

Firewalls vs. scrambling

Beni Yoshida

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

E-mail: byoshida@perimeterinstitute.ca

ABSTRACT: Recently we pointed out that the black hole interior operators can be reconstructed by using the Hayden-Preskill recovery protocols. Building on this observation, we propose a resolution of the firewall problem by presenting a state-independent reconstruction of interior operators. Our construction avoids the non-locality problem which plagued the “ $A = R_B$ ” or “ER = EPR” proposals. We show that the gravitational backreaction by the infalling observer, who simply falls into a black hole, disentangles the outgoing mode from the early radiation. The infalling observer crosses the horizon smoothly and sees quantum entanglement between the outgoing mode and the interior mode which is distinct from the originally entangled qubit in the early radiation. Namely, quantum operation on the early radiation cannot influence the experience of the infalling observer since description of the interior mode does not involve the early radiation at all. We also argue that verification of quantum entanglement by the outside observer does not create a firewall. Instead it will perform the Hayden-Preskill recovery which saves an infalling observer from crossing the horizon.

KEYWORDS: AdS-CFT Correspondence, Black Holes, Holography and condensed matter physics (AdS/CMT)

ARXIV EPRINT: [1902.09763](https://arxiv.org/abs/1902.09763)

Contents

1	Introduction	1
2	Review of firewall argument	4
2.1	From the outside	4
2.2	From the inside	5
2.3	Interior operators	6
3	Review of Hayden-Preskill recovery	7
3.1	Setup	8
3.2	State representation of U	8
3.3	Reconstruction of operators	9
3.4	Quantum cloning puzzle	10
4	Scrambling and recovery	11
4.1	Scrambling	11
4.2	OTOCs in a black hole	13
4.3	Recovery protocols	14
5	Reconstruction of interior operators	15
5.1	Hayden-Preskill for interior operators	16
5.2	Young black hole	17
5.3	Monogamy and scrambling	20
6	State-independent interior operators	20
6.1	Long-throat AdS black hole	21
6.2	State-independence	22
6.3	Code subspace	24
7	Backreaction by infalling observer	25
7.1	Including Alice	25
7.2	Sending probes	27
7.3	Geometric interpretation	29
8	On definition of state-independence	30
8.1	No-go arguments	30
8.2	Observer-dependence	31
9	Resolution of AMPS puzzle	32
9.1	AMPS thought experiment	32
9.2	Revisiting firewalls	33
9.3	Puzzle on non-local interactions (and its resolution)	35

10 Discussions	36
10.1 Early time	36
10.2 Peeking in through the horizon	37
10.3 Entanglement wedge reconstruction	38
A Energy measurement	38
B Verlinde-Verlinde proposal	40
C Fault-tolerant entanglement	41
D Alice’s quantum field theory	43

1 Introduction

The question of how a quantum black hole forms and evaporates unitarily remains among the most puzzling mysteries in theoretical physics [1]. Following earlier works, Almheiri, Marolf, Polchinski and Sully (AMPS) argued that by careful quantum processing of the early Hawking radiation, one can distill a qubit that is perfectly entangled with an outgoing particle from the black hole [2]. According to Hawking’s semiclassical calculation [3], however, that same particle is entangled with some field mode inside the black hole, leading to a violation of quantum mechanics, namely the monogamy of quantum entanglement. This contradiction, called the AMPS puzzle, sharply illustrates the fundamental difficulty in describing the interior of a quantum black hole.

The AMPS problem becomes most precise for the outside observer where the difficulty is constructing operators for interior modes. This viewpoint led to approaches bundled under “ $A = R_B$ ” or “ER = EPR” where the interior operators are reconstructed in the early radiation instead of the remaining black hole [4–8]. Unfortunately this proposal runs into various paradoxes [9–11]. (See [4, 12] for summaries). The reconstruction may lead to apparent non-local interaction between the interior and distant radiation.¹ Relatedly, the interior operators seem to be sensitive to perturbations added to the early radiation. Also, previous proposals construct interior operators in a *state-dependent* manner that suffers from a number of potential inconsistencies with quantum mechanics [13, 14]. To overcome these difficulties, further proposals have been put forward. A concrete state-dependent proposal was made in the context of the AdS/CFT correspondence [5, 15, 16] and the Sachdev-Ye-Kitaev (SYK) model [17, 18], but these proposals have some caveats [19].

Preliminary studies suggest that the idea of quantum error-correction [20, 21] plays key roles in avoiding some of the identified problems [18, 22–25]. The ultimate resolution of the

¹The fact that some description of interior operator requires the early radiation does not necessarily mean that information can be sent non-locally. However, if every possible reconstruction of the interior operators requires the early radiation, then it is problematic as it implies that the infalling observer may interact non-locally with the early radiation by seeing interior modes.

AMPS puzzle, however, would be obtained only by finding an explicit way of reconstructing interior operators which does not suffer from the aforementioned problems.

Recently we pointed out that the procedure of reconstructing the interior operators can be viewed as the Hayden-Preskill recovery [26], a phenomenon of reconstructing an infalling quantum state from an old black hole by collecting the Hawking radiation [27]. Input quantum states can be recovered when the time-evolution of the system scrambles quantum information in a sense of decay of out-of-time ordered correlations [28]. A quantum black hole exhibits such dynamics in the most dramatic manner [29]. In parallel with the conclusion by Hayden and Preskill, this observation suggests that the interior operators can be constructed “almost” inside the remaining black hole with a few extra qubits from the early radiation. Also, the reason why interior operators are insensitive to perturbations on the early radiation can be explained by scrambling dynamics without relying on speculative quantum circuit complexity arguments [6, 30]. However, our observation in [26] still suffers from the non-locality problem and state-dependence.

In this paper, we expand on the aforementioned observation and present a protocol to reconstruct interior operators in a *state-independent* way while at the same time avoiding the non-locality problem. In particular, we show that quantum entanglement between the outgoing mode and the reconstructed interior mode is insensitive to any perturbation on the early radiation. In fact, we argue that our construction of state-independent interior operators is applicable to generic black holes, young or old, in quasi thermal equilibrium. We will explain how our construction *avoids* previous arguments which appear to suggest that interior operators must be state-dependent.² Namely, we will find that construction of interior operators does not depend on the initial state of the black hole (i.e. how it is entangled with the early radiation), but *depend* on the infalling observer.

Based on the state-independent reconstruction, we suggest the following resolution of the AMPS puzzle. Let us distill a qubit \tilde{D} from the early radiation R that is entangled with the outgoing mode D . We show that an infalling observer leaves non-trivial gravitational backreaction which disentangles the outgoing mode D from the distilled qubit \tilde{D} *no matter how she falls into a black hole*. The outgoing mode D will be entangled with the interior mode \bar{D} which is *distinct* from the originally entangled qubit \tilde{D} . The infalling observer crosses the horizon smoothly and can observe entangled modes D and \bar{D} . Namely, any quantum operation on the early radiation R *cannot* influence the experience of the infalling observer since the construction of the interior mode \bar{D} does not involve any degrees of freedom on R . The disentangling phenomenon is a universal feature of a quantum black hole which can be studied quantitatively by using out-of-time order correlation functions.

The aforementioned proposal provides an account for a physical mechanism where the infalling observer sees quantum entanglement while the outside observer does not. It is intriguing to consider an opposite limit of the monogamy relation where the outside observer can see quantum entanglement while the infalling observer does not. Previous works, especially the one by Maldacena and Susskind [6], suggested that verification of quantum entanglement, namely the distillation of the entangled qubit \tilde{D} on the early

²I thank Ahmed Almheiri and Suvrat Raju for useful discussions on this.

radiation R , generates a high-energy gravitational shockwave which would become a firewall and prevent the infalling observer from crossing the horizon smoothly. This proposal, however, can be refuted by observing that the interior mode \bar{D} is distinct from the distilled qubit \tilde{D} and its construction does not involve the early radiation R at all. Here, we argue that the verification of quantum entanglement will generate “safety barrier” which stops the infalling observer from falling into a black hole. The key observation is that the outside observer can retrieve quantum entanglement back from the infalling observer by performing the Hayden-Preskill recovery on the infalling observer. Since the infalling observer does not cross the horizon, she will not be able to see quantum entanglement.³ This proposal resonates with the recovery protocols for the Hayden-Preskill thought experiment proposed by the author and Kitaev [31]. In fact, these recovery protocols proceed by protecting quantum entanglement between the outgoing mode D and the early radiation R from the backreaction by the infalling observer.

While we rely on intuitions from the AdS/CFT correspondence, our arguments are applicable to generic quantum black holes. There are two main assumptions behind our proposals. First, the outside observer has a quantum description of a black hole with a pure quantum state on his Hilbert space. Second, out-of-time ordered correlation functions decay around/after the scrambling time. The proposed state-independent construction of interior operators work for generic quantum black holes that satisfy these two assumptions. Much of our argument, however, is based on simple quantum information theoretic description of a quantum black hole. While detailed calculations on realistic black holes are beyond the scope of the paper, we will try to be clear about our assumptions behind the use of simplified toy models.⁴

The rest of the paper is organized as follows. In section 2, we briefly review the firewall argument. In section 3, we briefly review the Hayden-Preskill recovery and the quantum cloning puzzle. In section 4, we discuss the notion of quantum information scrambling and its relation to out-of-time ordered correlation functions. In section 5, we show that the interior operators can be reconstructed by using the Hayden-Preskill recovery protocol. In section 6, we present a state-independent reconstruction of the interior operators. In section 7, we discuss the effect of backreaction by the infalling observer. In section 9, we propose the resolution of the AMPS puzzle. In section 10, we conclude with discussions. In appendix A, we show that measurement of energy density suffices to entangle the infalling observer with the outgoing radiation. In appendix B, we make comments on the proposal by Verlinde and Verlinde [22]. In appendix 8, we discuss how our construction avoids previous arguments for the necessity of state-dependence. In appendix C, we discuss the state-independent reconstruction from the perspective of dynamics in quantum many-body systems. In appendix D, we discuss geometric interpretations of the Hayden-Preskill recovery protocols.

³Here the infalling observer should be thought of as some light probe particle. If it were a macroscopic object with large energy, the outside observer wouldn't be able to save it by low quantum complexity operations. Also the amount of entanglement the outside observer needs to verify depends on the entropy associated with the infalling observer. See [31] for details.

⁴Our attitude is to treat a black hole as a quantum system and try to explain the AMPS puzzle without giving up with unitarity, smoothness (equivalence principle) or locality.

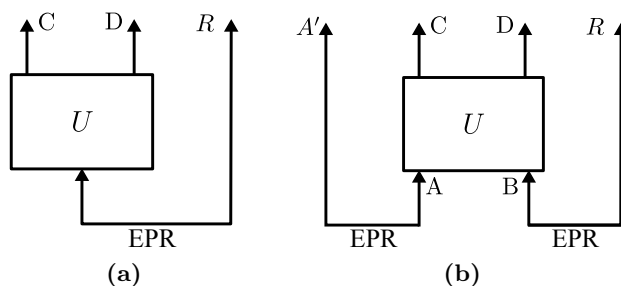


Figure 1. The notation used in the paper. Here U represents the unitary time-evolution of the black hole. (a) The AMPS thought experiment. C : the remaining black hole. D : the outgoing mode. R : the early radiation. (b) The Hayden-Preskill thought experiment. A' : the reference. A : the input quantum state. B : the initial black hole. R : the early radiation. C : the remaining black hole. D : the outgoing mode. Horizontal lines represent that two subsystems are in the EPR pairs: $\frac{1}{\sqrt{d}} \sum_{j=1}^d |j\rangle \otimes |j\rangle$.

2 Review of firewall argument

We begin with a brief overview of the original argument of AMPS [2]. The argument has a number of technical assumptions and fine details, and we will not cover all of them. Our goal is to remind readers of the form of quantum entanglement and the associated Hilbert space structure postulated by AMPS, as well as to introduce some notation. We also summarize some of the proposed resolutions of the puzzle, and potential difficulties and concerns that may arise from them. Our goal is to motivate some of the desired properties in reconstructing the interior operators.

The notation in the present paper is slightly non-standard. For convenience, our notation is summarized in figure 1. Throughout the paper, we will denote the infalling and outside observers by Alice and Bob respectively.

2.1 From the outside

We first discuss the description of a quantum black hole from the perspective of Bob, the outside observer.

The AMPS thought experiment relies on the following assumption on the Hilbert space structure. In Bob’s description, the formation and evaporation of the black hole is described by a unitary operator. Namely, there is a pure quantum state $|\Psi\rangle$ in the outside Hilbert space $\mathcal{H}_{\text{outside}}$ at any given time. The Hilbert space $\mathcal{H}_{\text{outside}}$ can be factorized into (figure 2a)

$$\mathcal{H}_{\text{outside}} = \mathcal{H}_C \otimes \mathcal{H}_D \otimes \mathcal{H}_R. \tag{2.1}$$

Here \mathcal{H}_R consists of the radiation field outside of the black hole with roughly $r > 3GM$ while \mathcal{H}_D are the modes confined in $2GM + \epsilon < r < 3GM$. Here $r = 2GM + \epsilon$ corresponds to the location of the stretched horizon, and ϵ is of order the Planck length. This region is often called the zone. It is common to restrict \mathcal{H}_D to include only modes with Schwarzschild energy less than the black hole temperature since higher energy modes are not confined.

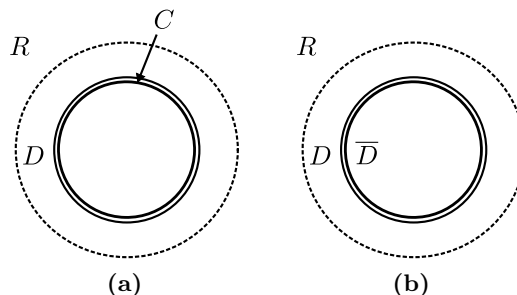


Figure 2. The Hilbert space structures. The dotted line represents $r = 3GM$ and the double lines represent $r = 2GM \pm \epsilon$. Roughly, D corresponds to the zone. (a) The outside Hilbert space $\mathcal{H}_{\text{outside}}$. (b) The infalling Hilbert space $\mathcal{H}_{\text{inside}}$.

Finally, \mathcal{H}_C are the remaining degrees of freedom which can be interpreted as some entities sitting at the stretched horizon at $r = 2GM + \epsilon$. It is common to restrict \mathcal{H}_R to be a finite dimensional Hilbert space for convenience of discussions. Precise distinctions among these subsystems are not essential for the AMPS puzzle.

Following AMPS, we consider an old black hole that has already emitted more than half of its initial entropy: $|R| \gg |C||D|$. If $|\Psi\rangle$ is a Haar random quantum state in $\mathcal{H}_{\text{outside}}$ [32, 33], the density operator in CD is approximately proportional to the identity operator:

$$\rho_{CD} \approx \frac{1}{|C||D|} I_C \otimes I_D. \quad (2.2)$$

This implies D is maximally entangled with R . Of course, approximating $|\Psi\rangle$ with a Haar random state is an oversimplification. A more careful analysis suggests that thermal correlations are present in ρ_{CD} [22] although these are purely classical ones resulting from the conservation of the energy [26]. For the time being, we proceed with eq. (2.2). We emphasize that this detail does not matter for discussions except appendix A.

2.2 From the inside

Next we discuss the description from the perspective of Alice, the infalling observer.

Following AMPS, we assume that Alice has quantum mechanical description of her experiences in terms of a quantum state on her time slice. Namely, the Hilbert space $\mathcal{H}_{\text{outside}}$ can be factorized into (figure 2b)

$$\mathcal{H}_{\text{inside}} = \mathcal{H}_{\bar{D}} \otimes \mathcal{H}_D \otimes \mathcal{H}_R \otimes \mathcal{H}_S. \quad (2.3)$$

Here \mathcal{H}_D and \mathcal{H}_R correspond to the confined modes (the zone) and the radiation fields respectively as defined above. These are subsystems which both Alice and Bob can access. $\mathcal{H}_{\bar{D}}$ are the modes just inside the horizon roughly with $GM < r < 2GM - \epsilon$. Finally, \mathcal{H}_S is all the remaining degrees of freedom that is not accessible to Alice. The modes at the stretched horizon \mathcal{H}_C are often included in \mathcal{H}_S , but this detail does not matter in the

AMPS argument. Finally, we postulate that the quantum description of ρ_{DR} for Alice and Bob are identical.⁵

The AMPS puzzle concerns an apparent inconsistency between Alice and Bob’s descriptions. In Alice’s description, $\mathcal{H}_{\bar{D}}$ and \mathcal{H}_D contain Rindler modes which are close to maximally entangled:

$$\propto \sum_n e^{-\frac{\beta E_n}{2}} |n\rangle_{\bar{D}} \otimes |n\rangle_D. \tag{2.4}$$

Alice’s Hilbert space may contain some degrees of freedom other than Rindler modes. The important point for us is that two modes in $\mathcal{H}_{\bar{D}}$ and \mathcal{H}_D must be entangled. On the other hand, in Bob’s description, D is maximally entangled with the radiation R , which would imply that D and \bar{D} are not entangled with each other due to the monogamy of entanglement. In particular, by using the strong subadditivity, AMPS, as well as earlier works [34, 35], arrived at $I(D, \bar{D}) \equiv S_D + S_{\bar{D}} - S_{D\bar{D}} \approx 0$. This suggests $\rho_{\bar{D}D} = \rho_{\bar{D}} \otimes \rho_D$, the existence of a firewall.^{6,7}

2.3 Interior operators

The AMPS puzzle can be also stated in the outside observer’s description. This viewpoint becomes most precise in the AdS/CFT correspondence where R is interpreted as degrees of freedom on the other side of a large two-sided AdS black hole. In Bob’s description, or in the boundary CFT, the puzzle can be identified with the fact that $I(C, D) = 0$. This suggests that the interior mode in \bar{D} cannot be described as an operator inside \mathcal{H}_C . Namely, its description must involve the radiation R . This viewpoint is often advocated under the slogan “ $\bar{D} = R$ ” (or “ $A = R_B$ ” in accordance with the standard notation in the literature) and also “ER = EPR”. This proposal, however, runs into various paradoxes.

First, it may lead to non-local signalling from the interior of the black hole to the radiation R (and vice versa). To be concrete, one may consider a large AdS black hole with very long wormhole throat. The description of the interior mode, which an infalling observer would be able to measure, *always* requires degrees of freedom from a far distant boundary of the black hole, which sounds highly unphysical. The non-local signalling problem is referred to as the commutator problem since operators for the interior mode \bar{D} do not commute with operators in the radiation R . To avoid the non-local signalling problem, the reconstruction of the interior operator in the outside should be insensitive to perturbations

⁵By “identical”, we mean that $\rho_{DR}^{(\text{Alice})} = \rho_{DR}^{(\text{Bob})}$. It may be possible that $\rho_{DR}^{(\text{Alice})}$ and $\rho_{DR}^{(\text{Bob})}$ differ up to local unitary transformations with factorized form $U_D \otimes U_R$. The key assumption for us is that non-local properties such as quantum entanglement between D and R are identical for two observers.

⁶More precisely, this suggests a strong breakdown of the equivalence principle (or effective quantum field theory description) in a regime of low curvature.

⁷The term “firewalls” generically denote various forms of high energy densities that may appear near the horizon. In addressing the AMPS puzzle, we will focus on firewalls that may arise from the loss of quantum entanglement between pairs of modes inside and outside the black hole. In other words, our primary focus is on an apparent violation of monogamy of quantum entanglement and how to restore it by properly understanding the breakdown of semiclassical arguments through the lens of quantum information theory. Note that there may be other physical mechanisms for firewalls, see [34] for instance, but these are beyond the scope of this paper.

on R . The problem may be resolved by employing the idea of quantum error-correction where the interior operator may be encoded in R via some sort of quantum error-correcting code [20, 21]. This viewpoint has been considered seriously by several authors [18, 22–25]. Relatedly, arguments based on quantum circuit complexity have been attempted [6, 30]. The key question, however, is the physical origin of fault-tolerant encoding of the interior operators.⁸

Second, it leads to state-dependent construction of interior operators [5, 15, 16]. To see the origin of state-dependence, suppose that a quantum state of a black hole $|\Psi\rangle$ is given by some maximally entangled state with $S_{CD} = \log_2 |C||D|$. Such a state can be written as

$$|\Psi\rangle = (I \otimes K)|\Phi^{\text{EPR}}\rangle, \quad |\Phi^{\text{EPR}}\rangle_{CDR} \equiv \frac{1}{\sqrt{|C||D|}} \sum_{i,j} |i,j\rangle_{CD} \otimes |i,j\rangle_R \quad (2.5)$$

where K is some unitary operator and $|\Phi^{\text{EPR}}\rangle_{CDR}$ are EPR pairs between CD and R .⁹ Then construction of operators for the interior modes \bar{D} on the early radiation R depends on the unitary operator K . But these state-dependent interior operators are problematic in describing quantum physics of the infalling observer due to a number of potential inconsistencies with quantum mechanics [8–10, 13, 14, 19].¹⁰

To summarize, our goal is to reconstruct interior operators in a state-independent manner with fault-tolerance against perturbations on R and without the non-locality problem.

3 Review of Hayden-Preskill recovery

In this section, we provide a brief review of the Hayden-Preskill thought experiment [27]. We try to be clear about what the recoverability of quantum states means in information-theoretic terms. We will mostly focus on a situation which mimics the AdS eternal black hole while the scope of the original argument by Hayden and Preskill is broader. Part of our goal is to remind readers of how the quantum cloning puzzle can be resolved by using the idea of backreaction within the context of the AdS/CFT correspondence.

⁸Some previous works argue that the involvement of the early radiation R is fine as long as information cannot be sent non-locally. We think that this is a serious problem if all the possible representations of the interior mode require degrees of freedom from R since the measurement of the interior mode by the infalling observer would require non-local coupling between the remaining black hole C and the early radiation R . In this paper, we will show that the construction of the interior mode and its measurement by the infalling observer can be achieved without ever involving degrees of freedom on the early radiation R .

⁹Harlow and Hayden suggested that there is a certain obstruction for the infalling observer to verify the quantum entanglement between the outgoing mode D and the radiation R , hence the AMPS thought experiment cannot be carried out [30]. Namely, they conjectured that quantum circuit complexity of verifying the quantum entanglement is exponential in the entropy S_{BH} by relying on computational complexity theoretical assumptions. This viewpoint has been further strengthened by Aaronson and Susskind [36, 37]. However, the complexity theoretic obstruction itself can be avoided by performing the so-called precomputation [38]. Also, the argument does not apply to the eternal AdS black hole where the verification of entanglement can be performed via simple quantum operations.

¹⁰Again, we think that this is a serious problem if all the possible representations of the interior mode are state-dependent.

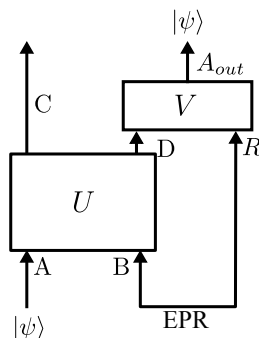


Figure 3. The Hayden-Preskill thought experiment. U represents the time-evolution of the black hole and V represents a recovery unitary.

3.1 Setup

As in the firewall argument, Hayden and Preskill considered an old black hole which has emitted more than half of its initial entropy. In the Hayden-Preskill thought experiment, Alice throws a quantum state $|\psi\rangle$ into a black hole and Bob, the outside observer, attempts to reconstruct it by collecting the Hawking radiation. To simplify the argument further, we treat the initial state as $n_B = \log |B|$ copies of EPR pairs between the black hole B and the radiation R :

$$|\Phi^{EPR}\rangle_{BR} = \frac{1}{\sqrt{|B|}} \sum_j |j\rangle_B \otimes |j\rangle_R. \quad (3.1)$$

Let us append a subsystem A to the black hole to account for the Hilbert space of the infalling quantum state $|\psi\rangle$. Following Hayden and Preskill, let us assume that the black hole evolves by a Haar random unitary U that acts on AB . Let C and D be the remaining black hole and the outgoing mode respectively (figure 3). Bob’s goal is to reconstruct $|\psi\rangle$ by catching D and having access to the early radiation R . The surprising result is that $|D| \gtrsim |A|$ is sufficient to achieve this goal. Namely, if $|\psi\rangle$ is an n_A -qubit quantum state, catching $n_D = n_A + \epsilon$ qubits of the outgoing mode with $\epsilon = O(1)$ suffices. By “reconstruction”, we mean the existence of some recovery unitary V_{DR} that, for any given input state $|\psi\rangle$, reconstructs the original state $|\psi\rangle$ (figure 3):

$$(I_C \otimes V_{DR})(U_{AB} \otimes I_R)(|\psi\rangle_A \otimes |\Phi^{EPR}\rangle_{BR}) \approx |\psi\rangle_{A_{out}} \otimes |\text{something}\rangle \quad \text{for all } |\psi\rangle \in \mathcal{H}_A. \quad (3.2)$$

Here A_{out} is some subsystem in DR with $|A_{out}| = |A|$, and “ \approx ” is measured in terms of the fidelity of the output state on A_{out} with $|\psi\rangle$ (i.e., the expectation value of the output state with respect to the projector $|\psi\rangle\langle\psi|_{A_{out}}$).

3.2 State representation of U

Instead of throwing an unknown quantum state $|\psi\rangle$ into a black hole, it is more convenient to introduce an ancillary subsystem A' which has the same dimensionality as A . We will see that this trick enables us to discuss the recoverability in a quantitative manner with

information-theoretic measures. Prepare an EPR pair on A' and A and throw qubits from A into the black hole B . After the time-evolution by a unitary operator U that acts on AB , the system is given by

$$|\Psi\rangle = (I_{A'} \otimes U_{AB} \otimes I_R)(|\text{EPR}\rangle_{A'A} \otimes |\Phi^{\text{EPR}}\rangle_{BR}) = \tag{3.3}$$

The diagram shows a central rectangular box labeled 'U'. Two vertical lines enter the box from the bottom: the left one is labeled 'A' and the right one is labeled 'B'. Below these two lines, the text 'EPR' is written. Two vertical lines enter the box from the top: the left one is labeled 'C' and the right one is labeled 'D'. Two vertical lines exit the box: one on the far left labeled 'A'' and one on the far right labeled 'R'. Below these two lines, the text 'EPR' is written again.

which is supported on $A'CDR$. The wavefunction in eq. (3.3) is often called the “state representation” (or the “Choi representation”) of U [28]. To remind readers of the notation, B is the initial black hole, R is the early radiation, C is the remaining black hole and D is the late radiation.

Information theoretically, the recoverability of unknown quantum states can be quantitatively addressed by the amount of entanglement between A' and RD , e.g., the mutual information $I(A', DR)$. If $I(A', DR)$ is close to its maximal value, nearly perfect EPR pair can be distilled by applying some recovery unitary operator V_{DR} :

$$(I_{A'C} \otimes V_{DR})|\Psi\rangle \approx |\text{EPR}\rangle_{A'A_{\text{out}}} \otimes |\text{something}\rangle. \tag{3.4}$$

Here “ \approx ” is measured in terms of the fidelity with the perfect EPR pair (i.e., the expectation value with respect to the EPR projector on $A'A_{\text{out}}$).¹¹

Distillation of EPR pair can be related to recovery of infalling quantum states as follows. By projecting the reference system A' onto $|\psi^*\rangle$, we will have $|\psi\rangle$ on the original input Hilbert space A since

$$\langle\langle\psi^*|_{A'} \otimes I_A\rangle\rangle|\text{EPR}\rangle_{A'A} \propto |\psi\rangle_A. \tag{3.5}$$

If $|\text{EPR}\rangle_{A'A_{\text{out}}}$ is distilled on $A'A_{\text{out}}$, the above projection on A' will generate $|\psi\rangle_{A_{\text{out}}}$ on the output Hilbert space A_{out} , implying successful reconstruction of the input state. One merit of introducing the reference system A' is that we do not need to keep track of how the recovery operation works for each choice of input states.

3.3 Reconstruction of operators

We will use the Hayden-Preskill recovery protocols to construct interior operators in the AMPS problem. For this purpose, it is convenient to state the Hayden-Preskill recovery in terms of operators. Given an arbitrary unitary operator $O_{A'}$ acting on A' , it is possible to identify a partner operator on DR if $I(A', DR)$ is nearly maximal (i.e., OTOCs are small). Recall the following relation:

$$(O_{A'} \otimes I)|\text{EPR}\rangle_{A'A_{\text{out}}} = (I \otimes O_{A_{\text{out}}}^T)|\text{EPR}\rangle_{A'A_{\text{out}}} \tag{3.6}$$

¹¹The fidelity of the distilled EPR pair can be quantitatively related to out-of-time ordered correlation function [31].

or graphically

$$\begin{array}{c} \uparrow \\ \boxed{O_A} \\ \text{EPR} \end{array} = \begin{array}{c} \uparrow \\ \boxed{O_A^T} \\ \text{EPR} \end{array} \quad (3.7)$$

which holds for any unitary operator $O_{A'}$ and its transpose $O_{A_{\text{out}}}^T$. Define \tilde{O}_{DR}^T as follows:

$$\tilde{O}_{DR}^T \equiv V_{DR}^\dagger O_{A_{\text{out}}}^T V_{DR}. \quad (3.8)$$

Then, the partner operator \tilde{O}_{DR}^T on DR satisfies

$$(O_{A'} \otimes I_{CDR})|\Psi\rangle \approx (I_{A'C} \otimes \tilde{O}_{DR}^T)|\Psi\rangle \quad (3.9)$$

or graphically

$$\begin{array}{c} \uparrow A' \\ \boxed{O_{A'}} \\ \uparrow A \quad \uparrow B \\ \text{EPR} \quad \text{EPR} \\ \boxed{U} \\ \uparrow C \quad \uparrow D \\ \uparrow R \end{array} \approx \begin{array}{c} \uparrow A' \\ \uparrow A \quad \uparrow B \\ \text{EPR} \quad \text{EPR} \\ \boxed{U} \\ \uparrow C \quad \uparrow D \\ \uparrow R \\ \boxed{\tilde{O}_{DR}^T} \end{array}. \quad (3.10)$$

3.4 Quantum cloning puzzle

The Hayden-Preskill recovery leads to an apparent cloning of quantum states since Alice possesses $|\psi\rangle$ inside the black hole whereas Bob possesses the reconstructed copy of $|\psi\rangle$ outside the black hole.¹² A traditional approach to resolve this puzzle is to resort to the idea of complementarity; we decide not to be bothered by violation of quantum mechanics as long as there is no observer who can verify it [39–42]. In order to verify the quantum cloning, Bob needs to wait for a black hole to implement a highly-complicated unitary U , catch the outgoing mode, reconstruct $|\psi\rangle$, and jump into a black hole to meet Alice who possesses the other copy. For the time being, assume that reconstructing $|\psi\rangle$ is instantaneous once the outgoing mode is collected. Then if it takes more than $t \sim r_s \log r_s$ in the Schwarzschild time (or $t \sim \log r_s$ at the stretched horizon) for the black hole to implement U , Alice will reach the black hole singularity before meeting Bob. Quantum information theoretic studies suggest that time scale to “delocalize” quantum information must be indeed longer than $t \sim \log r_s$ in a clock at the stretched horizon [43]. Hence, the quantum cloning puzzle is avoided.

While the above viewpoint provides a tentative resolution of the quantum cloning puzzle, a refined resolution has been recently proposed within the context of the AdS/CFT

¹²By introducing the reference system A' , the quantum cloning puzzle can be stated as potential violation of monogamy of entanglement. This observation hints that the Hayden-Preskill thought experiment is indeed related to the firewall problem.

correspondence. The original argument via complementarity assumes that the black hole horizon is an absolute entity and its location never moves regardless of quantum operations applied from the outside. A modern viewpoint on this puzzle is to postulate that Bob’s recovery operation may have a non-trivial backreaction to the black hole geometry. This viewpoint is largely motivated from recent developments in studies of backreactions in the AdS/CFT correspondence and related toy models such as the SYK model [44–48]. The key insight is that Bob’s operation from the outside may shift the location of the event horizon and pull the quantum state $|\psi\rangle$ from the interior to the exterior. A concrete physical realization of this scenario is the so-called traversable wormhole phenomena discovered by Gao, Jafferis and Wall [49]. In particular, for the eternal AdS black hole, they identified certain forms of interactions between CFTs on opposite boundaries which send negative energy to the bulk and shift the event horizon so that signals can traverse the wormhole.¹³ The relation between traversable wormholes and the Hayden-Preskill thought experiment is also discussed in [31, 50].

4 Scrambling and recovery

We have reviewed the Hayden-Preskill thought experiment and the quantum cloning puzzle from the perspective of backreaction within the context of the AdS/CFT correspondence, namely for the eternal AdS black hole. In this section, we extend the argument to more generic quantum systems by reviewing the relation between the Hayden-Preskill recovery and quantum information scrambling as diagnosed by out-of-time ordered correlation (OTOC) functions. We also discuss concrete recovery protocols for the Hayden-Preskill thought experiment.

4.1 Scrambling

Let us continue our discussion on the Hayden-Preskill recovery. Again, we will focus on a quantum black hole whose initial state is represented by $|\Phi^{\text{EPR}}\rangle_{BR} = \frac{1}{\sqrt{|B|}} \sum_j |j\rangle_B \otimes |j\rangle_R$. We are primarily interested in quantum entanglement between A' (the reference system) and DR (the late radiation and the early radiation) in the state representation of U as defined in eq. (3.3).

While Hayden and Preskill considered Haar random unitary U , a recovery operation can be performed in strongly-interacting quantum systems which delocalize quantum information over the whole system. This phenomena, often called quantum information scrambling, can be probed by the out-of-time ordered correlation (OTOC) function [44, 51–53]:

$$\langle O_A(0)O_D(t)O_A^\dagger(0)O_D^\dagger(t) \rangle \equiv \frac{1}{d} \text{Tr} (O_A U^\dagger O_D U O_A^\dagger U^\dagger O_D^\dagger U) \quad (4.1)$$

where $d = |A||B| = |C||D|$ denotes the total Hilbert space dimension. Two different bipartitions of the total Hilbert space \mathcal{H} , into $\mathcal{H}_A \otimes \mathcal{H}_B$ and $\mathcal{H}_C \otimes \mathcal{H}_D$, are considered. Here U is an arbitrary unitary operator that accounts the time-evolution of the system.

¹³Since their interaction couples two boundaries, null geodesics through the wormhole is no longer achronal. Hence the average null energy condition (ANEC) is obeyed for achronal geodesics.

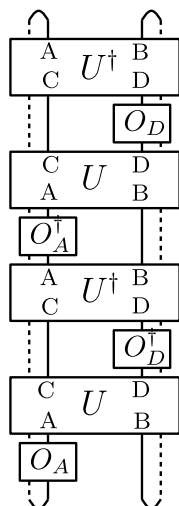


Figure 4. Out-of-time ordered correlation (OTOC) functions.

Note that O_A and O_D are operators that act on subsystems A and D respectively. For concreteness, we take O_A, O_D to be basis operators (such as Pauli operators) supported on A, D . For A, D with no overlap, the OTOC starts at 1 and then decays under chaotic time-evolution.

The decay of OTOCs implies that the Hayden-Preskill recovery can be performed [28, 54]. Specifically, the following equality was proven:

$$2^{-I^{(2)}(A', RD)} = \int dO_A dO_D \langle O_A(0) O_D(t) O_A^\dagger(0) O_D^\dagger(t) \rangle \quad (4.2)$$

where integrals with dO_A, dO_D take averages over all the basis operators supported on A, D respectively. Here $I^{(2)}(A', RD) \equiv S_{A'}^{(2)} + S_{RD}^{(2)} - S_{A'RD}^{(2)}$ represents the Rényi-2 mutual information.¹⁴ It is worth reminding that the left hand side of eq. (4.2) is defined for the state representation of U (eq. (3.3)) which is supported on $\mathcal{H}_{A'} \otimes \mathcal{H}_C \otimes \mathcal{H}_D \otimes \mathcal{H}_R$ whereas OTOCs on the right hand side are calculated on $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B = \mathcal{H}_C \otimes \mathcal{H}_D$. See figure 4 for graphical representation of OTOCs. We emphasize that the aforementioned relation holds for any unitary operator U .

¹⁴While we have $I(A', RD) \geq I^{(2)}(A', RD)$ in the above setting, the Rényi-2 mutual information does not lower bound the ordinary mutual information in general. In fact, from the perspective of quantum information theory, the Rényi-2 mutual information is not a proper entanglement monotone. Instead one must use the Sandwiched Rényi- α divergence. A similar duality relation is known for two of them with α and β satisfying $\frac{1}{\alpha} + \frac{1}{\beta} = 2$ [55]. In [56], a certain average of OTOCs is expressed as the $\alpha = 2$ Sandwiched Rényi divergence and is proven to lower bound the mutual information $I(A', RD)$.

The Hayden-Preskill recovery and quantum information scrambling can be explicitly related by the above formula in eq. (4.2).¹⁵ Namely, smallness of OTOCs implies large $I^{(2)}(A', RD)$ which implies the existence of reconstruction procedures.¹⁶

4.2 OTOCs in a black hole

Having seen that scrambling implies the recoverability in the Hayden-Preskill thought experiment, let us discuss OTOCs for black holes. OTOCs for gravitational systems can be explicitly computed by taking O_A to be a weak perturbation. Under the gravitational blueshift, such a perturbation is greatly amplified and becomes a gravitational shockwave as it falls into a black hole. For the Schwarzschild black hole, the effect of gravitational shockwaves on correlation functions were discussed by 't Hooft and Dray [59, 60], and also by Kiem, Verlinde and Verlinde [61]. Kitaev pointed out that these pioneering works actually computed OTOCs in disguise [44]. For the AdS black hole, OTOCs have been computed by Shenker and Stanford [52]. Shockwave geometries in the Reissner-Nordström black hole were studied by Sfetsos [62]. Recently shockwave geometries in the Kerr black hole were studied by BenTov and Swearngin [63].

These calculations suggest that OTOCs start to decay from its initial value significantly around the scrambling time, which is of order $\approx r_s \log r_s$ in the Schwartzshild black hole. As such, after the scrambling time, a quantum state that has fallen into a black hole can be retrieved as long as the black hole is maximally entangled. We emphasize that this argument is not restricted to black holes that are described in the AdS/CFT correspondence. Namely, we only assumed that the outside observer has a quantum description of a black hole with a pure quantum state on his Hilbert space, and OTOCs decay after the scrambling time.

¹⁵An alternative definition of quantum information scrambling was previously considered [43, 57]. This definition considers initial quantum states with low entanglement (e.g., product states) and asks if the states become thermalized. We think that the definition based on OTOCs is a more appropriate characterization of quantum information scrambling as its relation to black hole physics and Hayden-Preskill thought experiment is clear. For interested readers, a definition of scrambling at finite temperature was proposed in [31] by using OTOCs. Recently Shor presented an argument showing that scrambling in the alternative definition cannot be achieved within the scrambling time $\approx r_s \log r_s$ [58].

¹⁶Note that the above formula is restricted to quantum systems at infinite temperature (i.e., the system is finite-dimensional and the quantum state in OTOCs is given by the maximally mixed state $\rho = \frac{1}{d}I$). As such, the conclusion on the recoverability is applicable only to quantum black holes whose initial states are approximated by maximally entangled states $\approx (I \otimes K)|\Phi^{\text{EPR}}\rangle_{BR}$. Essentially, these states are simplified descriptions of maximally entangled black holes as an S_{BH} -qubit system where S_{BH} is the Bekenstein-Hawking (or the coarse-grained) entropy. Some readers might be worried about the validity of the relation between OTOCs and the recoverability for more realistic black holes at finite temperature. Further generalization of this relation for black holes with more generic quantum states has been discussed in [31]. Namely it was shown that the decay of OTOCs implies the recoverability as long as quantum correlations between A and B as well as C and D are small. For this reason, we do not view the simplification with $\approx (I \otimes K)|\Phi^{\text{EPR}}\rangle_{BR}$ as fundamental limitation of our argument. For simplicity of discussion, we proceed with this simplified description of a quantum black hole as a S_{BH} -qubit system. We will discuss its validity from other perspective in section 6.3.

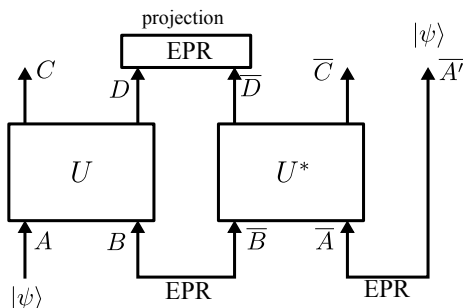


Figure 5. A recovery protocol for the Hayden-Preskill thought experiment from [31]. The input quantum state $|\psi\rangle$ on A is teleported to \bar{A}' upon postselecting $D\bar{D}$ to be EPR pairs. In the diagram, $|A| = |\bar{A}| = |\bar{A}'|$, $|B| = |\bar{B}|$, $|C| = |\bar{C}|$, and $|D| = |\bar{D}|$. Here $D\bar{D}$ are interpreted as outgoing modes. For scrambling unitary U , the recovery protocol works when $|D| \gtrsim |A|$. In particular, D can be chosen as any subsystem as long as OTOCs $\langle O_A(0)O_D(t)O_A(0)O_D(t) \rangle$ are small.

4.3 Recovery protocols

The above relation in eq. (4.2) provides a formal proof of the existence of recovery protocols in a scrambling system. The author and Kitaev proposed two particular “simple” recovery protocols which work for any scrambling systems [31]. Here we briefly review these protocols.¹⁷

The first protocol works probabilistically. It applies the complex conjugate U^* to the entangled partner R (which is denoted as \bar{B} in figure 5), and then projects the outgoing modes D and \bar{D} from both sides into EPR pairs. Upon postselecting the measurement outcome to be EPR pairs, Alice’s quantum state $|\psi\rangle$ will be teleported to Bob’s register qubits on \bar{A}' . See figure 5 for summary of the whole process. Noting that $D\bar{D}$ were initially EPR pairs at $t = 0$ (if A and D do not overlap), this recovery operation can be interpreted as the outside observer’s attempt to restore $D\bar{D}$ to be original EPR pairs. The protocol achieves high recovery fidelity when OTOCs decay.

The second protocol is more involved and works deterministically. It incorporates the quantum Grover search algorithm where U^* and U^T are applied on the entangled partner multiple times with Grover rotation operators inserted. Interested readers should see the original paper for details. Unlike the first protocol, the second protocol applies a careful sequence of unitary operators which restores $D\bar{D}$ to be EPR pair in a deterministic manner. If $|A| \sim O(1)$, the quantum circuit complexity of both recovery protocols is roughly equal to that of U . Hence, if it takes time t to implement U , it takes $\approx t$ to reconstruct $|\psi\rangle$. Both protocols enable us to reconstruct partner operators, although the first protocol requires

¹⁷These recovery protocols work well after the scrambling time whereas the traversable wormhole picture is valid before the scrambling time. The nature of scrambling after the scrambling time is quite different from that before the scrambling time, see [64, 65] for instance. While this paper focuses on the physics after the scrambling time, we believe that a qualitatively similar implication on the AMPS puzzle can be made by relying on the physics of scrambling before the scrambling time.

us to normalize reconstructed operators properly.¹⁸ Geometric interpretations of these protocols are discussed in appendix D.

So far we have mostly focused on a quantum black hole whose quantum state can be approximately modeled by $\approx |\Phi^{\text{EPR}}\rangle_{BR}$. A generic quantum state for a maximally entangled black hole, however, should be modeled by $\approx (I \otimes K)|\Phi^{\text{EPR}}\rangle_{BR}$ as in eq. (2.5) with some unitary operator K . Here we would like to clarify the applicability of our argument.

As a matter of principle, the quantitative argument on the recoverability remains valid regardless of K since K does not affect the mutual information $I(A', DR)$. In particular, as long as the unitary time-evolution operator U forces OTOCs to decay, one can prove the existence of recovery protocols. Also, the aforementioned recovery protocols can be applied to generic initial states by replacing U with UK^* . This essentially cancels the unitary operator K and reduces the system back to $|\Phi^{\text{EPR}}\rangle_{BR}$. Performing recovery protocols with UK^* , however, may be problematic if K is a highly complex unitary operator. One might argue that recovery protocols cannot be performed before the black hole evaporates completely. The difficulties associated with highly complex K may be avoided by the idea of the “precomputation” [38]. Yet, it is not clear if such a procedure is physically plausible in realistic systems.

We will return to this issue of the K -dependence in section 6 when the Hayden-Preskill recovery is used to reconstruct interior operators. We will see that, in the context of the AMPS puzzle, we can actually avoid the K -dependence and achieve state-independent reconstruction of interior operators.

5 Reconstruction of interior operators

In this section, we point out that the black hole interior operators can be reconstructed by using a protocol similar to the Hayden-Preskill recovery procedure. This subsection is a slightly reorganized reprint from section 5 of our previous work [26]. We will focus on cases where the initial state of the old black hole is given by EPR pairs while deferring discussions on generic entangled states, as well as the issue of the state-dependence, to the next section.

¹⁸Some readers may be worried about the fact that Bob needs to understand the dynamics of quantum black hole for reconstruction. Given our little understanding of quantum gravity, this might sound like a valid concern. However we emphasize that this is not the issue intrinsically related to quantum gravity. In fact, a similar problem can be identified in classical systems. Suppose that Alice prepares a classical bit 0 or 1 and Bob tries to find out what is written in Alice’s bit. If Alice’s bit remains unchanged, Bob will immediately know the answer. If Alice has applied a Pauli- X operator, her bit is flipped; $0 \rightarrow 1$ and $1 \rightarrow 0$. Then, Bob needs to flip her bit back to find the original information. Now suppose Bob does not know if Alice has applied X or not. Then it is not possible for Bob to know Alice’s original bit. As this simple classical observation suggests, Bob knowing the dynamics U is a prerequisite to treat the information loss problem in a rigorous information theoretic setting, and our little knowledge of U is a separate problem.

5.1 Hayden-Preskill for interior operators

The quantum state of interest is as follows:

$$|\Psi\rangle = \begin{array}{c} \begin{array}{ccc} \uparrow D & \uparrow C & \uparrow R \\ \hline & U & \\ \hline & \uparrow & \\ & \text{EPR} & \end{array} \end{array} \quad (5.1)$$

where the black hole started in EPR pairs and time-evolved by U . Here R is the early radiation, C is the remaining black hole and D is the outgoing mode. The goal is to find the partner \tilde{O}_{CR}^T of an arbitrary operator O_D on the outgoing mode:¹⁹

$$(O_D \otimes I_{CR})|\Psi\rangle = (I_D \otimes \tilde{O}_{CR}^T)|\Psi\rangle \quad (5.2)$$

or graphically

$$\begin{array}{c} \begin{array}{ccc} \uparrow D & \uparrow C & \uparrow R \\ \hline \boxed{O_D} & & \\ \hline & U & \\ \hline & \uparrow & \\ & \text{EPR} & \end{array} = \begin{array}{c} \begin{array}{ccc} \uparrow D & \uparrow C & \uparrow R \\ \hline & & \boxed{\tilde{O}_{CR}^T} \\ \hline & U & \\ \hline & \uparrow & \\ & \text{EPR} & \end{array} \end{array} \quad (5.3)$$

Existence of such partner operators is guaranteed since the mutual information $I(D, CR)$ is maximal. However, since $I(D, C) = 0$, partners cannot be supported on C . That is, partner operators \tilde{O}_{CR}^T must have non-trivial support on R . We have seen that, from the perspective of the outside quantum mechanical description, this fact (the use of the early radiation R in writing \tilde{O}_{CR}^T) is the origin of the firewall problem (or relatedly the state-dependence problem and the non-locality problem).

Here our aim is to construct the partner \tilde{O}_{CR}^T by using as few qubits from R as possible. To be explicit, let A be a small subsystem of R , and B be the complementary subsystem inside R (so $R = AB$). Here we would like to construct a partner \tilde{O}_{CA}^T by using only qubits on A from R , as graphically shown below:

$$\begin{array}{c} \begin{array}{cccc} \uparrow D & \uparrow C & \uparrow A & \uparrow B \\ \hline \boxed{O_D} & & & \\ \hline & U & & \\ \hline & \uparrow & \uparrow & \\ & \text{EPR} & & \end{array} \approx \begin{array}{c} \begin{array}{cccc} \uparrow D & \uparrow C & \uparrow A & \uparrow B \\ \hline & & \boxed{\tilde{O}_{CA}^T} & \\ \hline & U & & \\ \hline & \uparrow & \uparrow & \\ & \text{EPR} & & \end{array} \end{array} \quad (5.4)$$

where R is split into two subsystems AB .

¹⁹Construction of \tilde{O}_{CR}^T is not unique.

A key observation in [26] is that the task of reconstructing \tilde{O}_{CR}^T for a given O_D can be performed by using the Hayden-Preskill recovery protocol. Namely, by rotating the above figures in eq. (5.4) by 180 degrees and bent some arrows, we obtain the following relation:

$$\begin{array}{c}
 \begin{array}{cccc}
 D' & B & A & C \\
 \square & & & \\
 O_D & & & \\
 \downarrow & & & \\
 \text{EPR} & \uparrow D & \uparrow \bar{C} & \\
 & U^T & & \\
 \end{array}
 \approx
 \begin{array}{cccc}
 D' & B & A & C \\
 & & \square & \\
 & & \tilde{O}_{AC}^T & \\
 & & \downarrow & \\
 \text{EPR} & \uparrow D & \uparrow \bar{C} & \\
 & U^T & & \\
 \end{array}
 .
 \end{array}
 \tag{5.5}$$

Here U^T represents the transpose of U as the diagram of U is rotated upside down. In the rotated diagram, D is interpreted as an “input” while its partner is constructed on the remaining black hole C and just a little bit of qubits on A from the early radiation R . It is immediate to see that the structure of the diagram is identical to the one for the Hayden-Preskill thought experiment.

As such, the partner operator \tilde{O}_{CA}^T can be constructed by using the Hayden-Preskill recovery protocol from [31], as long as A is bigger than D and OTOCs between A and D decay.²⁰ It is worth summarizing the discussion so far as follows:

HP recovery:	input A	reconstruction on BD	(5.6)
Interior operator:	input D	reconstruction on CA .	(5.7)

We will study actual expressions of reconstructed interior operators more concretely in a future work.

5.2 Young black hole

The Hayden-Preskill recovery, as in the original Hayden-Preskill thought experiment, and the construction of interior operators outlined above can be achieved by using the same quantum information theoretic technique, namely the reconstruction protocol from [31]. It is, however, worth emphasizing that these two phenomena, the Hayden-Preskill recovery and the construction of interior operators, are *physically distinct*. In the original Hayden-Preskill thought experiment, the presence of the early radiation R , which is maximally entangled with the black hole, was essential for the recoverability of input quantum states. As such, the recovery was possible only for old black holes after the Page time. On the other hand, the aforementioned construction of the interior operators is possible for young black holes too as we will see below. In fact, the age of the black hole is not so essential.

²⁰The partners of interior operators are often called mirror operators in the literature. It is an interesting coincidence that Hayden and Preskill referred to their findings as “black hole as mirror”.

To see this explicitly, let us consider a young black hole which is not yet maximally entangled. One may model such a black hole simply as follows:

$$= \frac{1}{\sqrt{d_R}} \sum_{j=1}^{d_R} |j\rangle_{DC} \otimes |j\rangle_R \quad (5.8)$$

where the black hole Hilbert space CD is 2^n -dimensional (in a sense that the Bekenstein-Hawking entropy is $S_{\text{BH}} = n$) while the early radiation R is supported on a d_R -dimensional Hilbert space with $d_R < 2^n$. The triangle in the figure denotes an isometric embedding, emphasizing that DC is bigger than R . Taking $d_R = 1$ would correspond to a one-sided black hole. The quantum state of interest is obtained by time-evolving the above state:

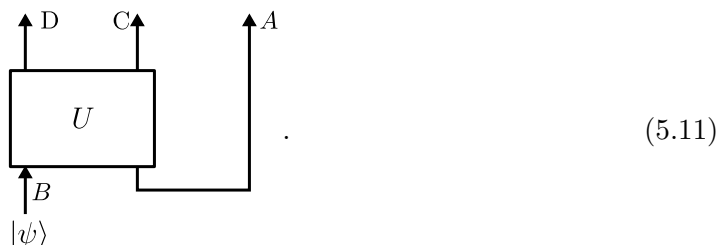
$$|\Psi\rangle = \quad (5.9)$$

For such a time-evolved young black hole, let us rotate the diagram to obtain

$$(5.10)$$

Then it is immediate to see that the partner can be reconstructed on C and some small subsystem A of R , as long as A is bigger than D . Hence, even when a black hole is young, the interior operator can be reconstructed by using the Hayden-Preskill recovery protocol (as long as R itself is bigger than D).

Another (perhaps more illuminating) way to see that the reconstruction method works for a young black hole is to start with an old black hole, as in eq. (5.1), and project a subsystem of R onto some fixed quantum state. Starting from a maximally entangled old black hole, let us split R into AB as before. Recall that we were able to reconstruct a partner operator \tilde{O}_{CA}^T on CA without using B . Let us apply a projection onto some quantum state $|\psi^*\rangle_B$ supported on B .²¹



After the projection, the system reduces to a young black hole which is entangled only with a small subsystem A . The reconstructed operator on CA works as good as before the projection since its expression does not involve B at all.²²

From these observations, we notice that the aforementioned reconstruction of interior operators is *fault-tolerant* against perturbations on R . Fault-tolerance is a quantum information theoretic term to say that encoding of quantum information is resilient against some types of errors. Here, the black hole interior operators are immune to rather drastic noises which would damage *all the qubits* on $B \subset R$.²³ In qubits count, the reconstruction is robust even if any subsystem of $S_{\text{BH}} - \epsilon$ qubits from the early radiation are damaged as long as $\epsilon \gtrsim |D|$. It is worth emphasizing that one can choose A to be any subsystem of R as long as OTOCs are small. Moreover, choosing different subsystems A leads to different expressions of the interior operators \tilde{O}_{CA}^T . The existence of multiple equivalent expressions of operators is a key characteristic of quantum error-correcting code. The necessity of such fault-tolerance was previously pointed out by Maldacena and Susskind [6]. The above observation provides a physical account on how fault-tolerance is achieved due to scrambling dynamics of black holes.

So far we have shown that the partner operators can be constructed on CA where A is a small subsystem in R . In a sense, our construction is only “slightly” non-local as it uses only a small portion from the early radiation R . However, our construction does not resolve some of important problems identified in the context of the firewall problem. First, the reconstruction still requires an access to distant radiation R albeit very little. Second, the reconstruction is state-dependent. For a generic initial state $(I \otimes K)|\Phi^{\text{EPR}}\rangle$, the recovery protocols should be run with $U \rightarrow UK^*$, so the reconstructed operators depend on K . Relatedly, the quantum circuit complexity of reconstruction can be huge depending on the complexity of K . We will remove these undesirable features in the next section by considering the effect of including the infalling observer.

²¹Note $(I \otimes |\psi^*\rangle\langle\psi|)|\text{EPR}\rangle = |\psi\rangle \otimes |\psi^*\rangle$.

²²More precisely, it is because D is (nearly) maximally entangled with CA when OTOC between A and D decay. In quantum information theoretic term, the subsystem B is said to be decoupled from D .

²³More precisely, it is an erasure error.

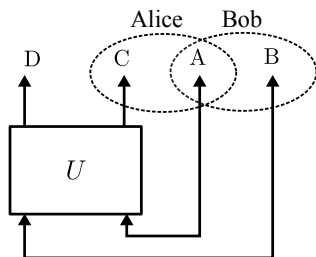


Figure 6. The AMPS puzzle from the perspective of the monogamy of entanglement. Alice can distill an EPR pair by accessing AC while Bob can distill an EPR pair by accessing AB . Note that A can be any subsystem of the early radiation R as long as $|A| \gtrsim |D|$ due to scrambling property of U .

5.3 Monogamy and scrambling

Our finding in this section sheds a new light on the AMPS puzzle through the lens of quantum information scrambling. The important lesson from the aforementioned reconstruction is that whoever possesses a tiny portion A of the early radiation R will be able to see some degrees of freedom which are entangled with the outgoing mode D . Namely, if A is included to C as degrees of freedom which Alice can access, then she is able to see an EPR pair between D and AC . On the other hand, if A is left untouched, the outgoing mode D is entangled with $R = AB$ and Bob is able to see an EPR pair from the outside. See figure 6 for schematic illustration of the situation.

These statements can be made quantitative by using the monogamy relation of the mutual information. Decay of OTOCs implies that $I^{(2)}(D, AC)$ is nearly maximal. Since $I^{(2)}(D, AC) + I^{(2)}(D, B) = 2S_D^{(2)}$, this suggests that $I^{(2)}(D, B)$ is close to zero. Namely, Alice reconstructs \bar{D} on AC while Bob cannot reconstruct it on B .

It is worth emphasizing that A can be any subsystem of R as long as $|A| \gtrsim |D|$.²⁴ Hence, from the perspective of the infalling observer, it is rather “easy” to steal the EPR pair from the outside observer as she needs to take a few qubits in R from Bob. We will make this intuition more precise in section 6 and section 7. In fact, we will show that generic perturbations to the black hole by the infalling observer will disentangle the outgoing mode D from the early radiation R *without* ever accessing R .

6 State-independent interior operators

In the previous section, we have pointed out that the interior operators can be constructed on the remaining black hole C , together with a little bit of qubits A from the early radiation R . This construction, however, still suffers from the state-dependence and non-locality problems.

In this section and the next, we discuss the effect of explicitly including the infalling observer to the firewall problem. We will show that inclusion of the infalling observer to the system, no matter how it is done, leads to significant gravitational backreaction

²⁴A common misunderstanding is to think A as qubits which are entangled with D (i.e., the distilled qubits in the AMPS thought experiment). They are not.

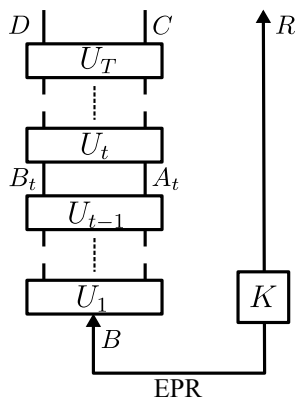


Figure 7. Discretized time evolution of a large AdS black hole with long throat. K represents some complex unitary and A_t are the confined modes. R is far distant on the right-hand side.

which disentangles the outgoing mode D from the early radiation R . This decoupling phenomena enables us to construct interior operators in a way independent of the initial state of the black hole without using any degrees of freedom from the early radiation R . Hence, the inclusion of the infalling observer enables state-independent interior operators and resolves the non-locality problem. Yet, the construction *depends* on how the infalling observer is introduced to the system. In this sense, our construction is *state-independent*, but *observer-dependent*.

In this section, we will demonstrate that a certain small perturbation to the black hole is sufficient to disentangle the outgoing mode from the early radiation. The main goal of this section is to simply point out that it is possible to disentangle the outgoing mode D from the early radiation R and construct the interior partner without using R . Discussions on physical interpretations of such a perturbation are given in the next section. Readers may find the particular scenario we consider in this section rather fine-tuned and artificial. In the next section, we will see that a similar phenomena occurs without any fine-tuning, in a rather universal manner, regardless of how perturbations are added to the black hole.

6.1 Long-throat AdS black hole

Let us consider a large two-sided AdS black hole with long throat (i.e., K is highly complex). The AdS assumption is used to ensure that the black hole does not evaporate. Let us imagine that the black hole initially starts with $(I \otimes K)|\Phi^{\text{EPR}}\rangle_{BR}$ and time-evolves by U acting on B . Imagine that U is discretized into small time steps:

$$U = U_T \cdots U_1. \tag{6.1}$$

We emphasize that details of the discretization are not important in the following discussion. The quantum state at $t = T$ is depicted in figure 7. Here intermediate Hilbert spaces are labelled by A_t, B_t where A_t corresponds to the modes on the AdS boundary while B_t corresponds to other degrees of freedom including modes between the boundary and the stretched horizon as well as other entities living at the stretched horizon.

We now show that the partner of interior operators D can be reconstructed by quantum operations which are strictly localized on the left-hand side without ever accessing R . Consider the quantum state of the black hole at time t :

$$(U_{t-1} \cdots U_1 \otimes K) |\text{EPR}\rangle_{BR}. \quad (6.2)$$

To begin, we append two subsystems E, \bar{E} , which have the same dimensionality as A_t at time t , and prepare EPR pair on $E\bar{E}$ as shown in figure 8a. The resulting state can be written explicitly as follows:

$$|\Psi(t)\rangle = (U_{t-1} \cdots U_1 \otimes K) |\text{EPR}\rangle_{BR} \otimes |\text{EPR}\rangle_{E\bar{E}}. \quad (6.3)$$

Note that the EPR pair on E, \bar{E} is added to the black hole Hilbert space by hand, and is not a part of the original Hilbert space. Let us apply SWAP operator between A_t and E as shown in figure 8a:

$$|\Psi'(t)\rangle = \text{SWAP}_{A_t \bar{E}} ((U_{t-1} \cdots U_1 \otimes K) |\text{EPR}\rangle_{BR} \otimes |\text{EPR}\rangle_{E\bar{E}}). \quad (6.4)$$

Here SWAP is a unitary operator which acts as $\text{SWAP}(|i\rangle \otimes |j\rangle) = |j\rangle \otimes |i\rangle$.

The outcome of these operations is depicted in figure 8b. We find that \bar{E} is entangled with A_t . Namely, we notice that \bar{E} can play the role of A in section 5. Hence, the interior operators (the partner of operators on D) can be reconstructed on $\bar{E}C$ by applying the recovery unitary from the Hayden-Preskill recovery protocol with $U' = U_T \cdots U_t$ as long as $T - t \gtrsim t_{\text{scr}}$. More explicitly, at time $t = T$, the quantum state of the black hole is given by

$$|\Psi'(T)\rangle = (U_T \cdots U_t \otimes I_{E\bar{E}R}) \text{SWAP}_{A_t \bar{E}} ((U_{t-1} \cdots U_1 \otimes K) |\text{EPR}\rangle_{BR} \otimes |\text{EPR}\rangle_{E\bar{E}}). \quad (6.5)$$

Given an arbitrary operator O_D on D , one can construct a partner operator $\tilde{O}_{C\bar{E}}^T$ on $C\bar{E}$ such that

$$(O_D \otimes I) |\Psi'(T)\rangle \approx (I \otimes \tilde{O}_{C\bar{E}}^T) |\Psi'(T)\rangle. \quad (6.6)$$

Due to the additional of the $E\bar{E}$ system and the SWAP operation, the outgoing mode D is no longer entangled with the early radiation R . Namely, the partner $\tilde{O}_{C\bar{E}}^T$ can be reconstructed without involving the early radiation R ever.

6.2 State-independence

Having constructed a partner operator without using the early radiation R , let us discuss the problem of the state-dependence. As is evident from figure 8b, the reconstruction of interior operators depends only on $U' = U_T \cdots U_t$. Importantly, it has no dependence on the unitary K , and hence is state-independent. As such, the reconstruction is not only efficient with $O(|T - t|)$ quantum circuit complexity, but also independent from the choice of the initial state $(I \otimes K) |\Phi^{\text{EPR}}\rangle_{BR}$.

The aforementioned protocol works for the case of one-sided AdS black holes. Let us represent the initial state of the black hole by $K|0\rangle^{\otimes n}$ and time-evolve it by $U = U_T \cdots U_1$. This is equivalent to projecting the right-hand side R in figure 7 onto $|0\rangle^{\otimes n}$. Since the

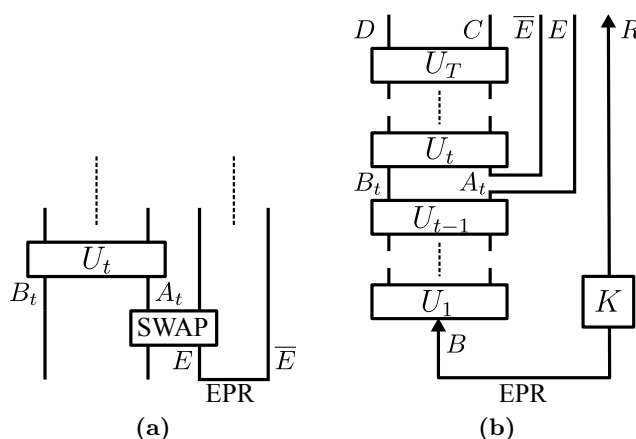


Figure 8. A protocol for an infalling observer to access A_t . EPR pair is prepared on $E\bar{E}$. (a) SWAP operation between A_t and E . (b) The outcome. \bar{E} and A_t are entangled.

recovery protocol does not depend on any quantum operation on R , be it unitary or non-unitary, the reconstructed interior operators are independent from the initial state.²⁵ The important point for us is that OTOCs between A and D decay for one-sided black holes too.

Finally let us return to the discussion of evaporating black holes as originally considered by AMPS. We model the evaporation dynamics as a sequential application of the procedures depicted in figure 9. Here A_t are the confined mode inside the zone whereas R_t are the modes which escape from the zone and never return. As time passes, R_t joins the radiation R , and the entropy of the black hole decreases.²⁶ The aforementioned protocol works successfully by swapping E and A_t . Here it is crucial for an infalling observer E to interact with the confined modes A_t rather than the escaping modes R_t as R_t is completely decoupled from $A_t B_t$. Again, the interior operators can be reconstructed in a state-independent and low-complexity manner without ever accessing R .

Readers might find that the above construction of the interior operator is rather artificial and fine-tuned. Obviously, this setup was devised in order to mimic the situation considered in section 5. One may interpret $E\bar{E}$ as certain measurement apparatus or physical probe which is introduced to the black hole Hilbert space.²⁷ In the next section, we will consider corresponding physical situations and how our reconstruction resolves the AMPS

²⁵When the initial state of the black hole is given by n copies of EPR pairs, projecting R onto $|\psi^*\rangle$ will project the black hole on B onto $|\psi\rangle$. Naively, this appears to suggest that it might be possible to influence the black hole on B and its interior by non-unitary operation on R , contradicting with our claim. The important point is that the 2^n -dimensional subspace where the black hole is initially entangled with R should be interpreted as a coarse-grained Hilbert space with roughly equal energy. As long as projection on R picks some quantum states of the black hole in thermal equilibrium within the typical energy window, reconstructed interior operators will behave well. Yet, if projection on R brings the black hole on B to a quantum state with much higher/lower temperatures, then reconstruction won't work well. See section 6.3 for further discussion on this point.

²⁶The AdS black hole can be viewed as a limit where R_t disappears.

²⁷Situations analogous to the $E\bar{E}$ -system plus SWAP can be realized by performing informationally complete measurements (POVMs).

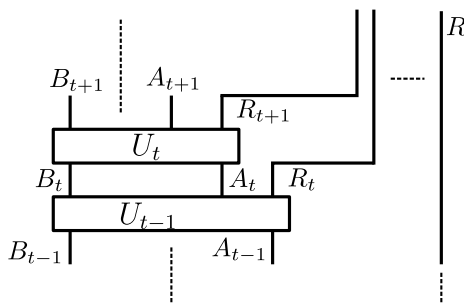


Figure 9. The dynamics of an evaporating black hole. R_t represents the modes which escape and join R while A_t are the confined modes.

puzzle. We will also argue that the underlying physical phenomenon is quite universal in gravitational systems and does not require fine-tuning.

6.3 Code subspace

We have presented state-independent reconstruction of interior operators that are completely insensitive to perturbations on the early radiation R . Strictly speaking, this conclusion on state-independence is a bit of exaggeration. For instance, large perturbations added to A_t will influence the Hamiltonian under which U_t evolves. As such, it is important to note that the state-independence is an approximate statement.

To understand the validity of the approximation, let us return to the underlying assumption behind using simplified descriptions of quantum black holes as S_{BH} -qubit systems. In static geometries, the Bekenstein-Hawking entropy S_{BH} can be interpreted as the coarse-grained entropy. Namely it suggests that there are $\approx 2^{S_{\text{BH}}}$ -dimensional Hilbert space, denoted by $\mathcal{H}_{\text{code}}$, which have an identical classical black hole geometry. This subspace is spanned by quantum states which correspond to black holes in thermal equilibrium with the same thermodynamic parameters, such as mass M , angular momentum J and charge Q . Those black hole microstates, however, differ in a subtle manner. They may differ in terms of matter content on a fixed geometry and other high energy objects living at the stretched horizon. Black holes in thermal equilibrium with the same background classical geometry span the subspace $\mathcal{H}_{\text{code}}$.

The very motivation in reconstructing interior operators in a state-independent manner is to understand quantum field theory on a curved space-time experienced by the infalling observer. This suggests that reconstructed operators must be state-independent on a fixed black hole geometry, but can be dependent on choices of geometries. Our claim on the state-independence is restricted to the validity of S_{BH} -qubit toy models, and hence is an approximate statement inside the subspace $\mathcal{H}_{\text{code}}$ which is determined by the classical black hole geometry.

These interpretations become sharp in the context of the AdS/CFT correspondence. In [20], Almheiri, Dong and Harlow proposed that the subspace $\mathcal{H}_{\text{code}}$ can be interpreted as a codeword subspace of a quantum error-correcting code. Their insight, as well as concrete

toy model of such a scenario [21], suggests that bulk operators can be interpreted as logical operators which label different codeword states in $\mathcal{H}_{\text{code}}$.²⁸

From the perspective of the boundary Hilbert space, one may also interpret the origin of the codeword subspace $\mathcal{H}_{\text{code}}$ by relying on the Eigenstate Thermalization Hypothesis (ETH) [66]. In particular, for strongly interacting quantum many-body systems, eigenstates from a tiny energy window are expected to look thermal in small subsystems. One may think that the subspace $\mathcal{H}_{\text{code}}$ is spanned by these wavefunctions contained in the small energy window.

Keeping this caveat in mind, our claim in a more precise term is that the reconstruction of interior operators is fault-tolerant against perturbations added to the entangled partner R as well as any quantum operations added before time t , provided that the black hole is initially in thermal equilibrium.²⁹ We also believe that the above claim includes cases of evaporating black holes where the system reaches thermal equilibrium quickly. The typical time scale for our reconstruction procedure is the scrambling time which is of order $\approx r_S \log r_S$ in the Schwartzschild black hole whereas the thermalization time for small perturbation is of order $\approx r_S \approx \frac{1}{T}$.

7 Backreaction by infalling observer

In this section, we show that gravitational backreaction by the infalling observer, who simply fall into a black hole, disentangles the outgoing mode D from the early radiation R . As a result, the infalling observer crosses the horizon smoothly and observes quantum entanglement between the outgoing mode D and the interior mode \bar{D} . We also show that the interior mode \bar{D} is insensitive to any perturbations to the early radiation R from geometric perspective. Our conclusions will be supported quantitatively by using decay of OTOCs.

7.1 Including Alice

Our reconstruction protocol required an access to the earlier mode. Readers might think that we have just traded the spatial non-locality with the temporal non-locality. The infalling observer can access earlier modes by simply falling into the black hole earlier. To be concrete, let us focus on the two-sided eternal AdS black hole. Let D be the outgoing mode that reaches the boundary at $t = 0$ as shown in figure 10a. As an infalling observer, one may consider some apparatus M which departs the boundary at $t = -\Delta t$. To minimize the effect of adding M , one may consider a limit where M assumes the smallest possible energy, say $\approx 1/\beta$. Here the apparatus M travels along the infalling mode A at $t = -\Delta t$.

To understand the effect of M , it is convenient to draw the partner mode \tilde{D} in the Penrose diagram. The reason why we used \tilde{D} instead of \bar{D} will become clear later. Here the trajectory of \tilde{D} can be constructed by rotating the trajectory of D by 180 degrees as in figure 10a. Near the horizon, D and \tilde{D} can be viewed as Rindler modes in the opposite

²⁸When perturbations being the system completely outside $\mathcal{H}_{\text{code}}$, the Hayden-Preskill recovery phenomenon may not be possible as discussed in [56].

²⁹By thermal equilibrium, we mean that time-ordered correlation functions reach their stationary values.

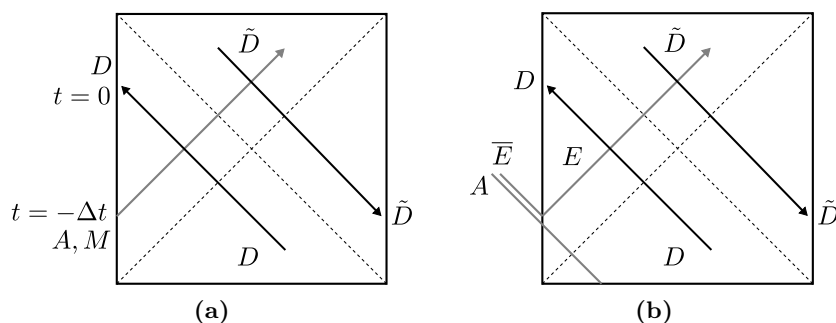


Figure 10. (a) The effect of an apparatus M . (b) The protocol from section 6.

regions. Near the boundary, D and \tilde{D} can be viewed as entangled degrees of freedom in the boundary conformal field theory which is in the thermofield double state. In the absence of the apparatus M , Bob can verify the entanglement between D and \tilde{D} by simply measuring the correlation functions between them. Also, one may explicitly compute the mutual information $I(D, \tilde{D})$ by using the Ryu-Takayanagi formula [67]. Here the interior operator \tilde{D} is constructed exclusively on the early radiation R (or degrees of freedom in the right hand side).

However, in the presence of the measurement apparatus M , Bob’s quantum entanglement between D and \tilde{D} is disturbed. Precise form of interactions between the apparatus M and other degrees of freedom is not so important. In fact, the interaction is universal in a sense that it is a gravitational effect. Namely, since M becomes a gravitational shockwave which shifts the geometry near the horizon, the correlation function between D and \tilde{D} decays. The two-point correlation functions between D and \tilde{D} in the presence of a shockwave is identical to OTOCs defined for a thermal state [52]. As such, when $\Delta t \gtrsim t_{\text{scr}}$, Bob can no longer verify the entanglement between the outgoing mode D and the early radiation R by measuring D and \tilde{D} . This hints that Bob’s quantum entanglement is destroyed by the infalling observer who has crossed the horizon.

One might think that this observation does not fully deny a possibility that the outgoing mode D is entangled with some other degrees of freedom in the early radiation R . In order to make a more quantitative argument, let us explicitly study the effect of adding the apparatus M . Suppose that the quantum state of a black hole is given by EPR pairs $|\Phi^{\text{EPR}}\rangle_{ABR}$ at $t = -\Delta t$. We then append the subsystem for the apparatus M to the black hole. We assume that M starts with some particular pure state, say $|0\rangle$. Under the time-evolution by a unitary operator U , we obtain the following state:

$$(U_{BAM} \otimes I_R)(|0\rangle_M \otimes |\Phi^{\text{EPR}}\rangle_{BAR}) = \text{Diagram} \quad (7.1)$$

Here we increased the Hilbert space dimension of the black hole in order to account for the increase of the coarse-grained entropy due to the inclusion of M .³⁰ We assume that this extension of the Hilbert space via equilibration occurs in thermalization time which is much shorter than the time scale of applying U (or scrambling time). If U is Haar random unitary operator, one can show that $I(C, D)$ becomes nearly maximal when $|M| \gtrsim |D|^2$. This suggests that quantum entanglement between D and R , which Bob would have seen, disappears as a result of including M . We emphasize that this argument on $I(C, D)$ is not limited to the AdS/CFT correspondence as it only requires the quantum state of a black hole to be maximally entangled.

A similar setting was previously considered by Verlinde and Verlinde in a slightly different context [22], and we will make further comments on their work in appendix B.

While the aforementioned estimate assumes that U is a Haar random unitary, one can relate certain average of OTOCs to the mutual information $I(C, D)$ for the state defined in eq. (7.1). Again, generic decay of OTOCs is sufficient to show that $I(C, D)$ is nearly maximal.³¹ Also, further analysis with energy conservation into consideration suggests that $|M| \gtrsim |D|$ is sufficient to restore nearly maximal correlation in $I(C, D)$ as discussed in appendix A.

7.2 Sending probes

In the previous subsection, we discussed the effect of including the infalling observer (or a measurement apparatus). The situation, however, differs from the reconstruction method in section 6 in a subtle detail. Also we had to change the size of the black hole Hilbert space by including M . Here we consider a situation which exactly mimics our reconstruction method.

Instead of adding the apparatus M , we think of replacing the infalling mode A at $t = -\Delta t$ with some “probe” mode. Let us prepare the EPR pair on $E\bar{E}$. Let the outgoing mode A at $t = -\Delta t$ escape from the boundary and replace the infalling mode A with E while keeping \bar{E} as a reference qubit. See figure 10b for the illustration. Here the earlier outgoing mode A and the reference \bar{E} remain outside the boundary whereas E probes the bulk. This is essentially identical to the protocol with SWAP operation discussed in section 6. Decay of OTOCs implies that $I^{(2)}(D, \bar{E}C)$ becomes nearly maximal which suggests that the outgoing mode D and $\bar{E}C$ are indeed entangled. This also implies that D is no longer entangled with R due to the monogamy of quantum entanglement. Instead, it is entangled with the interior mode \bar{D} which we constructed on section 6. Again, this argument is not particularly limited to the AdS/CFT correspondence.

We have seen that including $E\bar{E}$ also disentangled the outgoing mode D from the early radiation R . One merit of introducing the probe mode $E\bar{E}$ as in section 6 is that the size of the Hilbert space for the black hole itself does not change. Note that this protocol differs from the scenario in the previous subsection where an apparatus M was introduced. Nevertheless, we believe that these differences, which result from how we

³⁰The increase of the entropy is approximately given by E/T where E is the energy of M including its rest mass and T is the temperature.

³¹This statement can be shown by a slight generalization of results from [28, 31].

include the apparatus M in the black hole Hilbert space, do not affect our main conclusion significantly. Some readers may find inclusion of extra objects, such as the apparatus M and the ancilla qubits $\bar{E}E$, uncomfortable. The reason why we need to introduce extra degrees of freedom is that a black hole is treated as a system of discrete finite-dimensional qubits, and hence, there is no room for adding probes. In a more realistic field theoretic formulation, one may simply turn on the source term which creates the “apparatus” or the “probe mode” as an excitation.

Regardless of how the probes or the apparatus are introduced, what is important for us is that OTOCs between M (or A) and D decay after the scrambling time. From gravitational perspective, this is a universal phenomenon resulting from the surface gravity and shockwave geometries. From quantum information theory perspective, this is sufficient to prove that the outgoing mode D is disentangled from the early radiation R and \bar{D} can be reconstructed on A and C without involving R . Hence, the disentangling phenomenon occurs universally no matter how she falls.³²

Finally we speculate on non-uniqueness of reconstruction of interior operators. From the perspective of the outside observer, the interior operator is realized as \tilde{D} supported on R . On the other hand, from the perspective of the infalling observer who travels along the infalling mode A , the interior operator is realized as \bar{D} which is supported on the remaining black hole C and the infalling mode A . Hence, we propose that reconstruction of \bar{D} from section 6 corresponds to quantum mechanical operators for the infalling observer who departed at $t = -\Delta t$.³³ In principle, the infalling observer may choose to fall into a black hole at any time by following any trajectory as long as it does not violate the causality. We speculate that different infalling trajectories correspond to different reconstructions of interior operators. An explicit relation between trajectories and reconstructions, however, is beyond the scope of this paper.³⁴

³²One might wonder whether an observer can be introduced in a way which introduces no backreaction. From the outside quantum mechanical viewpoint, this appears to be not possible if the perturbations are added on one side of the black hole. Our argument relies on decay of OTOCs which is a rather generic feature of interacting quantum many-body systems. It is in principle possible that the value of OTOCs returns to $O(1)$ value after a very long time (of the order the quantum recurrence time). But in time scales which are relevant to the AMPS puzzle, we believe that OTOCs decay monotonically to a small stationary value and then oscillates around it. We believe that such a conclusion can be mathematically proven by utilizing ETH. One possible way to undo the effect of backreaction is to perform the Hayden-Preskill recovery protocol by accessing both sides of the black hole as we discuss later.

³³Although we have included an observer or an apparatus, one can discuss the monogamy of entanglement purely from the operator algebraic perspective. The important point is that representations of basis operators for \bar{D} on AC and AB do not necessarily commute.

³⁴In the aforementioned scenarios, infalling observers were introduced to the system by appending ancilla Hilbert spaces. Readers might wonder if addition of ancilla systems are necessary to achieve the disentangling phenomena or not. One can actually avoid the use of ancilla systems simply by considering a more physical realistic description of black holes, at least within the context of the AdS/CFT correspondence. Let \mathcal{H}_{CFT} be the full Hilbert space of the CFT and consider a black hole at temperature β which lives on a subspace $\mathcal{H}_\beta \subset \mathcal{H}_{\text{CFT}}$ consisting of microstates at the same temperature. Inclusion an infalling observer or measurement apparatus increases the entropy of the black hole by $\Delta S = \frac{E}{T}$. The point is that physics of adding infalling observers can be discussed in the full CFT Hilbert space \mathcal{H}_{CFT} .

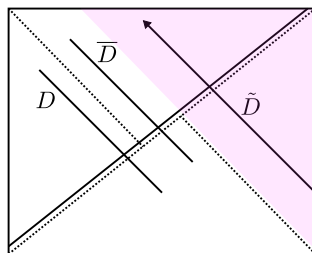


Figure 11. The infalling observer as a gravitational shockwave. The interior mode \bar{D} is outside a shaded region that may be influenced by quantum operations on the right hand side. The original mode \tilde{D} emerges from the right hand side, and is not longer entangled with D .

7.3 Geometric interpretation

We have argued that the outgoing mode D is disentangled from the early radiation R (or the right hand side) due to the decay of OTOCs caused by the infalling observer. Our primary focus was on assessing the effect of backreaction from the outside quantum mechanical description. Here we present geometric interpretation of our discussion.³⁵

For simplicity of discussion, let us consider the AdS eternal black hole. Given the outgoing mode D , one possible representation of interior partner operators can be constructed by time-evolving a corresponding mode on the right hand side. This operator, constructed exclusively on the degrees of freedom on the right hand side, was denoted by \tilde{D} . We now include the effect of the infalling observer M as a gravitational shockwave and draw the backreacted geometry where the horizon is shifted as depicted in figure 11. If the time separation between the outgoing mode D and the infalling observer M is longer than or equal to the scrambling time, the interior mode \bar{D} can be found across the horizon and is outside the causal influence of any operations on the right hand side. This is the geometric explanation of the fault-tolerance and state-independence of the interior operator \bar{D} . It is worth noting that a similar observation applies to generic black holes in thermal equilibrium.³⁶ Here it is important to note that the interior mode \bar{D} in a backreacted geometry is different from the original partner operator \tilde{D} .

This geometric interpretation enables us to deduce the experience of the infalling observer. From the backreacted geometry, we see that the infalling observer can cross the horizon smoothly and observe the entangled pair D and \bar{D} . According to the causal structure, after observing the interior mode \bar{D} , Alice eventually enters a region which may be affected by quantum operations on the right hand side. Inside this region, shaded in figure 11, Alice may be hit by a gravitational shockwave created by the outside observer, according to the backreacted geometry. It is, however, unclear to us whether perturbations on the right hand side can actually influence the experience of the infalling observer at all

³⁵I thank Xiaoliang Qi and Ying Zhao for useful discussions on this.

³⁶Here we treated the effect of the perturbation by Alice first on the left hand side, and then discussed the effect of the perturbations added on the right hand side. One might wonder why we did not treat the effect of the perturbation on the right hand side first. According to the structure of quantum entanglement implied by decay of OTOCs, we should treat the effect of the perturbation on the left hand side first when addressing the experience of the infalling observer from the left hand side.

in the absence of non-local interactions between the left and right hand sides (or between the remaining black hole and distant early radiation). In fact, we do not think that perturbation on R influences Alice’s infalling experience at all if it is strictly localized on R . In any cases, quantum operations on the right hand side do not affect Alice’s experience of crossing a smooth horizon and seeing D and \bar{D} .

8 On definition of state-independence

Some previous studies suggest that black hole interior operators must be state-dependent. In this section, we revisit these no-go arguments for state-independence and discuss how our construction avoids them. The main message is that our construction does not depend on the initial state of the black hole, but does depend on how the infalling observer is introduced, being *state-independent*, but *observer-dependent*.

8.1 No-go arguments

Let us briefly recall some versions of the no-go arguments for state-independence for an old black hole. Consider a maximally entangled black hole:

$$|\Psi\rangle = (I \otimes K)|\text{EPR}\rangle_{CDR} \tag{8.1}$$

where C is the remaining black hole, D is the outgoing mode and R is the early radiation. Here R can be decomposed into \bar{C} and \bar{D} such that C and D are entangled with \bar{C} and \bar{D} respectively. Given a unitary operator O_D acting on D , a partner operator \tilde{O}_{CR} must satisfy

$$(O_D \otimes I_{CR})|\Psi\rangle \approx (I_D \otimes \tilde{O}_{CR})|\Psi\rangle. \tag{8.2}$$

One possible representation can be constructed as follows:

$$\tilde{O}_{CR} = KO_D^T K^\dagger \tag{8.3}$$

by using the fact $(O_D \otimes I_{C\bar{C}\bar{D}})|\text{EPR}\rangle = (O_D^T \otimes I_{CD\bar{C}})|\text{EPR}\rangle$. However, this construction clearly depends on the initial state of the black hole.

Next, let us recall the argument for a young black hole where the early radiation R is absent. Consider a “typical” pure quantum state $|\Psi_{\text{young}}\rangle_{CD}$. For the sake of discussion, one may assume that $|\Psi_{\text{young}}\rangle_{CD}$ is drawn randomly from the Haar measure. If $|D| \ll |C|$, the mutual information $I(C, D)$ is nearly maximal (for typical states). Hence, given a unitary operator O_D , a partner operator \tilde{O}_C does exist:

$$(\tilde{O}_C \otimes O_D)|\Psi_{\text{young}}\rangle_{CD} \approx |\Psi_{\text{young}}\rangle_{CD}. \tag{8.4}$$

For a given $|\Psi_{\text{young}}\rangle_{CD}$ and O_D , let us fix \tilde{O}_C and ask if \tilde{O}_C is also a partner operator for some other typical pure quantum state $|\Psi_{\text{young}'}\rangle_{CD}$. The answer is clearly no since C and D are entangled in a way which depends on each quantum state.

A similar conclusion can be also obtained by considering an average over typical quantum states. Observe the following relation

$$\int d|\Psi\rangle \langle\Psi|_{CD}(\tilde{O}_C \otimes O_D)|\Psi\rangle_{CD} = \frac{1}{2^n} \text{Tr}(\tilde{O}_C \otimes O_D) \quad (8.5)$$

where the integral is taken over the Haar measure. If \tilde{O}_C were partner operators of O_D for $|\Psi\rangle_{CD}$, we would have $\langle\Psi|_{CD}(\tilde{O}_C \otimes O_D)|\Psi\rangle_{CD} \approx 1$. But the r.h.s. becomes zero for traceless unitary operators, such as Pauli operators. Hence one cannot choose the partner operators in a state-independent manner.³⁷

8.2 Observer-dependence

Our construction of the black hole interior operators avoids the above no-go arguments simply due to the fact that it does not give partner operators for the initial state of the black hole, but for a slightly perturbed state after the inclusion of the infalling observer. To see this point explicitly, let us recall how the interior operators were reconstructed after including the infalling observer, specifically the scenario discussed in section 7.1. Let $|\Phi(0)\rangle$ be the initial state of the black hole, which can be two-sided or one-sided (old or young). In section 7.1, we explicitly appended an infalling observer to the system:

$$|\Phi'(0)\rangle = |\Phi(0)\rangle \otimes |0\rangle_M. \quad (8.6)$$

After the inclusion of the infalling observer, the whole system time-evolves by some chaotic unitary U :

$$|\Phi'(t)\rangle = U(t)|\Phi'(0)\rangle. \quad (8.7)$$

What we have constructed, by the use of the Hayden-Preskill recovery protocol, are the interior partner operators for the quantum state $|\Phi'(t)\rangle$, not for the unperturbed original quantum state $|\Phi(t)\rangle$. As such, previous no-go arguments do not apply to our construction.

It is, however, worth emphasizing that we are not entirely free from dependence on specifics of the system. Namely, our construction manifestly depends on how the infalling observer is introduced to the system. For instance, if a different initial state of the infalling observer is chosen (say $|1\rangle$) in eq. (8.6), the expression of interior operators would be different. Hence, while our construction of interior operators does not depend on the initial state of the black hole $|\Phi(0)\rangle$, it does depend on *observers*.

Similar justifications hold for other scenarios of introducing infalling observers to the black hole. For instance, we have illustrated the use of an ancilla mode \bar{E} as a physical probe to the black hole. In this case, the reconstructed interior operators are of course the ones for perturbed states with such ancilla modes. It is unclear to us how to treat the inclusion of the infalling observers on a unified footing. But formally, this procedure can be expressed as an isometric quantum operation:

$$|\Phi'(0)\rangle = \Omega(|\Phi(0)\rangle). \quad (8.8)$$

³⁷Arguments along this line are also used as an evidence for firewalls in typical black hole microstates [9].

Here the isometry Ω accounts for all the quantum operations associated with the inclusion of the infalling observer. Generally speaking, Ω consists of some simple (low complexity) unitary operations as well as addition of ancilla systems. Obviously, $|\Phi'(0)\rangle$ depends on the choice of the initial state $|\Phi(0)\rangle$ as well as how the infalling observer is introduced, namely the map Ω . After the inclusion of the infalling observer, the black hole time-evolves by some chaotic dynamics: $|\Phi'(t)\rangle = U(t)|\Phi'(0)\rangle$. Our claim is that construction of interior operators for such a slightly perturbed black hole $|\Phi'(t)\rangle$ does not depend on the initial state $|\Phi(0)\rangle$, but depends on the map Ω . Discussions on actual expressions of interior operators for each possible scenario of introducing infalling observers will be given in a future work.

9 Resolution of AMPS puzzle

In this section, we present a resolution of the AMPS puzzle.

9.1 AMPS thought experiment

Let us consider a version of the AMPS thought experiment. Given an old black hole that is maximally entangled with the early radiation R , consider an outgoing mode D which has been just emitted from the black hole. The outgoing mode D is entangled with some degrees of freedom in the early radiation R . The outside observer Bob distills a qubit (or qubits) \tilde{D} that is entangled with the outgoing mode D by performing some careful quantum operation on R . Bob may isolate the qubit \tilde{D} from the black hole or hand it to Alice who is about to fall into a black hole. After crossing the horizon smoothly, Alice will see an interior mode \bar{D} which is entangled with the outgoing mode D . This, however, leads to a contradiction because D is also entangled with \tilde{D} .

The fallacy of the above argument is clear. When Alice falls into a black hole, the outgoing mode D is disentangled from the distilled qubit \tilde{D} due to the backreaction which makes OTOCs decay. Furthermore, Alice will observe the interior mode \bar{D} which is distinct from the original partner mode \tilde{D} and is independent of initial states of the black hole. Hence, the monogamy of quantum entanglement is not violated. Recall that the approaches bundled under “ $A = R_B$ ” try to interpret \tilde{D} and \bar{D} as the same physical degrees of freedom. Here we argued that they must be different. In fact, \bar{D} and \tilde{D} have no relevance at all as \bar{D} is state-independent. Furthermore, \bar{D} can be constructed in the same manner for young black holes.³⁸

³⁸The AMPS argument relies on three key assumptions; a) unitarity b) no drama (i.e., smooth horizon) and c) effective quantum field theory outside the horizon. Note that a) results from quantum mechanics while b) and c) result from general relativity, namely the equivalence principle. We have assumed unitary quantum evolution and argued that the horizon must be smooth. Hence, the assumption c) must be violated. The essential point in our argument is that the outgoing mode D is disentangled from the distilled qubit \tilde{D} . This conclusion was obtained from decay of OTOCs. Calculations of OTOCs generically go beyond naïve effective quantum field theory. In its simplest form at the massless limit, the perturbation becomes a gravitational shockwave and leaves significant backreaction to the geometry. Such a calculation goes beyond effective quantum field theory on a fixed background geometry.

It is worth comparing our proposal with the previous proposal by Maldacena and Susskind, an approach often called “ER = EPR”. They argue that the distillation of the qubit \tilde{D} , as well as quantum operations afterwards by Bob, creates perturbations which will become a high-energy density near the horizon. Furthermore, they argue that this perturbation spoils the quantum entanglement between D and \tilde{D} and creates a firewall. As such, Alice is not able to observe entanglement between D and \tilde{D} .

While we also concluded that D and \tilde{D} are not entangled due to perturbations, the underlying physical mechanism differs in a crucial manner. In our scenario, D and \tilde{D} are disentangled due to the backreaction by the infalling observer herself, not due to Bob. Also, despite the loss of entanglement between D and \tilde{D} , Alice still crosses the horizon smoothly and observe the entangled pair D and \bar{D} since \bar{D} is realized as a mode different from the original mode \tilde{D} . After observing the interior mode \bar{D} , Alice may be (or may not be) hit by a gravitational shockwave which Bob creates, but this has nothing to do with quantum entanglement between D and \bar{D} .

In summary, we propose the following resolution of the AMPS puzzle:

- 1) An infalling observer leaves non-trivial gravitational backreaction no matter how she falls. This disentangles the outgoing mode from the early radiation.
- 2) The infalling observer crosses the horizon smoothly and observes quantum entanglement between the outgoing mode and the interior mode.
- 3) The interior mode which the infalling observer sees is distinct from degrees of freedom that were originally entangled with the outgoing mode. Hence the monogamy of entanglement is not violated.
- 4) Reconstructions of interior operators are not unique. Different reconstructions of the interior operators correspond to different infalling trajectories that the infalling observer follows.

In figure 12, we summarized the original AMPS argument, previous proposals along the lines of ER = EPR, and our proposal.

9.2 Revisiting firewalls

We have argued that the infalling observer sees quantum entanglement while the outside observer does not due to the backreaction by the infalling observer. It is also intriguing to consider a possible scenario where the outside observer can see quantum entanglement while the infalling observer does not. We have already refuted previous proposals which suggested that verification of quantum entanglement, namely the distillation of the entangled qubit \tilde{D} on the early radiation R , generates a high-energy gravitational shockwave which prevents the infalling observer from crossing the horizon smoothly.

Here we propose a physical mechanism for Bob to protect quantum entanglement from the backreaction by Alice. Recall that Alice can see quantum entanglement by interacting with earlier mode A_t . Bob’s strategy, then, would be to retrieve A_t from Alice and bring

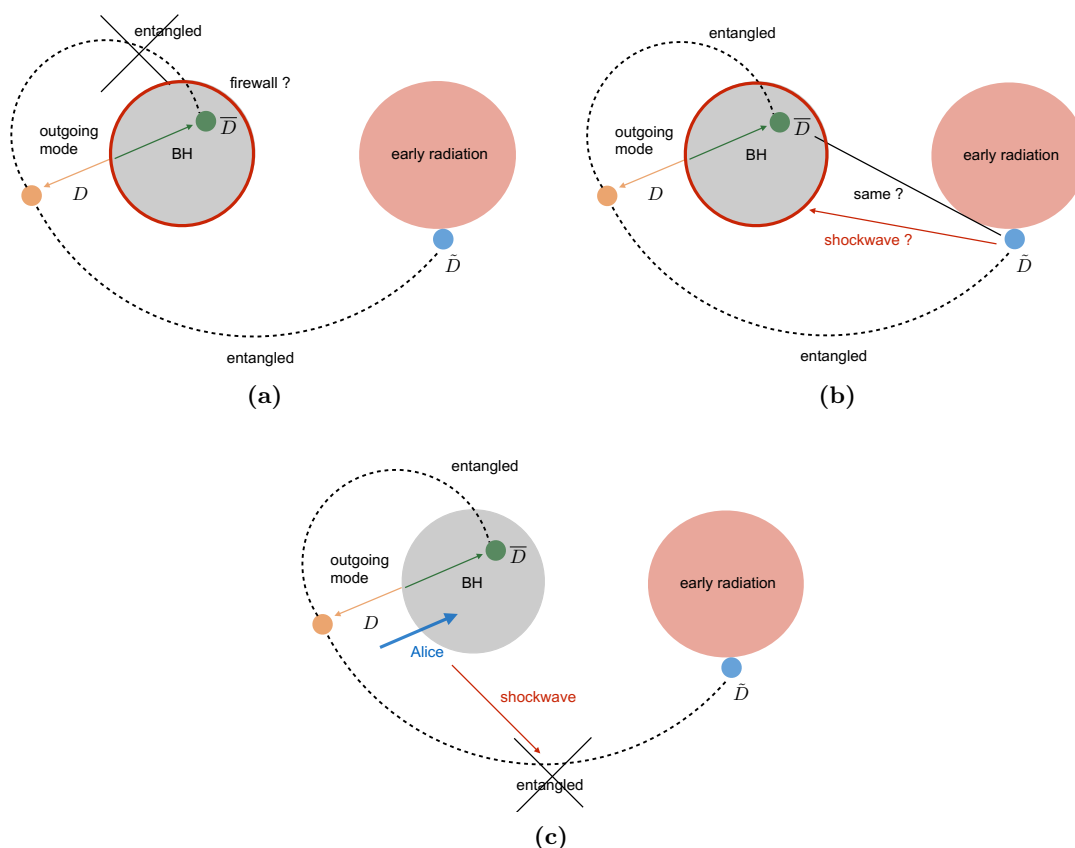


Figure 12. Schematic pictures of (a) the AMPS argument, (b) some previous proposals and (c) our proposal.

it back to his possession. One concrete (and peaceful) method is to run the Hayden-Preskill recovery protocol from the outside and pull Alice herself to the outside! That is, by preventing Alice from falling into a black hole, Bob is able to protect quantum entanglement between the outgoing mode and the early radiation. Since Alice does not cross the horizon, she will not see quantum entanglement.

At this point, it is pretty satisfying to recall how the recovery protocols of [31] work, reviewed in section 3. The first protocol proceeds by verifying the entanglement between the outgoing mode D and its partner \bar{D} . This is essentially Bob's attempt to observe the EPR pair between D and R by applying the EPR projection operator. The second protocol achieves this goal in a deterministic manner by unitarily restoring $D\bar{D}$ to be EPR pairs. Both protocols verify the EPR pair between D and R , and steal the EPR pair from Alice by preventing her from crossing the horizon.³⁹

³⁹The first protocol projects $D\bar{D}$ onto EPR pair. Since $I(D, \bar{D}) = \max$, we have $I(D, AC) = 0$ suggesting that \bar{D} cannot be reconstructed on the left hand side. Since $I(A, \bar{A}) \approx \max$, Alice traverses the wormhole. So, we conclude that Alice does not see the EPR pair and simply returns to the outside. If we project $D\bar{D}$ onto other entangled states, such as $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$, we still have $I(D, \bar{D}) = \max$ and $I(D, AC) = 0$. However, $I(A, \bar{A})$ is not necessarily large. Hence, we conclude that Alice does not see the EPR pair. And unfortunately, she does not return to the outside.

Based on this observation, we make the following proposal as physical interpretation of firewalls that may arise from the outside observer’s verification of quantum entanglement:

- 1) An outside observer creates a non-trivial backreaction for an infalling observer by verifying (or protecting) quantum entanglement between the outgoing mode D and the early radiation R .
- 2) However, it does not kill an infalling observer. Instead, it saves her from falling into a black hole recklessly.
- 3) Since an infalling observer does not cross the horizon, she will not be able to see quantum entanglement.

The backreaction from Bob’s verification of quantum entanglement is different from either high-energy barrier or some hard wall which would terminate the geometry at the horizon. It is worth noting that the Hayden-Preskill recovery, as well as the traversable wormhole, can be interpreted as a shockwave with negative energy.

9.3 Puzzle on non-local interactions (and its resolution)

One mysterious feature of the AdS/CFT correspondence is an apparent non-locality behind the horizon in the two-sided black hole. The CFT Hamiltonian on the boundary is given by $H = H_L + H_R$ and the Hilbert space factorizes into the left and right. (See [68] for problems related to the factorization of the Hilbert space). Imagine two observers who start from left and right boundaries respectively and meet inside the black hole. This can be achieved in principle by starting at sufficiently early time. Two observers may interact with each other when they meet. From the perspective of the boundary quantum mechanical descriptions, however, this seems to imply some non-local coupling between left and right.

We have encountered an analogous problem in the context of the AMPS puzzle where the infalling observer jumps into a black hole and sees the interior mode. Suppose that every possible representation of interior modes requires degrees of freedom on the right hand side. Suppose that she crosses the horizon smoothly and can observe the interior mode. This suggests that there must be some interaction between degrees of freedom on the left and right hand sides of the AdS black hole, leading to a contradiction.

The resolution of this non-locality puzzle can be obtained straightforwardly. By falling into a black hole, Alice leaves a backreaction and creates a replicated mode \bar{D} which consists exclusively of the left hand side degrees of freedom. Since she does not interact with \tilde{D} on the right hand side, there is no contradiction.⁴⁰ Our viewpoint is that the black hole

⁴⁰A common misguided approach to the non-locality problem in the two-sided AdS black hole goes as follows. “From the boundary perspective, whether Alice has interacted with the partner mode \tilde{D} on the other side or not cannot be verified from either boundary because the interaction would happen behind the horizon. Yet, the interaction between Alice and \tilde{D} can be seen by coupling two boundaries in a suitable manner, e.g. via the traversable wormhole effect. The interaction between Alice and \tilde{D} does not lead to contradiction because two boundaries need to be coupled for verification of the interaction”. We do not think this is a correct resolution of the non-locality problem. As we have rigorously demonstrated in this paper, the infalling observer will disentangle D from \tilde{D} and create a new interior mode \bar{D} exclusively on

interior for an infalling observer emerges due to scrambling dynamics, not due to quantum entanglement between the left and right hand sides. We hope to address this issue further elsewhere.

10 Discussions

In this paper, we argued that the AMPS puzzle does not lead to inconsistencies by presenting state-independent reconstructions of interior operators. It is an interesting future problem to verify the proposed scenario in concrete models of quantum black holes. We are currently working on writing the interior operators in an explicit manner, namely in the SYK model.

While we have focused on the AMPS puzzle in black hole horizons, it will be interesting to ask about cosmological horizons. In the dS space, a shockwave shifts the horizon in the opposite direction, making the Penrose diagram “taller”, in a way similar to sending a negative energy shockwave in the AdS space. As such, our proposed resolution for the firewall problem in black hole horizons does not apply to the problem in cosmological horizons. One possible scenario may be that the backreaction by an existing observer recedes the horizon away so that she cannot actually cross the cosmological horizon.

In the remainder of the paper, we present discussions and speculations on relevant topics.

10.1 Early time

Since OTOCs start to decay significantly around the scrambling time, Bob can verify quantum entanglement between the outgoing mode D and the early radiation R when $\Delta t \ll t_{\text{scr}}$ even if Alice has fallen into a black hole. This, however, suggests that Alice should not be able to observe quantum entanglement when crossing the horizon! For this reason, our proposal resolves the AMPS puzzle only when $\Delta t \gtrsim t_{\text{scr}}$. There are three possible physical explanations for the $D\bar{D}$ pair with $\Delta t \ll t_{\text{scr}}$.

The first explanation concerns the applicability of effective quantum field theory near the singularity. In the Penrose diagram of the AdS black hole, one notices that Alice meets \bar{D} very close to the singularity for $\Delta t \ll t_{\text{scr}}$. Hence the validity of quantum field theory is questionable. This observation seems to suggest that Alice may not be able to see quantum entanglement for small Δt .

The second explanation concerns the quality of quantum entanglement that Alice may be able to see. Recall that 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. As one goes away from the horizon, the density of thermal entropy becomes smaller. Namely, one needs to coarse-grain a larger volume in order to distill a single EPR pair. In order for Alice to meet the outgoing mode D near the horizon, she needs $\Delta t \gtrsim r_S \log r_S$ in the Schwartzschild black

her side of the boundaries. Namely, \tilde{D} has nothing to do with her infalling experience. Furthermore, we think that it is possible to see the interior mode \bar{D} via a careful quantum operation acting only on one side of the black hole, namely by performing the Hayden-Preskill recovery protocol on one side. See section 10.2 for further discussions.

hole which is of order the scrambling time. As such, we speculate that $\Delta t \gtrsim r_S \log r_S$ is necessary for Alice to see a good quality EPR pair.⁴¹

The third explanation is to assert that the loss of quantum entanglement for modes away from the horizon does not lead to significant violation of effective quantum field theory descriptions. The loss of entanglement in the Rindler modes near the horizon will create high energy density as entanglement is at short distance scale. For small Δt , however, the Rindler modes in question are separated by a long distance, and hence the loss of entanglement cannot be detected by simple local physical observables. Such tiny deviations may be superseded by e.g., curvature corrections.

See appendix C for further discussions on the reconstruction of interior operators for small Δt . The important point is that construction of \bar{D} is almost state-independent even when Δt is small. We think that this phenomena may be of independent interest in the context of quantum many-body physics.

10.2 Peeking in through the horizon

While we have used the Hayden-Preskill recovery protocols as a mathematical tool to reconstruct interior operators, it is intriguing to think of physically implementing them on a black hole as an actual physical process. Let us focus on the probabilistic reconstruction protocol outlined in section 4. The initial state of the black hole is given by $|\text{EPR}\rangle_{A'A}|\text{EPR}\rangle_{BR}$ where A' is the reference system of A . After the black hole evolves by a unitary operator U , the outgoing mode D is kept outside the black hole. We prepare an additional EPR pair. The half of the EPR pair is thrown into a black hole while the other half, denoted by \bar{D} , is kept outside the black hole. We then implement the inverse U^\dagger to the system. Finally, we project $\bar{A}'\bar{A}$ onto the EPR pair. This leaves $\bar{D}D$ to be a nearly ideal EPR pair. Figure 13 summarizes the reconstruction of the interior operator as a physical process.

Let us ponder over possible geometric interpretation of these protocols. Since $\bar{A}'\bar{A}$ is projected to the EPR pair, an input quantum state on A would remain invariant and reappear as the same quantum state on \bar{A} . This suggests that the infalling observer, who fell into a black hole, returns to the outside again safely.⁴² This is not surprising as we applied the inverse U^\dagger . What is surprising is that the outgoing mode D is entangled with another mode \bar{D} which is left outside the black hole. One possible interpretation is that, due to the backreaction of the Hayden-Preskill recovery protocol, the interior operator \bar{D} is somehow pulled to the outside of the black hole. Hence, the infalling observer appears to cross the horizon smoothly and returns safely with the interior mode in hand!

⁴¹This observation enables us to obtain “upper bound” on the scrambling time. If the scrambling time t_{scr} , in a sense of decay of OTOCs, is longer than $\approx r_S \log r_S$, Bob can verify the EPR pair when $\Delta t \approx r_S \log r_S$. However, Alice can also see the EPR pair, leading to the violation of monogamy of entanglement. As such, we arrive at $t_{\text{scr}} \lesssim r_S \log r_S$. Recalling that the Hayden-Preskill thought experiment says $t_{\text{scr}} \gtrsim r_S \log r_S$, one may conclude $t_{\text{scr}} \approx r_S \log r_S$.

⁴²More precisely, we applied a projection operator to pick an event where the infalling observer returns safely. This works only probabilistically. One may apply the deterministic version of the Hayden-Preskill recovery protocol to make this happen deterministically.

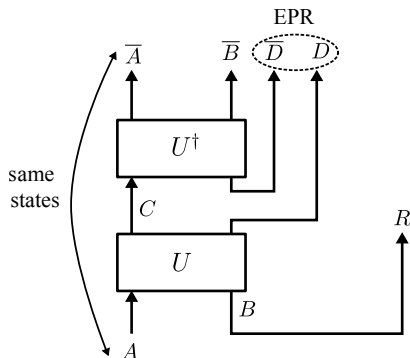


Figure 13. Reconstruction of the interior operator \bar{D} . If $D\bar{D}$ were to be measured in EPR pair, A and \bar{A} must be identical quantum states.

As this speculation suggests, reconstruction of the interior operators by the Hayden-Preskill recovery protocols enables us to probe the physics behind the horizon from the outside even when the black hole is one-sided. To turn the speculation into a concrete observation, it may be interesting to consider a similar process by using the Gao-Jafferis-Wall traversable wormhole.

10.3 Entanglement wedge reconstruction

We have presented a protocol to reconstruct interior operators by using the Hayden-Preskill recovery. Hence it is reasonable to expect that the Hayden-Preskill recovery is useful in reconstructing operators in the entanglement wedge. Progresses along this line have been recently made by Almheiri [18]. Since operators in the entanglement wedge must be interpreted as quantum field theory operators on a fixed geometry, state-independent reconstructions must be possible. We expect that our results are useful in achieving this goal.

Acknowledgments

I would like to thank Ahmed Almheiri, Yoni BenTov, Isaac Kim, Rob Myers and Herman Verlinde for useful discussions on related topics. I thank Yoni BenTov and Rob Myers for careful reading of the preliminary draft and a number of useful feedbacks. For the second version of the paper, I would like to thank Ning Bao, Raphael Bousso, Juan Maldacena, Xiaoliang Qi, Suvrat Raju, Lenny Susskind and Ying Zhao for useful discussions. Acknowledgment of discussions with a person does not necessarily imply the person’s agreement. 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.

A Energy measurement

We have presented a protocol to reconstruct the interior operator \bar{D} without involving the early radiation R . Our protocol requires EPR pairs in the ancilla system $E\bar{E}$ which

interact with the earlier mode A in a particular manner. Here we argue that much simpler and generic interaction can achieve this goal.

To set the stage, we elaborate on the issue briefly mentioned in section 2.1. Let us denote the orthonormal basis of the outgoing mode D by $|n\rangle \in \mathcal{H}_D$. According to the standard rules of statistical physics, the reduced density operator on D should look thermal:

$$\rho_D = \sum_n w_n |n\rangle\langle n|, \quad w_n = \frac{e^{-\beta E_n}}{Z} \tag{A.1}$$

with $Z = \sum_n e^{-\beta E_n}$. The discrepancy with ρ_D in eq. (2.2) originates from the total energy conservation. In particular, energy densities on C and D are correlated. An analysis with energy conservation taken into consideration indicates that ρ_{CD} sustains diagonal correlations in the $|n\rangle$ -basis while off-diagonal correlations decohere due to scrambling dynamics (i.e., correlations are classical) [26].

The above observation prompts us to consider two different classes of operators on D . The off-diagonal operators are the ones which may change the energy eigenstates (or the particle number) in D . The interior partners of those can be identified in C as operators which decrease and/or increase the energy density. This is due to diagonal correlations in ρ_{CD} . The diagonal operators are the ones which leave energy eigenstates unchanged up to phases. Such operators can be explicitly written as follows:

$$O_\theta = \sum_n e^{i\theta_n} |n\rangle\langle n|. \tag{A.2}$$

Here θ represents all the data about θ_n . Unlike off-diagonal operators, the interior partners of diagonal operators O_θ cannot be found in C .⁴³

Off-diagonal correlations can be restored by simply measuring the energy of the earlier mode A . The procedure of quantum measurement can be formulated as a unitary process in the following manner. We adjoin the system to a pointer system E whose dimensionality is equal to A . Denote the orthonormal basis by $|n\rangle_E$. If we start the system in an arbitrary pure state $|\psi\rangle = \sum_n c_n |n\rangle_A$, the measurement leads to the evolution:

$$|\psi\rangle_A |0\rangle_E \rightarrow \sum_n c_n |n\rangle_A |n\rangle_E \tag{A.3}$$

under the unitary operator

$$W = \sum_n |n\rangle\langle n|_A \otimes X_E^n \tag{A.4}$$

where $X = \sum_j |j+1\rangle\langle j|$ is the Weyl operator (a generalization of Pauli- X to multi-level systems).

⁴³Off-diagonal partners exist due to the existence of diagonal correlations. Diagonal partners do not exist due to the absence of off-diagonal correlations.

The important point here is that applying M allows Alice to “copy” O_θ from A to E . This can be most explicitly shown by noting the following identity:

$$\text{Diagram (A.5)} \tag{A.5}$$

which holds for any density operator σ_A . Hence, from the outside description, all Alice needs to do is measure the energy (or the particle number) of the earlier mode A . The interior operator \bar{D} can be reconstructed in the pointer E and the remaining black hole C .

B Verlinde-Verlinde proposal

Verlinde and Verlinde proposed an intriguing resolution of the AMPS puzzle by employing the idea of quantum error-correction [22]. To illustrate the idea in a simplified setting, suppose that a black hole B is entangled with the early radiation R only in the codeword subspace $\mathcal{H}_{B_0} \subset \mathcal{H}_B$ with $\mathcal{H}_{B_0} = \{|\tilde{j}\rangle\}_{j=1}^{|B_0|}$. Specifically, consider a quantum state given by

$$|\Psi\rangle_{BR} = \frac{1}{\sqrt{|B_0|}} \sum_{j=1}^{|B_0|} |\tilde{j}\rangle_B \otimes |\tilde{j}\rangle_R \tag{B.1}$$

where $|B_0| \ll |B|$. Here, by the “codeword subspace”, we simply mean that $|B_0| < |B|$. If the black hole evolves by a Haar random unitary, the mutual information $I(C, D)$ is nearly maximal, implying that \bar{D} can be reconstructed on C . Indeed, they performed a careful statistical physics analysis and presented explicit constructions of interior operators by using the error recovery procedure known in quantum error-correction theory. A related idea was considered [69] in a slightly different context.

In our interpretation, however, this observation itself does not resolve the AMPS puzzle. First, this proposal does not apply to black holes which have maximal entanglement such as the eternal AdS black hole. Let S_{BH} be the coarse-grained entropy of the black hole and S_{ent} be the entanglement entropy between the black hole and the radiation R . Then the black holes under consideration satisfy $S_{\text{BH}} > S_{\text{ent}}$ whereas the eternal AdS black hole corresponds to the case with $S_{\text{BH}} = S_{\text{ent}}$. The situation of eq. (B.1) resembles black holes before the Page time. If such a codeword subspace emerges, its origin should be explained. Second, the proposed reconstruction of interior operators depends on choices of the codeword subspace, and as such, is state-dependent. Third, as pointed out in [26], this proposal runs into another problem where the Hayden-Preskill recovery cannot be performed. Indeed, if the outgoing mode D can be used for the Hayden-Preskill recovery, then it should not be entangled with C .

Nevertheless, we will make an interesting link between our proposal and their viewpoint. Let us view the energy measurement discussed in the appendix A as a process of

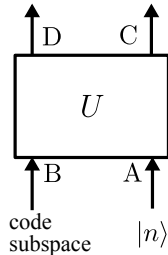


Figure 14. A black hole entangled through a codeword subspace. An infalling observer measures $|n\rangle$ on A , and quantum entanglement between C and D is restored.

“collapsing” A into some pure state without introducing the pointer system. If a particular energy eigenstate $|n\rangle$ is measured, the postselected black hole dynamics can be interpreted as an isometry from B to CD :

$$U|0\rangle_A : \mathcal{H}_B \rightarrow \mathcal{H}_C \otimes \mathcal{H}_D \quad |B| < |C||D|. \quad (\text{B.2})$$

An alternative interpretation of this dynamics is to think that the black hole is represented by $S_{\text{BH}} = \log_2 |A||B|$ qubits, but its entanglement entropy is $S_{\text{ent}} = \log_2 |B|$. In other words, measurement on A reduced the entanglement entropy of the black hole. Here B is viewed as the subspace where the black hole is entangled with the partner. This situation resembles the scenario considered by Verlinde and Verlinde [22]. By having $S_{\text{ent}} < S_{\text{BH}}$, quantum entanglement can be restored between C and D . We have treated a similar situation in section 7 by including a measurement apparatus M .

In fact, collapsing the input Hilbert space by generic projective measurements on A restores quantum entanglement between C and D . For simplicity of discussion, let us assume that U is Haar random. An explicit calculation based on Haar random unitary suggests that projective measurements on A with $|A| \gtrsim |D|^2$ leads to nearly maximal mutual information $I(C, D)$ [26]. When thermal correlations are taken into account, only the off-diagonal correlations need to be restored, so taking $|A| \gtrsim |D|$ suffices. The important implication of this observation is that (essentially) any measurement by an infalling observer effectively creates a situation where the black hole is entangled through a codeword subspace. This restores quantum entanglement in C and D , and does not require any fine-tuning.

The work by Verlinde and Verlinde did not provide concrete account for the physical origin of the codeword subspace. We think that inclusion of the infalling observer effectively creates situations similar to theirs.

C Fault-tolerant entanglement

The state-independent reconstruction of interior operators may be of independent interest from the perspective of quantum many-body physics as it reveals a certain universal aspect of quantum entanglement in chaotic quantum dynamics.

Let us consider an initial quantum state of the following form:

$$\rho_{\text{in}} = |\psi\rangle\langle\psi|_A \otimes \sigma_B \tag{C.1}$$

where $|\psi\rangle$ is a fixed pure state on A and σ_B is an arbitrary quantum state, pure or mixed, on B . The system evolves under a chaotic unitary dynamics U :

$$\rho_{\text{out}} = U\rho_{\text{in}}U^\dagger. \tag{C.2}$$

Here we make two assumptions. First, we assume that OTOCs for $O_D(t)$ and $O_A = |\psi\rangle\langle\psi|$ decay to small values. Second, we assume that the system thermalizes, in a sense that time-ordered correlation functions, such as $\langle O_D(t)O_A(0)\rangle$, decay to small values. From these assumptions, we can find that D is nearly maximally entangled with its complement C as long as $|A| \gtrsim |D|^2$. Here our primary interest is in how C and D are entangled.

In this paper, we argued that the partner operators can be constructed on C in a way independent of σ_B . This implies that D is entangled with some fixed degrees of freedom, denoted by \bar{D} inside C , regardless of the choice of the initial state σ_B . Here \bar{D} may be interpreted as a code subspace where entanglement with D is fault-tolerantly stored in such a way that is protected from any perturbations on σ_B . In other words, this subspace is determined by the dynamics U and $|\psi\rangle_A$, but is independent of σ_B . As such, the entanglement structure between C and D is a universal feature of the chaotic dynamics U .

The above observation assumed that OTOCs have decayed to small values. Next, let us consider situations where OTOCs have not decayed to smaller values yet, e.g., before the scrambling time. Given an operator O_D on D , let us reconstruct a partner operator $O_{\bar{D}}$ supported on C by using the Hayden-Preskill recovery protocol. Consider the maximally mixed state $\sigma_B = \frac{1}{d_B}I_B$ as the initial state. Since C and D are entangled only weakly, we will have

$$\alpha \equiv \langle O_D \otimes O_{\bar{D}} \rangle_{\sigma_B = \frac{1}{d_B}I_B} \ll 1 \tag{C.3}$$

where the expectation value is taken with respect to ρ_{out} . Note that the expectation value would be close to 1 if C and D were nearly maximally entangled.

Despite the fact that O_D and $O_{\bar{D}}$ are only weakly correlated, the construction of $O_{\bar{D}}$ is (almost) state-independent in the following sense. To be specific, let us consider pure states $\sigma_B = |\phi\rangle\langle\phi|$ on B as initial states. We then typically have

$$\int d|\phi\rangle \langle O_D \otimes O_{\bar{D}} \rangle_{\sigma_B = |\phi\rangle\langle\phi|} = \alpha \tag{C.4}$$

where the integral is taken uniformly over all $|\phi\rangle$ on B . Hence, if $|\phi\rangle$ is a typical pure state sampled from the Haar measure (or an ensemble forming 2-design), then we have

$$\langle O_D \otimes O_{\bar{D}} \rangle_{\sigma_B = |\phi\rangle\langle\phi|} \approx \alpha \tag{C.5}$$

where the deviation from α is exponentially suppressed. As such, the structure of entanglement between D and \bar{D} is universal even when OTOCs have not decayed to small values yet.

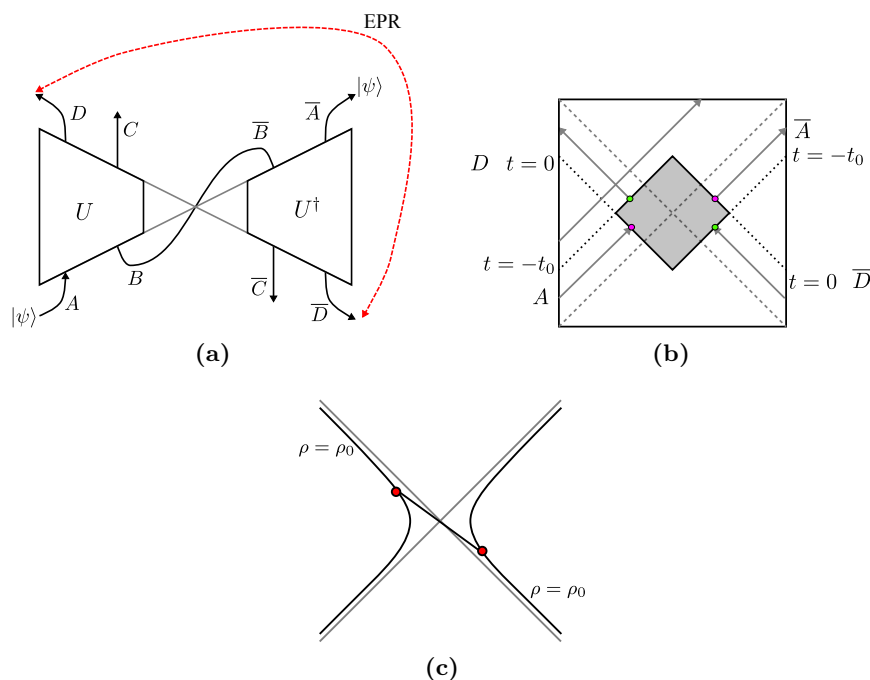


Figure 15. (a) The Hayden-Preskill recovery protocol. (b) Backreaction from the Hayden-Preskill recovery. (c) Alice’s quantum field theory near the horizon. Two surfaces in the opposite regions are identified.

D Alice’s quantum field theory

We discuss effective quantum field theory that the infalling observer may experience. According to the equivalence principle, we expect that Alice will have normal quantum mechanical descriptions until she reaches the singularity. The fundamental difficulty in addressing this question lies in the very fact that Alice never returns to the outside. To circumvent this, we may think of performing the Hayden-Preskill recovery protocol from the outside so that Alice can have effective quantum field theory without reaching the singularity. To understand quantum field theory under the effect of the Hayden-Preskill recovery, we must understand its effect on the black hole geometry. At this moment, we are unsure about what is the corresponding geometric picture. It is even unclear to us if it admits any classical geometric description or not. Below we present some speculation on the backreacted geometry.

Let us focus on the first recovery protocol from [31] which projects $D\bar{D}$ into the EPR pair. To gain some insights, it is useful to draw the process in the Penrose diagram of the eternal two-sided AdS black hole. Let us take the time directions to be upward and downward on left and right hand sides respectively. One may flip the diagram of the recovery procedure in accordance with the convention of the Penrose diagram. The flipped diagram is shown in figure 15a. This cartoon picture enables us to depict the recovery protocol schematically in the Penrose diagram as in figure 15b. Here two modes $D\bar{D}$ on two sides are collapsed into EPR pairs. This lets A traverse across the black hole and reach the other boundary at \bar{A} . The backreaction by the Hayden-Preskill recovery

becomes significant when the separation of the time between the infalling observer A and the outgoing mode D becomes of order the scrambling time. In the Penrose diagram, let us take D and \bar{D} at $t = 0$ on the left and right hand sides respectively. The geometry will be significantly modified only inside the shaded diamond shown in figure 15b which is constructed by considering modes at $t = -t_0$ with $t_0 \approx t_{\text{scr}}$. If a signal is sent between $t = -t_0$ and $t = 0$, it will cross the horizon and reach the singularity. If a signal is sent before $t = -t_0$, it will traverse to the other side. These observations suggest that the portion inside the diamond effectively disappears, and antipodal points should be identified as shown in figure 15b so that signals entering the diamond can traverse. Precise value of t_0 will depend on the size of D as well as the amount of quantum information in A we wish to transmit.⁴⁴

Suppose that Alice jumps into a black hole at $t < -t_0$. She will enter the shaded region near the horizon which is approximated by the Rindler space. Letting ρ be the distance from the horizon, Alice sees the backreaction from the Hayden-Preskill recovery at the surface $\rho = \rho_0$. The recovery protocol prevents Alice from seeing EPR pairs at $\rho \leq \rho_0$ and lets her traverse to the opposite region. This situation can be realized by cutting the Rindler space at $\rho = \rho_0$ and glue two regions together. If the time separation between A and D is of order the scrambling time, ρ_0 is small compared to the black hole radius, but is still large compared to the Planck length. We speculate that Alice's effective quantum field theory is defined on this truncated and traversable Rindler space. If the separation becomes larger than the scrambling time, ρ_0 eventually becomes of order the Planck length. In this regime, we expect that the effect of the Hayden-Preskill recovery can be simply treated as identification of opposite sides by neglecting physics below the Planck scale.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] S.W. Hawking, *Breakdown of Predictability in Gravitational Collapse*, *Phys. Rev. D* **14** (1976) 2460 [[INSPIRE](#)].
- [2] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black Holes: Complementarity or Firewalls?*, *JHEP* **02** (2013) 062 [[arXiv:1207.3123](#)] [[INSPIRE](#)].
- [3] S.W. Hawking, *Particle Creation by Black Holes*, *Commun. Math. Phys.* **43** (1975) 199 [*Erratum ibid.* **46** (1976) 206] [[INSPIRE](#)].
- [4] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford and J. Sully, *An Apologia for Firewalls*, *JHEP* **09** (2013) 018 [[arXiv:1304.6483](#)] [[INSPIRE](#)].

⁴⁴For simplicity of discussion, consider a deeply scrambling regime with $t_0 \gtrsim t_{\text{scr}}$. An analysis from [31] suggests that $|D| \gtrsim |A|$ suffices to transmit all the quantum information contained inside A . If quantum information in the subspace $A_0 \subset A$ are to be transmitted, we only need $|D|^2 \gtrsim |A||A_0|$. So, even when A is a big subsystem, $|D|^2 \gtrsim \text{const} \cdot |A|$ suffices if A_0 is small. In a qubit count, the sufficient condition would be $n_D \gtrsim \frac{1}{2}n_A$.

- [5] K. Papadodimas and S. Raju, *An Infalling Observer in AdS/CFT*, *JHEP* **10** (2013) 212 [[arXiv:1211.6767](#)] [[INSPIRE](#)].
- [6] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortsch. Phys.* **61** (2013) 781 [[arXiv:1306.0533](#)] [[INSPIRE](#)].
- [7] L. Susskind, *New Concepts for Old Black Holes*, [arXiv:1311.3335](#) [[INSPIRE](#)].
- [8] R. Bousso, *Complementarity Is Not Enough*, *Phys. Rev. D* **87** (2013) 124023 [[arXiv:1207.5192](#)] [[INSPIRE](#)].
- [9] D. Marolf and J. Polchinski, *Gauge/Gravity Duality and the Black Hole Interior*, *Phys. Rev. Lett.* **111** (2013) 171301 [[arXiv:1307.4706](#)] [[INSPIRE](#)].
- [10] R. Bousso, *Firewalls from double purity*, *Phys. Rev. D* **88** (2013) 084035 [[arXiv:1308.2665](#)] [[INSPIRE](#)].
- [11] S.B. Giddings, *Nonviolent nonlocality*, *Phys. Rev. D* **88** (2013) 064023 [[arXiv:1211.7070](#)] [[INSPIRE](#)].
- [12] D. Harlow, *Jerusalem lectures on black holes and quantum information*, *Rev. Mod. Phys.* **88** (2016) 015002 [[arXiv:1409.1231](#)] [[INSPIRE](#)].
- [13] R. Bousso, *Violations of the Equivalence Principle by a Nonlocally Reconstructed Vacuum at the Black Hole Horizon*, *Phys. Rev. Lett.* **112** (2014) 041102 [[arXiv:1308.3697](#)] [[INSPIRE](#)].
- [14] D. Marolf and J. Polchinski, *Violations of the Born rule in cool state-dependent horizons*, *JHEP* **01** (2016) 008 [[arXiv:1506.01337](#)] [[INSPIRE](#)].
- [15] K. Papadodimas and S. Raju, *Black Hole Interior in the Holographic Correspondence and the Information Paradox*, *Phys. Rev. Lett.* **112** (2014) 051301 [[arXiv:1310.6334](#)] [[INSPIRE](#)].
- [16] K. Papadodimas and S. Raju, *State-Dependent Bulk-Boundary Maps and Black Hole Complementarity*, *Phys. Rev. D* **89** (2014) 086010 [[arXiv:1310.6335](#)] [[INSPIRE](#)].
- [17] I. Kourkoulou and J. Maldacena, *Pure states in the SYK model and nearly-AdS₂ gravity*, [arXiv:1707.02325](#) [[INSPIRE](#)].
- [18] A. Almheiri, *Holographic Quantum Error Correction and the Projected Black Hole Interior*, [arXiv:1810.02055](#) [[INSPIRE](#)].
- [19] D. Harlow, *Aspects of the Papadodimas-Raju Proposal for the Black Hole Interior*, *JHEP* **11** (2014) 055 [[arXiv:1405.1995](#)] [[INSPIRE](#)].
- [20] A. Almheiri, X. Dong and D. Harlow, *Bulk Locality and Quantum Error Correction in AdS/CFT*, *JHEP* **04** (2015) 163 [[arXiv:1411.7041](#)] [[INSPIRE](#)].
- [21] 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 [[arXiv:1503.06237](#)] [[INSPIRE](#)].
- [22] E. Verlinde and H. Verlinde, *Black Hole Entanglement and Quantum Error Correction*, *JHEP* **10** (2013) 107 [[arXiv:1211.6913](#)] [[INSPIRE](#)].
- [23] E. Verlinde and H. Verlinde, *Behind the Horizon in AdS/CFT*, [arXiv:1311.1137](#) [[INSPIRE](#)].
- [24] E. Verlinde and H. Verlinde, *Passing through the Firewall*, [arXiv:1306.0515](#) [[INSPIRE](#)].
- [25] A. Goel, H.T. Lam, G.J. Turiaci and H. Verlinde, *Expanding the Black Hole Interior: Partially Entangled Thermal States in SYK*, *JHEP* **02** (2019) 156 [[arXiv:1807.03916](#)] [[INSPIRE](#)].

- [26] B. Yoshida, *Soft mode and interior operator in Hayden-Preskill thought experiment*, [arXiv:1812.07353](#) [INSPIRE].
- [27] P. Hayden and J. Preskill, *Black holes as mirrors: Quantum information in random subsystems*, *JHEP* **09** (2007) 120 [[arXiv:0708.4025](#)] [INSPIRE].
- [28] P. Hosur, X.-L. Qi, D.A. Roberts and B. Yoshida, *Chaos in quantum channels*, *JHEP* **02** (2016) 004 [[arXiv:1511.04021](#)] [INSPIRE].
- [29] J. Maldacena, S.H. Shenker and D. Stanford, *A bound on chaos*, *JHEP* **08** (2016) 106 [[arXiv:1503.01409](#)] [INSPIRE].
- [30] D. Harlow and P. Hayden, *Quantum Computation vs. Firewalls*, *JHEP* **06** (2013) 085 [[arXiv:1301.4504](#)] [INSPIRE].
- [31] B. Yoshida and A. Kitaev, *Efficient decoding for the Hayden-Preskill protocol*, [arXiv:1710.03363](#) [INSPIRE].
- [32] D.N. Page, *Average entropy of a subsystem*, *Phys. Rev. Lett.* **71** (1993) 1291 [[gr-qc/9305007](#)] [INSPIRE].
- [33] D.N. Page, *Information in black hole radiation*, *Phys. Rev. Lett.* **71** (1993) 3743 [[hep-th/9306083](#)] [INSPIRE].
- [34] S.D. Mathur, *The Information paradox: A Pedagogical introduction*, *Class. Quant. Grav.* **26** (2009) 224001 [[arXiv:0909.1038](#)] [INSPIRE].
- [35] S.L. Braunstein, S. Pirandola and K. Życzkowski, *Better Late than Never: Information Retrieval from Black Holes*, *Phys. Rev. Lett.* **110** (2013) 101301 [[arXiv:0907.1190](#)] [INSPIRE].
- [36] L. Susskind, *Black Holes and Complexity Classes*, [arXiv:1802.02175](#) [INSPIRE].
- [37] S. Aaronson, *The Complexity of Quantum States and Transformations: From Quantum Money to Black Holes*, [arXiv:1607.05256](#) [INSPIRE].
- [38] J. Oppenheim and B. Unruh, *Firewalls and flat mirrors: An alternative to the AMPS experiment which evades the Harlow-Hayden obstacle*, *JHEP* **03** (2014) 120 [[arXiv:1401.1523](#)] [INSPIRE].
- [39] G. 't Hooft, *The black hole interpretation of string theory*, *Nucl. Phys.* **B 335** (1990) 138 [INSPIRE].
- [40] G. 't Hooft, *On the Quantum Structure of a Black Hole*, *Nucl. Phys.* **B 256** (1985) 727 [INSPIRE].
- [41] L. Susskind and L. Thorlacius, *Gedanken experiments involving black holes*, *Phys. Rev.* **D 49** (1994) 966 [[hep-th/9308100](#)] [INSPIRE].
- [42] L. Susskind, L. Thorlacius and J. Uglum, *The Stretched horizon and black hole complementarity*, *Phys. Rev.* **D 48** (1993) 3743 [[hep-th/9306069](#)] [INSPIRE].
- [43] N. Lashkari, D. Stanford, M. Hastings, T. Osborne and P. Hayden, *Towards the Fast Scrambling Conjecture*, *JHEP* **04** (2013) 022 [[arXiv:1111.6580](#)] [INSPIRE].
- [44] A. Kitaev, unpublished.
- [45] A. Almheiri and J. Polchinski, *Models of AdS₂ backreaction and holography*, *JHEP* **11** (2015) 014 [[arXiv:1402.6334](#)] [INSPIRE].

- [46] J. Maldacena and D. Stanford, *Remarks on the Sachdev-Ye-Kitaev model*, *Phys. Rev. D* **94** (2016) 106002 [[arXiv:1604.07818](#)] [[INSPIRE](#)].
- [47] J. Engelsöy, T.G. Mertens and H. Verlinde, *An investigation of AdS_2 backreaction and holography*, *JHEP* **07** (2016) 139 [[arXiv:1606.03438](#)] [[INSPIRE](#)].
- [48] D. Stanford, J. Maldacena and Z. Yang, *Conformal symmetry and its breaking in two-dimensional nearly anti-de Sitter space*, *Prog. Theor. Exp. Phys.* **2016** (2016) 12C104 [[arXiv:1606.01857](#)] [[INSPIRE](#)].
- [49] P. Gao, D.L. Jafferis and A. Wall, *Traversable Wormholes via a Double Trace Deformation*, *JHEP* **12** (2017) 151 [[arXiv:1608.05687](#)] [[INSPIRE](#)].
- [50] J. Maldacena, D. Stanford and Z. Yang, *Diving into traversable wormholes*, *Fortsch. Phys.* **65** (2017) 1700034 [[arXiv:1704.05333](#)] [[INSPIRE](#)].
- [51] A. Larkin and Y.N. Ovchinnikov, *Quasiclassical method in the theory of superconductivity*, *JETP* **28** (1968) 1200.
- [52] S.H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *JHEP* **03** (2014) 067 [[arXiv:1306.0622](#)] [[INSPIRE](#)].
- [53] D.A. Roberts, D. Stanford and L. Susskind, *Localized shocks*, *JHEP* **03** (2015) 051 [[arXiv:1409.8180](#)] [[INSPIRE](#)].
- [54] D.A. Roberts and B. Yoshida, *Chaos and complexity by design*, *JHEP* **04** (2017) 121 [[arXiv:1610.04903](#)] [[INSPIRE](#)].
- [55] S. Beigi, *Sandwiched Rényi divergence satisfies data processing inequality*, *J. Math. Phys.* **54** (2013) 122202.
- [56] B. Yoshida and N.Y. Yao, *Disentangling Scrambling and Decoherence via Quantum Teleportation*, *Phys. Rev. X* **9** (2019) 011006 [[arXiv:1803.10772](#)] [[INSPIRE](#)].
- [57] Y. Sekino and L. Susskind, *Fast Scramblers*, *JHEP* **10** (2008) 065 [[arXiv:0808.2096](#)] [[INSPIRE](#)].
- [58] P.W. Shor, *Scrambling Time and Causal Structure of the Photon Sphere of a Schwarzschild Black Hole*, [arXiv:1807.04363](#) [[INSPIRE](#)].
- [59] T. Dray and G. 't Hooft, *The Effect of Spherical Shells of Matter on the Schwarzschild Black Hole*, *Commun. Math. Phys.* **99** (1985) 613 [[INSPIRE](#)].
- [60] G. 't Hooft, *Strings from gravity*, *Phys. Scripta* **T 15** (1987) 143 [[INSPIRE](#)].
- [61] Y. Kiem, H. Verlinde and E. Verlinde, *Black hole horizons and complementarity*, *Phys. Rev. D* **52** (1995) 7053 [[hep-th/9502074](#)] [[INSPIRE](#)].
- [62] K. Sfetsos, *On gravitational shock waves in curved space-times*, *Nucl. Phys. B* **436** (1995) 721 [[hep-th/9408169](#)] [[INSPIRE](#)].
- [63] Y. BenTov and J. Swarngin, *Gravitational shockwaves on rotating black holes*, *Gen. Rel. Grav.* **51** (2019) 25 [[arXiv:1706.03430](#)] [[INSPIRE](#)].
- [64] A. Kitaev and S.J. Suh, *The soft mode in the Sachdev-Ye-Kitaev model and its gravity dual*, *JHEP* **05** (2018) 183 [[arXiv:1711.08467](#)] [[INSPIRE](#)].
- [65] Y. Gu and A. Kitaev, *On the relation between the magnitude and exponent of OTOCs*, *JHEP* **02** (2019) 075 [[arXiv:1812.00120](#)] [[INSPIRE](#)].

- [66] M. Srednicki, *Chaos and quantum thermalization*, *Phys. Rev. E* **50** (1994) 888 [[cond-mat/9403051](#)].
- [67] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys. Rev. Lett.* **96** (2006) 181602 [[hep-th/0603001](#)] [[INSPIRE](#)].
- [68] D. Harlow, *Wormholes, Emergent Gauge Fields and the Weak Gravity Conjecture*, *JHEP* **01** (2016) 122 [[arXiv:1510.07911](#)] [[INSPIRE](#)].
- [69] N. Bao, S.M. Carroll, A. Chatwin-Davies, J. Pollack and G.N. Remmen, *Branches of the Black Hole Wave Function Need Not Contain Firewalls*, *Phys. Rev. D* **97** (2018) 126014 [[arXiv:1712.04955](#)] [[INSPIRE](#)].
- [70] K. Papadodimas and S. Raju, *Local Operators in the Eternal Black Hole*, *Phys. Rev. Lett.* **115** (2015) 211601 [[arXiv:1502.06692](#)] [[INSPIRE](#)].