcut in these approaches. It is unclear to us if there would be a simpler way to make two observers meet inside the black hole. To influence two

[8] The bulk picture, where the infalling observer A herself is treated as a shockwave, suggests that A will encounter the partner mode $\widetilde{D}$ shortly after crossing the horizon. By the time A encounters the outgoing mode D, the partner mode dynamically changes from R_D to $\widetilde{D}$ due to her backreaction. Some readers might think that A would eventually encounter R_D after her encounter with $\widetilde{D}$. Observe, however, that this encounter would happen almost at the singularity. It is not surprising if the naive bulk picture makes a false prediction near the singularity. See appendix A for further discussions.

[9] A similar puzzle was identified in [15] which asked why two observers from the two sides appear to be able to meet inside the black hole. Our resolution of this rendezvous puzzle, however, differs from the scenario proposed in [15].

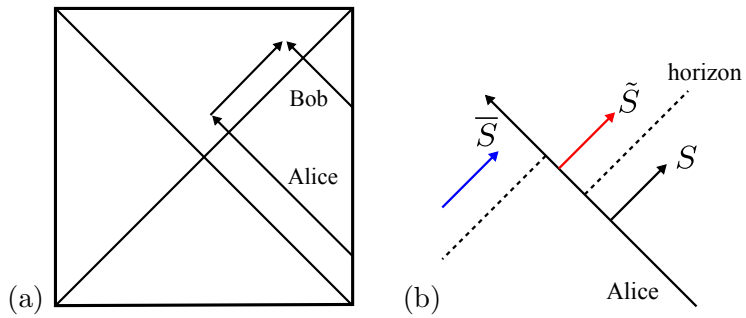


Figure 3: (a) A setup by Hayden and Preskill. Alice sends a signal to Bob. (b) A refinement of the Hayden-Preskill argument. Alice's signal does not reach Bob due to backreaction.

decoupled degrees of freedom deep in the bulk, it is natural to expect that high complexity quantum operations are needed on boundary CFTs [10].

Next let us turn to BFV's protocol. Here we shall focus on a certain subroutine of their protocol. BFV's protocol requires a pair of observers who send signals to each other inside the black hole. Since two observers from opposite sides cannot communicate as discussed above, they should come from the same side of the black hole. Let us imagine that one observer (Alice) jumps into the black hole, and then the other observer (Bob) jumps later. Upon crossing the horizon, Alice sends a signal to Bob. Can they communicate? Hayden and Preskill asked exactly the same question in [17]. Their argument goes as follows. If the time separation Δt between Alice and Bob is longer than the scrambling time, Alice needs large amount of energy, possibly larger than the Planck energy, to send a signal to Bob. Otherwise, Bob will reach singularity before Alice's message arrives. Hence two observers will not be able to communicate with each other unless they introduce drastic backreaction to the underlying geometry. This suggests that there seems to be a fundamental upper bound on efficiently measurable volume or length near the horizon.

Here we would like to present a modern perspective on the Hayden-Preskill's observation by considering the effect of backreaction by Alice. Suppose that Alice was planning to send a signal by using the mode $\overline{S}$ behind the horizon, which is the partner mode of S (Fig. 3(b)). Once Alice crosses the horizon, a new partner mode $\tilde{S}$ is dynamically created. Hence Alice cannot send a signal to Bob by using the original partner mode $\overline{S}$. In order to avoid backreaction, Alice would need to use the interior mode $\overline{S}$ which is not close to the horizon. However this means that the time separation Δt between Alice and Bob is significantly shorter than the scrambling time.

While we have arrived at the same conclusion as Hayden and Preskill's, there is a subtle (but

[10]By applying certain quantum operations on the boundary, it is possible to cancel backreaction from bulk observers. Indeed, the traversable wormhole phenomena and the Hayden-Preskill decoding protocol are examples of such operations. In the language of quantum error-correction, a naive bulk description, which mistakingly ignore backreaction, is valid only inside some codeword subspace. Backreaction by bulk observers can be viewed as "errors" to a quantum error-correcting code which brings the system outside the codeword subspace. By performing suitable quantum error-correction which cancels backreaction, one can keep the system inside the codeword subspace where the bulk effective description remains valid.

important!) refinement in our argument. The reason why Alice cannot communicate with Bob is because Alice has intersected with the mode S, and as a result, $\overline{S}$ is disentangled from S. Namely, the disentangling phenomena occurs automatically, regardless of whether Alice attempts to send a signal to Bob with trans-Planckian energy or not.

Finally, we critically comment on Susskind's proposal for a resolution of the complexity puzzle [2]. Susskind proposed that the qECT thesis may be violated behind the black hole horizon, but the violation should not be communicated to the outside efficiently. This proposal is in accord with the black hole complementarity; one chooses not to be bothered by the violation of the qECT thesis if the violation cannot be ever found out. As we have already demonstrated in the main body of this note, backreaction from infalling observers invalidate any shortcuts to measuring the volume of the wormhole. Hence we do not need to invoke the black hole complementarity to save the qECT thesis and resolve the puzzle. The effect of gravitational backreaction prevents the violation from happening. It is wrong to say that the violation is invisible; the violation simply did not happen.

Another counterargument can be obtained by actually saving the infalling observer from a black hole. Recall that an infalling observer in a two-sided black hole can be saved from the black hole by quantum operations which act solely on one side of the black hole in principle [11]. The quantum circuit complexity of quantum operations is independent of the complexity of the initial state (*i.e.* the wormhole volume $\mathcal{V}$ of the two-sided black hole). Hence the infalling observer can report her experience after performing quantum operations with a complexity much smaller than $\mathcal{V}$. This does not invalidate the qECT thesis; the rescued observer will not give us any useful information on the wormhole volume $\mathcal{V}$ or $\frac{d\mathcal{V}}{dt}$ as she was not hit by the shockwave.

6 Discussions

We have presented a possible resolution of the black hole complexity puzzle by considering the effect of gravitational backreaction by an infalling observer. Our observation suggests that there is no shortcut in measuring the volume of the wormhole, and thus the qECT thesis remains valid . We also revisited the thought experiment by Hayden and Preskill who asked whether two infalling observers from the same side of a black hole can communicate with each other or not. Namely we presented a refinement of their argument (but with the same conclusion) by examining gravitational backreaction without invoking the black hole complementarity.

A similar puzzle has been recently pointed out concerning the measurability of entanglement entropy [18]. Perhaps appropriate considerations of gravitational backreaction may provide a resolution of this puzzle [11].

Our results also provide a resolution on the unboundedness puzzle concerning the diverging Hilbert space dimension of the black hole interior and the apparent violation of the area law in the maximal

[11] [19] presents a certain argument concerning the measurability of entanglement entropy.

volume slice. Namely, our argument suggests that volume-law degrees of freedom can be regulated to give an area-law entropy in a dynamical manner. We speculate that a similar reasoning may resolve the puzzle concerning Wheeler's "bag of gold" spacetime. It is an interesting future problem to make this perspective more quantitative. We also believe that proper understanding of such phenomena will be useful in better understanding the physics of de Sitter space. Relatedly, it will be interesting to ask if the qECT remains valid for quantum gravity in de Sitter space.

Recently Herman Verlinde proposed an interesting argument on a possible loophole in the ER = EPR slogan [20]. The idea is that bulk effective theories emerge after projecting the system into a certain codeword subspace, and the preexisting quantum entanglement between two sides, as well as the complexity, is not essential for the wormhole geometry. We believe that our treatment of backreaction by infalling observers provides explicit construction of the projection operator in a dynamical manner. It is an interesting future problem to understand dynamical emergence of bulk effective theories.

BFV demonstrated the complexity puzzle in a slightly different manner by considering the complexity of the bulk-boundary dictionary and arrived at the following conclusion; 1) the qECT thesis is violated or 2) the complexity of the holographic dictionary is high. In our argument, this no-go result has been avoided by the fact that the dictionary is observer-dependent and thus is not invariant. Indeed, the expression of the interior partner mode depends on the initial state of the infalling observer.

6.1 Extended firewall puzzle

Motivated by the black hole complexity puzzle, we would like to point out another interesting challenge to the qECT thesis in the context of the firewall puzzle. The challenge concerns a certain puzzle on the decoding complexity of interior partner modes. We view this as an extension of the firewall puzzle from the computational complexity perspective.

The gravitational backreaction by an infalling observer provides a potential resolution of the firewall puzzle since the outgoing mode is decoupled from the early radiation due to scrambling dynamics of a black hole. As such, the fact that the infalling observer can cross a smooth horizon and see interior partner modes behind it does not lead to any inconsistency from the perspective of the monogamy of quantum entanglement [11]. From the computational complexity perspective, however, we find some trouble in this scenario. From the boundary viewpoint, finding the expression of interior partner operators appears to be computationally difficult especially when the Hilbert space size accounting for interior modes is large and OTOCs have decayed to asymptotic values at late times. The best known algorithm for interior reconstruction at late times uses the Grover search algorithm whose runtime is proportional to the Hilbert space size of the interior modes (hence is exponential in the number of qubits) [13] [12]. Yet, the infalling observer can see the interior Hawking modes by simply jumping into

[12]The problem of reconstructing partner operators can be interpreted as the Hayden-Preskill problem (or the information loss problem) running backwards in time. This is the reason why one can utilize the Hayden-Preskill recovery protocol to

a black hole and crossing the smooth horizon without doing any difficult quantum computation! This leads to an apparent violation of the qECT thesis.

Below we propose a possible resolution of the extended firewall puzzle. When the time separation Δt between the infalling observer and the outgoing mode is longer than the scrambling time $t_{\rm scr}$, the infalling observer will encounter the interior partner degrees of freedom near the horizon, possibly at the Planck length distance (or less) from the horizon. Such interior modes at short distance scale should not be visible to infalling observer since these would be realized as a part of the geometry itself, instead of matter fields propagating freely on the geometry, due to some short distance quantum gravity effect. Hence it is reasonable to assume that interior partners at short distance scales are computationally difficult to decode for the infalling observer too. On the other hand, when the time separation Δt is shorter than $t_{\rm scr}$, one may utilize decoding protocols based on traversable wormhole effects to efficiently construct interior partner operators as in [16]. These protocols utilize the fact that black holes are the fastest scrambler with the Lyapunov exponent $\frac{2\pi}{\beta}$ and work reliably only when a black hole scramble quantum information coherently until $\Delta t \leq t_{\rm scr}$. In the bulk, this corresponds to the fact that the infalling observer will encounter these interior partner modes away from the horizon, and can observe them easily. Hence, the qECT thesis will remain valid.

A resolution of the original firewall puzzle, in a sense of the monogamy of quantum entanglement, required the fact that a black hole scrambles quantum information. This resolution, however, is unsatisfactory as it did not utilize the fact that a black hole scrambles quantum information in the fastest possible manner. After all, a piece of burning wood will scramble quantum information at late times too! The extended firewall puzzle, strengthened by the qECT thesis and the computational complexity, is intrinsic to black holes and naturally calls for the very fact that a black hole is the fastest scrambler. Namely, the separation of the computational complexity in reconstructing interior operators before and after the scrambling time results from the fast scrambling nature of the black hole. On the bulk, these correspond to different physical objects; the matter and the geometry.

Acknowledgment

I thank Adam Bouland and Nick Hunter-Jones for discussions (v1). I thank Hrant Gharibyan, Junyu Liu, Geoff Penington and Douglas Stanford for discussions (v2). Research at the Perimeter Institute is supported by the Government of Canada through Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade.

construct interior operators. See [14] for details.

A Miscellaneous matters

In this appendix, we provide justifications of two assumptions in the decoupling theorem. We also present answers to some frequently asked questions.

We claimed that two signals from opposite sides of a black hole will not encounter inside the black hole as long as two sides are not coupled. From the bulk perspective, this was because naive predictions from bulk effective descriptions break down due to gravitational backreaction from an infalling observer. The crux of this argument was that a new partner mode $\widetilde{D}$ is dynamically generated as a result of scrambling dynamics of a black hole in an observer-dependent manner.

This decoupling mechanism relies on two assumptions; 1) the time separation Δt between an observer A and the outgoing mode D satisfies $\Delta t \geq t_{\rm scr}$ and 2) $d_A \gtrapprox d_D$ where d_A, d_D are Hilbert space sizes of A, D. Below we discuss the cases which do not satisfy these assumptions [13].

Let us begin with the cases with $\Delta t \leq t_{\rm scr}$. We present three observations which support our conclusions.

a) *Singularity:* If the time separation Δt is much shorter than the scrambling time $\Delta t \ll t_{\rm scr}$, the trajectories of A and R_D will intersect near the black hole singularity. It is then not surprising if bulk effective descriptions become invalid in the proximity of the singularity. This observation, however, does not provide a satisfactory justification for intermediate time scales $t_{\rm th} \ll \Delta t \lessapprox t_{\rm scr}$ where $t_{\rm th}$ is the thermalization time.

b) *Entanglement quality:* If $\Delta t \lessapprox t_{\rm scr}$, an infalling observer will interact with the outgoing mode D away from the horizon. The proper temperature near the Rindler horizon is given by $T = \frac{1}{2\pi\rho}$ where ρ is the proper distance from the horizon. This suggests that, as one moves away from the horizon, the density of thermal entropy becomes small and the quality of quantum entanglement, which an infalling observer may be able to detect via free fall, may deteriorate. Hence, there is no serious inconsistency with our conclusions even if the decoupling phenomena is weak.

c) *Measurement strength:* Recall that the decoupling phenomena occurs as a result of the interaction between the observer A and the outgoing mode D. Its strength is quantified by the decay of OTOCs. This suggests that the decoupling, as well as creation of a new partner mode $\widetilde{D}$, occurs as much as the observer A measures D. Then, even if D and $\widetilde{D}$ are not perfectly entangled, there is no contradiction with the monogamy of entanglement since A is observing D (and its partner $\widetilde{D}$) only weakly.

   If A wishes to measure D more directly or strongly, she may increase the size of herself (d_A) to enhance gravitational scattering or perform actual physical measurement on D upon encountering

[13] These discussions are copied from section 6.4 in v2 of [12] with minor modifications.

it. We expect that such strong interactions or direct measurements will lead to significant decay of OTOCs and create $\widetilde{D}$ which is nearly perfectly entangled with D.

Next, let us justify the condition of $d_A \gtrapprox d_D$ by presenting three observations.

d) *AMPS:* For an application to the firewall puzzle, namely the AMPS thought experiment, we can justify the requirement. In the thought experiment, the outside observer distills R_D and hand it to the infalling observer A who jumps into the black hole to verify the entanglement between D and $\widetilde{D}$. This effectively realizes a situation with $d_A = d_D$. See [12] for details.

e) *Metaphysical explanation:* In order to experience some physics, the infalling observer A herself should carry some amount of entropies, at least as much as the objects D she is going to measure. While being metaphysical, this explanation can be further justified from the aforementioned observation c) which suggests the decoupling occurs as much as the observer A measures D.

f) *QFT:* One may interpret A and D as infalling and outgoing modes of low energy QFT on the LHS wedge respectively while B and C can be viewed as all the other high-energy degrees of freedom associated with the future and past horizons respectively. It is thus natural to assume $d_A = d_D$ in a non-evaporating black hole as in the AdS space. (For an evaporating black hole, one can model the dynamics by taking d_D to be slightly larger than d_A). Here A and D are DOFs which an outside observer can easily see.

   On the other hand, the infalling observer, who travels nearly at the speed of light, will propagate along the infalling mode A. Hence it is natural to expect that the initial state of A herself becomes irrelevant as long as the initial state is chosen from typical states in the low energy subspace of the QFT. Namely, regardless of the initial state of A, she should be able to see the same bulk effective QFT description near the horizon. This is indeed the case as the creation of $\tilde{D}$ occurs for any initial state of A. However, it is worth emphasizing that the map from the bulk QFT to the boundary is dependent on the initial state of A. Here the infalling observer will have an access to the outdoing mode D and the interior mode $\widetilde{D}$ instead of A.

Let us move to a different topic. There seem to be a significant number of researchers who currently believe that two signals from opposite sides should encounter inside the black hole. It is hence worth summarizing evidences against such scenario. While we do not claim to disprove the encountering scenario, we believe that these evidences are sufficiently strong to force us to revisit some of previous popular proposals, such as [21], in a critical manner.

1) *Non-locality:* The first, and the most severe issue, is that two sides on the boundary are not coupled. Hence, an encounter inside the black hole would lead to rather strange and puzzling non-locality

inside the black hole. Indeed, this apparent non-locality is the origin of various conceptual puzzles, such as the firewall puzzle and the complexity puzzle.

There have been a number of arguments which essentially try to say that such non-locality remains inside the black hole and hidden from the outside. However, it is more natural to make every attempt to avoid relying on such non-locality. Also, whether an infalling observer will be hit by a signal from the other side or not is an actual physical question, and its apparent non-measurability from the outside does not provide strong support of such scenario.

Thus, we are investigating a less troublesome resolution that two signals will not encounter inside the black hole by using the boundary quantum mechanics as a guiding principle.

2) *Rescuing Alice:* The second counterargument, which has already been mentioned in the main text, can be obtained by actually saving the infalling observer from a black hole and asking her if she was hit by a signal from the opposite side or not. An infalling observer in a two-sided black hole can be saved from the black hole by quantum operations which act solely on one side of the black hole in principle [11]. We believe that the rescued observer will not be hit by the shockwave as the whole process does not depend on the other side of the black hole at all.

3) *Partner mode:* In the scenario which favours the encounter inside the black hole, it is often argued that touching R_D will create a deadly shockwave which kills A, preventing her from seeing the interior partner mode. In the light of this explanation, A should be able to cross the horizon with no drama when R_D remains unperturbed. However, the new entangled partner mode $\widetilde{D}$ emerges regardless of whether R_D is touched or not. Let us assume that R_D was not touched. If one insists that A sees R_D, instead of $\widetilde{D}$, then there should be a firewall even if R_D was not perturbed, as QFT is not in the vacuum state with D and R_D not being entangled. This suggests that such scenario should be rejected on the ground that it does not provide a resolution of the firewall puzzle, at least when the infalling observer A carries a finite entropy or energy.

Finally we would like to make a few comments on the bulk interpretation of the decoupling phenomena. By treating the infalling observer as a gravitational shockwave, we drew a backreacted Penrose diagram as in Fig. 2(b), to characterize the emergence of a new partner mode. One subtlety of this interpretation is that we need to think that the observer A is located just "above" the shockwave so that she encounters $\widetilde{D}$ near the horizon and R_D near the singularity. Then, it is consistent with our conclusion that A will actually not encounter R_D as this encounter would happen near the singularity. Here, one may think of the infalling observer A throwing a few qubits beforehand, guarding herself from R_D and jumps into the black hole.

On the other hand, if we think that Alice is located just "below" the shockwave, A will encounter R_D instead of $\widetilde{D}$ [14]. It is not easy to tell which interpretation is the correct one by naively staring at

[14]It is worth recalling again that the interior reconstruction is identical to the information recovery problem running backwards in time.

the backreacted Penrose diagram. One can pick the correct interpretation in an unambiguous manner by using predictions from the boundary as a guiding principle.

One might think of evolving $\widetilde{D}$ backward to the past and might hope to conclude that $\widetilde{D}$ was indeed a mode on the LHS degrees of freedom. This observation might appear to give a reconstruction of $\widetilde{D}$ on the LHS degrees of freedom via standard techniques by solving wave equations on the bulk. However, this naive expectation seems to have one serious problem. From the boundary analysis, we already know that the construction of $\widetilde{D}$ depends on the initial state of A whereas the above prescription does not have any dependence on A. The correct prescription would require some non-trivial processing of the mode $\widetilde{D}$ or non-trivial change of the holographic dictionary upon crossing the shockwave.

References

[1] A. Bouland, B. Fefferman, and U. Vazirani, "Computational pseudorandomness, the wormhole growth paradox, and constraints on the ads/cft duality," `arXiv:1910.14646`.

[2] L. Susskind, "Horizons protect church-turing," `arXiv:2003.01807`.

[3] D. Deutsch, "Quantum theory, the church–turing principle and the universal quantum computer," *Proc. R. Soc. Lond. A* **400** (1985) 97–117.

[4] D. Stanford and L. Susskind, "Complexity and shock wave geometries," *Phys. Rev. D* **90** (2014) 126007.

[5] S. Aaronson, "The complexity of quantum states and transformations: From quantum money to black holes," `arXiv:1607.05256`.

[6] D. A. Roberts and B. Yoshida, "Chaos and complexity by design," *JHEP* **4** (2017) 121.

[7] F. G. Brandão, W. Chemissany, N. Hunter-Jones, R. Kueng, and J. Preskill, "Models of quantum complexity growth," `arXiv:1912.04297`.

[8] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, "Holographic quantum error-correcting codes: toy models for the bulk/boundary correspondence," *JHEP* **06** (2015) 149.

[9] J. Cotler, N. Hunter-Jones, J. Liu, and B. Yoshida, "Chaos, complexity, and random matrices," *JHEP* **11** (2017) 48.

[10] J. S. Cotler, G. Gur-Ari, M. Hanada, J. Polchinski, P. Saad, S. H. Shenker, D. Stanford, A. Streicher, and M. Tezuka, "Black holes and random matrices," *JHEP* **5** (2017) 118.

[11] B. Yoshida, "Firewalls vs. scrambling," *JHEP* **10** (2019) 132.

[12] B. Yoshida, "Observer-dependent black hole interior from operator collision," `arXiv:1910.11346`.

[13] B. Yoshida and A. Kitaev, "Efficient decoding for the hayden-preskill protocol," arXiv:1710.03363.

[14] B. Yoshida, "Soft mode and interior operator in the hayden-preskill thought experiment," *Phys. Rev. D* **100** (2019) 086001–.

[15] D. Marolf and A. C. Wall, "Eternal black holes and superselection in ads/cft," *Class. Quant. Grav.* **30** (2012) 025001.

[16] P. Gao, D. L. Jafferis, and A. C. Wall, "Traversable wormholes via a double trace deformation," *JHEP* **12** (2017) 151.

[17] P. Hayden and J. Preskill, "Black holes as mirrors: quantum information in random subsystems," *JHEP* **09** (2007) 120.

[18] A. Gheorghiu and M. J. Hoban, "Estimating the entropy of shallow circuit outputs is hard," arXiv:2002.12814.

[19] N. Bao, J. Pollack, and G. N. Remmen, "Wormhole and entanglement (non-)detection in the er=epr correspondence," *JHEP* **11** (2015) 126.

[20] H. Verlinde, "ER = EPR revisited: On the entropy of an einstein-rosen bridge," arXiv:2003.13117.

[21] J. Maldacena and L. Susskind, "Cool horizons for entangled black holes," *Fortsch. Phys.* **61** (2013) 781–811.