Unitarity of black hole evaporation in final-state projection models

Seth Lloyd
Department of Mechanical Engineering, MIT, Cambridge MA 02139, USA

John Preskill
Institute for Quantum Information and Matter, Caltech, Pasadena CA 91125, USA

Abstract

Almheiri *et al.* have emphasized that otherwise reasonable beliefs about black hole evaporation are incompatible with the monogamy of quantum entanglement, a general property of quantum mechanics. We investigate the final-state projection model of black hole evaporation proposed by Horowitz and Maldacena, pointing out that this model admits cloning of quantum states and polygamous entanglement, allowing unitarity of the evaporation process to be reconciled with smoothness of the black hole event horizon. Though the model seems to require carefully tuned dynamics to ensure exact unitarity of the black hole S-matrix, for a generic final-state boundary condition the deviations from unitarity are exponentially small in the black hole entropy. We argue that an observer inside the black hole need not detect any deviations from standard quantum mechanics, and explain how verifying the entanglement between the early and late radiation emitted by an old black hole may create particles that can be detected by infalling observers who cross the event horizon. Final-state projection models illustrate how inviolable principles of standard quantum mechanics might be circumvented in a theory of quantum gravity.

1 Introduction

The quantum physics of black holes has caused great puzzlement since Stephen Hawking discovered [1] nearly 40 years ago that black holes evaporate. The crux of the puzzle is this: if a pure quantum state collapses to form a black hole, the geometry of the evaporating black hole contains spacelike surfaces crossed by both the collapsing body inside the event horizon and nearly all of the emitted Hawking radiation outside the event horizon. If this process is unitary, then the quantum information encoded in the collapsing matter must also be encoded (perhaps in a highly scrambled form) in the outgoing radiation; hence the infalling quantum state is *cloned* in the radiation, violating the linearity of quantum mechanics.

This puzzle has spawned many audacious ideas, beginning with Hawking's bold proposal [2] that unitarity fails in quantum gravity. Efforts to rescue unitary led to the formulation of black hole complementarity [3, 4], the notion that the inside and outside of a black hole are not really two separate subsystems of a composite quantum system, but rather two complementary views of the same system, related by a complex nonlocal map. Black hole complementary set the stage for the holographic principle [5, 6], and its eventual realization in AdS/CFT duality [7], which provides a pleasingly unitary picture of black hole evaporation in asymptotically AdS spacetimes, though the implications of this duality regarding the black hole interior remain unclear.

Black hole complementarity seeks to reconcile three reasonable beliefs: (1) An evaporating black hole scrambles quantum information without destroying it. (2) A freely falling observer encounters nothing unusual upon crossing the event horizon of a black hole. (3) An observer who stays outside a black hole detects no violations of relativistic effective quantum field theory. But Almheiri *et al.* (AMPS) recently argued [8] that these three assumptions are incompatible. They consider the Hawking radiation B emitted by a black hole which is nearly maximally entangled with an exterior system R. (For example, R could be the radiation so far emitted by an old black hole which has already radiated away more than half of its initial entropy [9].) Assumptions (1) and (3) require B to be highly entangled with a subsystem R_B of R, while assumption (2) requires B to be highly entangled with a subsystem A in the black hole interior. Taken together, then, the three assumptions violate the principle of monogamy of entanglement [10, 11], which asserts that if quantum systems A and B are maximally entangled, then neither can be correlated with any other system. This tension between unitarity and monogamy had been noted earlier in [12, 13].

Like Hawking's original black hole information loss puzzle, the AMPS puzzle has also spawned audacious ideas. AMPS themselves advocated relaxing assumption (2), arguing that an old black hole (and perhaps also a young one) has a singular horizon (a *firewall*) and no interior [8, 14]. Another possibility is that modifications of assumption (3) allow the entanglement of B with A to be transferred to entanglement of B with R_B as B propagates away from the black hole [15]. Or, clinging to a revised version of the complementarity principle, one can assert that R_B should be regarded as a complementary description of A [16, 17], possibly connected to the black hole interior via a wormhole [18]. All of these ideas will need to be fleshed out further before they can be accurately assessed.

Here we suggest another possible response to the AMPS puzzle, based on the final-state projection model of black hole evaporation proposed [19] by Horowitz and Maldacena (HM). In this scenario, the S-matrix relating the asymptotic incoming state of the collapsing matter and asymptotic outgoing state of the emitted radiation can be unitary; however unitarity can be temporarily violated during the black hole evaporation process, accommodating violations of monogamy of entanglement and the no-cloning principle [20, 21], and allowing assumptions (1), (2), and (3) to be reconciled. A type of black hole complementarity is realized, and there is no need for firewalls. Just as with other proposed ways to resolve the AMPS puzzle, the HM proposal requires further development before it can be fairly assessed, but we do not regard it as *a priori* much more outlandish than these other proposals.

HM proposed imposing a final-state boundary condition requiring a particular quantum state at the spacelike singularity inside the black hole, which allows information to escape from the black hole interior by *postselected teleportation*. Speaking fancifully, information residing in the collapsing matter propagates from past infinity to the spacelike singularity inside the black hole, where it is scrambled and reflected, then propagates backward in time from the singularity to the horizon, and forward in time from the horizon to future infinity. More concretely, HM consider the composite system $\mathcal{H}_M \otimes \mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{out}}$, where $\mathcal{H}_M$ is the Hilbert space of the infalling matter, $\mathcal{H}_{\text{in}}$ is the Hilbert space of infalling negative energy Hawking radiation behind the horizon, and $\mathcal{H}_{\text{out}}$ is the Hilbert space of outgoing positive energy Hawking radiation outside the horizon. What appears to be the vacuum to a freely falling observer crossing the horizon is a maximally entangled state of $\mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{out}}$, and the HM boundary condition projects onto a particular maximally entangled state of $\mathcal{H}_M \otimes \mathcal{H}_{\text{in}}$, which encodes the black hole S-matrix. While the horizon crosser sees nothing out of the ordinary, an observer who stays outside the black hole finds that the state of the infalling matter and the state of the outgoing radiation are related by a unitary map.

The HM proposal has the appealing feature that the new physics responsible for evading information loss occurs at the singularity, where we expect semiclassical physics to fail badly. Fur-

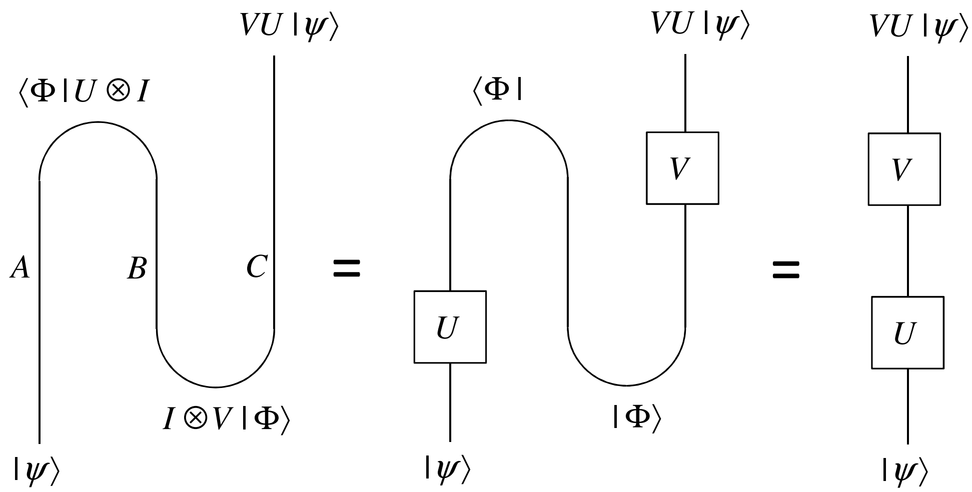

Figure 1: Quantum teleportation. To convey a quantum state $|\psi\rangle$ of system A to system C, first a maximally entangled state $|\Phi(V)\rangle$ of BC is prepared, and then AB is projected to a maximally entangled state $|\Phi(U^*)\rangle$. To recover $|\psi\rangle$, a party at C applies the unitary transformation $U^\dagger V^\dagger$.

thermore, if we are willing to impose initial-state boundary conditions at spacelike singularities in cosmological spacetimes [22], it may not be unreasonable to impose final-state boundary conditions at spacelike singularities in black hole spacetimes as well. But the proposal has other less pleasing features [23, 24]. In particular, unless appropriate constraints are imposed on the dynamics, postselected quantum mechanics can be afflicted with effective closed timelike curves and other causality paradoxes [25, 26, 27]; the dynamics may need to be carefully adjusted to protect the unitarity of the evaporation process.

Our attitude is that these potential bugs in the HM proposal may actually be welcome, helping to steer us toward a deeper understanding of quantum gravity. Therefore, we focus on delineating sufficient conditions for the proposal to work. In brief, we find that the evaporation process is unitary if the interactions between $\mathcal{H}_{\text{in}}$ and other systems are appropriately tuned. A deeper understanding of quantum gravity may be needed to decide whether black hole evaporation really fulfills these conditions, but we find that for a generic final-state boundary condition at the singularity, the deviations from exact unitarity scale like $e^{-S_{BH}/2}$ where S_{BH} is the black hole entropy. Such exponentially small violations of unitarity could well be regarded as a success for our semiclassical analysis of the HM model, since nonperturbative quantum gravity corrections of that order are expected and are beyond the scope of the analysis. We also argue, again assuming a generic final-state boundary condition, that deviations from standard quantum theory are unlikely to be detected by infalling observers as they approach the singularity.

Even if it turns out that the HM model is not realized in nature, the model is still quite instructive. It cautions us that inviolable consequences of standard quantum mechanics, such as the no-cloning principle and monogamy of entanglement, need not be respected in quantum gravity. Perhaps that is the proper lesson to be drawn from the AMPS puzzle.

After reviewing the HM model in Sec. 2, we comment on its relevance to the AMPS controversy in Sec. 3. In Sec. 4 we examine how unitarity of the black hole S-matrix might fail in the HM model, concluding that, for a generic final-state boundary condition at the singularity, the deviations from exact unitarity are exponentially small in the black hole entropy. In Sec. 5 we consider the

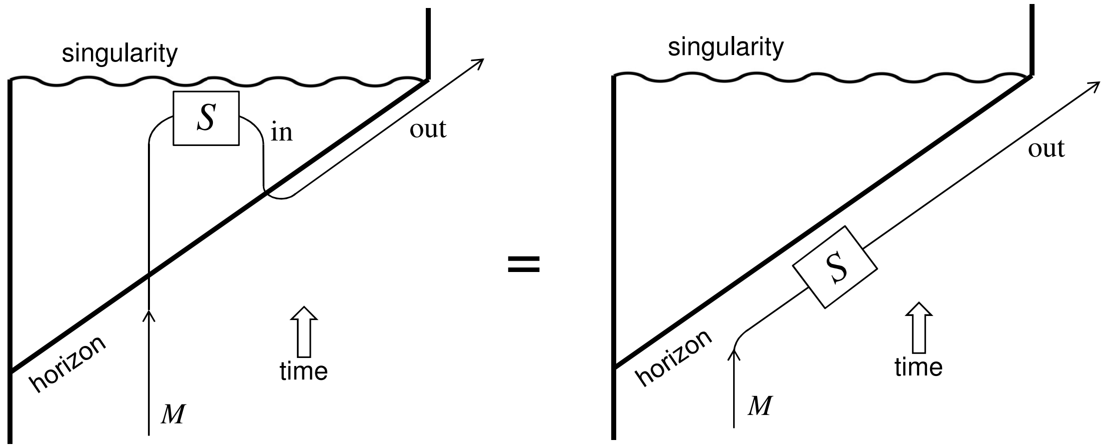

Figure 2: The Horowitz-Maldacena model, in which quantum information carried by the collapsing matter system M is teleported out of a black hole. Outgoing Hawking radiation is maximally entangled with infalling radiation, and a final-state boundary condition projects M and the infalling radiation to a maximally entangled state which encodes the unitary S-matrix S.

implications of postselection for observers inside the black hole, arguing that observers with limited access to the infalling Hawking radiation need not detect any deviations from standard quantum mechanics, and we explain how verifying the entanglement between the early and late radiation emitted by an old black hole can create particles which are visible to infalling observers who cross the horizon. Sec. 6 contains some concluding comments.

Connections between the AMPS puzzle and the HM model have also been discussed in [17, 28].

2 The Horowitz-Maldacena proposal

The HM proposal is based on quantum teleportation [29], which is illustrated in Fig. 1. Any maximally entangled pure state of two d-dimensional systems A and B can be expressed as

$$|\Phi(V)\rangle \equiv (I \otimes V)\,|\Phi\rangle = \left(V^T \otimes I\right)|\Phi\rangle. \tag{1}$$

Here $|\Phi\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^{d} |i\rangle_A \otimes |i\rangle_B$, where $\{|i\rangle_A\}$, $\{|i\rangle_B\}$ denote orthonormal bases, V is a unitary $d \times d$ matrix, and V^T is the transpose of V. To teleport the state $|\psi\rangle$ from A to C, we first prepare the entangled state $|\Phi(V)\rangle_{BC}$ of system BC, then perform an entangled measurement on AB. If the outcome of the measurement is $|\Phi(U^*)\rangle$, then up to normalization the state of C becomes

$$\left({}_{AB}\langle\Phi(U^*)|\Phi(V)\rangle_{BC}\right)|\psi\rangle_A = V_C\left({}_{AB}\langle\Phi|\Phi\rangle_{BC}\right)U_A|\psi\rangle_A = \frac{1}{d}VU|\psi\rangle_C, \tag{2}$$

where the factor $1/d$ indicates that the measurement outcome $|\Phi(U^*)\rangle$ occurs with probability $1/d^2$. Once known, this outcome can be transmitted to C by classical communication, and if the initial entangled state of BC is also known, then a party at C can apply $U^\dagger V^\dagger$ to recover the state $|\psi\rangle$ in system C. If either the initial state of BC or the projected state of AB were not maximally entangled, then either V or U would be non-unitary and hence unphysical; in that case the teleportation process would have imperfect fidelity.

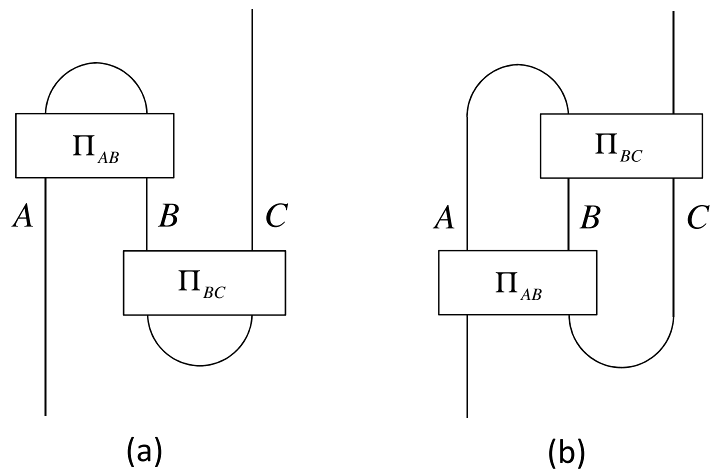

Figure 3: Polygamous entanglement in postselected teleportation. (a) Entanglement verifying projectors Π_{AB} and Π_{BC} both succeed with probability 1. (b) When their order is reversed, each projector succeeds with probability $1/d^2$, where d is the dimension of the teleported system.

In the HM proposal depicted in Fig. 2, quantum information is teleported from the collapsing matter system $\mathcal{H}_M$, the source for the black hole's classical geometry, to the outgoing Hawking radiation system $\mathcal{H}_{\rm out}$ that is emitted as the black hole evaporates. The dimension d is the number of distinguishable microstates for a black hole with specified total mass. Because the final-state boundary condition specifies that only one particular maximally entangled state is accepted at the singularity, there is no need for classical communication to convey the outcome of the entangled measurement.

The initial maximally entangled state used in the protocol is the Unruh state $|\Phi\rangle_{\rm in\otimes out}$, which looks like the vacuum state to a freely falling observer who crosses the horizon. Here $\mathcal{H}_{\rm in}$ is a system of infalling Hawking quanta behind the horizon. We use a microcanonical description, summing over all microstates with approximately the same energy, so that this state is maximally entangled rather than thermal. (The microcanonical ensemble is appropriate if we wish to consider the formation and evaporation of a black hole with sharply defined energy; of course, an observer with access to a small subsystem of $\mathcal{H}_{\rm out}$ will see a thermal state.) By a suitable basis choice, we set the unitary matrix specifying this maximally entangled state to the identity.

Loosely speaking, the basis state $|i\rangle_{\rm in}$ of $\mathcal{H}_{\rm in}$ is the negative energy Hawking state behind the horizon paired with the positive energy Hawking state $|i\rangle_{\rm out}$ outside the black hole. "Negative energy" is really a misnomer, because the timelike Killing vector of the exterior geometry becomes spacelike behind the horizon; hence "energy" inside the black hole is really momentum. In any case this description of the Unruh state is not precise because the evaporating black hole is not static and has no Killing vector. We take it for granted, though, that the notion of a maximally entangled state of $\mathcal{H}_{\rm in} \otimes \mathcal{H}_{\rm out}$ can be made precise.

If the entangled state of $\mathcal{H}_M \otimes \mathcal{H}_{\rm in}$ specified by the final-state boundary condition is $|\Phi(S^*)\rangle_{M\otimes {\rm in}}$, where S is unitary, then the infalling matter state and the outgoing radiation state are related by $|\varphi\rangle_{\rm out} = S|\psi\rangle_M$; thus S is the black hole S-matrix, presumed to be a highly nonlocal scrambling unitary transformation. S is required to rigorously satisfy conservation of energy and other exact gauge charges, but it need not respect global symmetries, which are expected to be broken in quantum gravity.

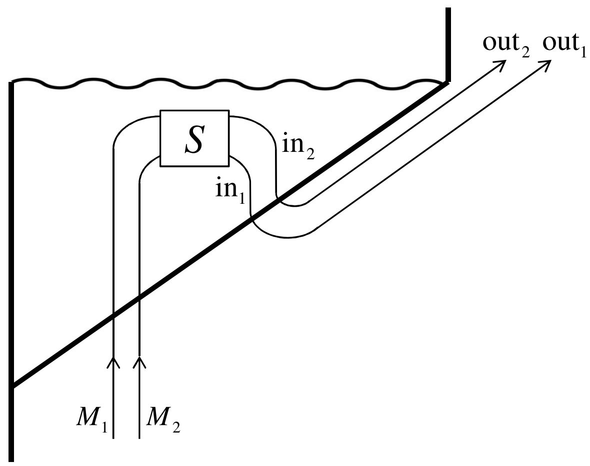

Figure 4: Polygamous entanglement in the HM model. The unitary transformation S entangles out_1 and out_2 with one another, yet both out_1 and out_2 are simultaneously entangled with their partner systems in_1 and in_2 behind the horizon.

Though analytically extended non-Schwarzschild black hole geometries can have timelike rather than spacelike singularities, the interior geometries of these solutions are unstable [30], and we assume the singularity is always spacelike and unavoidable in realistic collapse scenarios. Because the final-state boundary condition accepts any quantum state of the infalling matter system, observers approaching the singularity, particularly those with access to only a local subsystem, need not experience a reversal in the arrow of time or any departure from the usual laws of quantum mechanics.

3 Features of the model

As Fig. 2 indicates, the HM model supports a characteristic flow of information in spacetime, which ensures the unitarity of the black hole evaporation process. Information initially encoded in the collapsing matter flows forward in time from past infinity to the spacelike singularity, then backward in time from the singularity to the horizon, and finally forward in time from the horizon to future infinity. Despite the apparently acausal propagation backward in time, there is an equivalent description of the same process with a conventional causal ordering; the information flow can be "pulled tight" to "straighten out" the bends in the flow. This alternative description can be strictly justified only if the infalling radiation system $\mathcal{H}_{\mathrm{in}}$ is perfectly isolated from $\mathcal{H}_{\mathrm{out}}$ and $\mathcal{H}_M$, which may not be precisely true; therefore In Sec. 4 we will revisit the sufficient conditions for unitarity in a more general setting. But for now we will assume that the information flow admits a consistent causal ordering, and consider some of the consequences.

Once straightened, the overall process clearly preserves quantum information, with the unitary matrix S appearing in the final-state boundary condition playing the roll of the S-matrix relating the asymptotic incoming and outgoing states. But at intermediate times anomalous phenomena can occur, which would be disallowed in standard unitary quantum mechanics. For example, as Fig. 2 illustrates, cloning of quantum states can occur in postselected quantum mechanics. The quantum information encoded in $\mathcal{H}_M$ is also available, albeit in a highly scrambled form, in $\mathcal{H}_{\mathrm{out}}$ on the same spacelike slice. From the perspective of the causally ordered straightened process, the cloned state in the outgoing radiation is merely the same as the state of the infalling matter, except viewed at a later "time" and in a different basis.

Fig. 3 illustrates how monogamy of entanglement can be circumvented in postselected quantum mechanics. Monogamy [10, 11], a property of a quantum state with three parts A, B, and C, means that B can become highly entangled with A only at the cost of reducing its entanglement with C, and in particular that B can be maximally entangled with A only if it is uncorrelated with C. Conditioned on the postselected outcome of a final-state projection, however, monogamy may fail. As shown in Fig. 3a, if a maximally entangled state of AB is postselected, and Π denotes a projector onto a particular maximally entangled state, then the entanglement verifying projectors Π_{AB} and Π_{BC} may both succeed with probability one if the BC projector is applied before the AB projector. However, if the order of the projectors is reversed as in Fig. 3b, then each projector succeeds with probability $1/d^2$, where d is the dimension of the teleported system. Thus, the outcome of an entangled measurement on AB can be influenced by whether an entangled measurement on BC will be performed in the future. This sort of "causality paradox" is characteristic of postselected quantum mechanics [26, 27].

Fig. 4 illustrates polygamous entanglement in the HM model. Here two subsystems M_1 and M_2 of $\mathcal{H}_M$, initially unentangled, interact via the unitary transformation S encoded in the final-state boundary condition, resulting in an entangled state of two subsystems out_1 and out_2 of $\mathcal{H}_{\mathrm{out}}$. We may regard these two radiation subsystems as the early and late radiation emitted by an old black hole. We see that although out_1 and out_2 are entangled with one another, both out_1 and out_2 are simultaneously entangled with their partner systems in_1 and in_2 behind the horizon, violating monogamy. Postselection, then, can reconcile unitarity of black hole evaporation with smoothness of the horizon and conventional local physics outside the horizon.

Causality violation like that portrayed in Fig. 3b could arise if a measurement verifying the in-out entanglement might be performed after a measurement verifying the out_1-out_2 entanglement. To avoid potential closed timelike curves and protect unitarity we must restrict the entangled measurements that straddle the event horizon; we will return to this point in Sec. 4.

Black hole complementarity is realized in the HM model in the sense that observables inside and outside the horizon acting on the same spacelike slice do not commute. From the perspective of the causally ordered information flow, this failure of commutativity is expected, because the outside observables act on the same system as the inside observables, but at a later "time."

We may also consider a process in which we continually feed a black hole with additional matter to maintain its mass for a long time compared to its natural evaporation time, before finally allowing the evaporation to proceed to completion. In that case the S-matrix S, rather than being an arbitrary unitary transformation mapping the infalling matter to the outgoing radiation, must have a special structure enforced by the requirement that the entropy of the radiation should never exceed the Bekenstein-Hawking entropy of the black hole, if the overall state is pure. The information processing can be described by a quantum circuit whose bounded width is determined by the black hole entropy as in Fig. 5, and in particular the final-state boundary condition will respect the requirement that information cannot escape from the evaporating black hole before it falls in. If this circuit scrambles rapidly [31], then the "information mirror" phenomenon [32] will occur, in which, for a black hole highly entangled with its surroundings, information absorbed by the black hole returns in the emitted radiation after a Schwarzschild time $O(m \log m)$, where m is the black hole mass. We note that if additional mass is thrown into the black hole after it initially forms, to avoid firewalls we require a smooth Unruh vacuum at the apparent horizon, not at the global horizon whose position depends on the future history of the hole.

Quantum computation with final-state projection is known to be PP-complete [33]. Hence general final-state projection models allow very hard computational problems to be solved "efficiently"

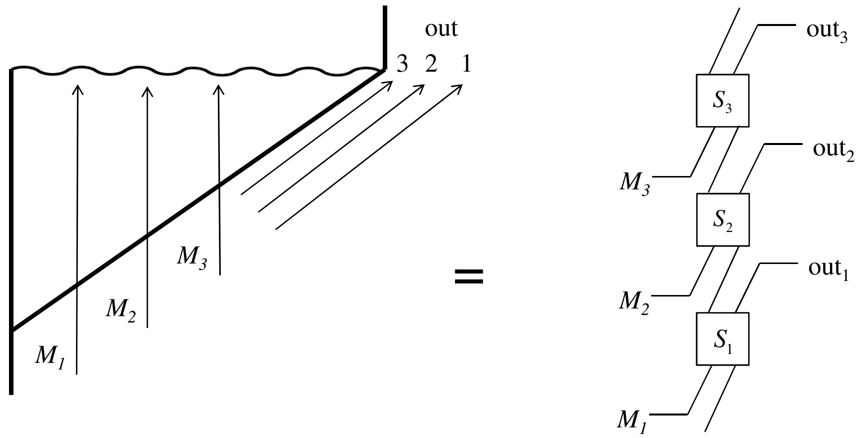

Figure 5: Information flow for a black hole that maintains its mass by accreting a steady stream of infalling matter.

(in particular, PP contains the complexity class NP, the class of problems for which a solution can be efficiently verified using a classical computer). Hence it is important to emphasize that the HM model admits only a restricted kind of postselection. The features that enforce unitarity of black hole evaporation (or an excellent approximation to unitarity), which we discuss in Sec. 4, seem to rule our using the final-state projection inside black holes for solving hard problems with unreasonable efficiency. Though it is still an open question whether quantum gravity can be simulated efficiently with a standard quantum computer, so it is at least possible in principle based on current knowledge that quantum gravity computers can solve problems which are beyond the reach of standard quantum computers, we see no reason why the computational power of the HM model should exceed that of other quantum gravity models.

4 Conditions for unitarity

The discussion in Sec. 3 was premised on the assumption that the postselected information flow in spacetime has a consistent causal ordering, ensuring the unitarity of the evaporation process. In general, though, interactions among the systems $\mathcal{H}_M$, $\mathcal{H}_{\rm in}$ and $\mathcal{H}_{\rm out}$ might disrupt this ordering; will the black hole S-matrix be unitary in that case?

4.1 Entanglement across the horizon

One important criterion for unitarity concerns the entangled state of $\mathcal{H}_{\rm in} \otimes \mathcal{H}_{\rm out}$ which is used as a resource in postselected teleportation.

Assuming $|\mathcal{H}_{\rm in}| = |\mathcal{H}_{\rm out}| = d$ (where $|\mathcal{H}|$ denotes the dimension of $\mathcal{H}$) we say that an entangled state of $\mathcal{H}_{\rm in} \otimes \mathcal{H}_{\rm out}$ is "full rank" if the marginal density operator on $\mathcal{H}_{\rm in}$ (and hence also $\mathcal{H}_{\rm out}$) has d nonzero (possibly degenerate) eigenvalues. Any full-rank bipartite entangled state can be expressed as $(U \otimes I)\, |\Phi\rangle$, where $|\Phi\rangle$ is a canonical maximally entangled state, and U is invertible (though not necessarily unitary).

If the initial state used in postselected teleportation is the full-rank entangled state $|\Psi\rangle_{\rm in\otimes out} = (I \otimes U)\, |\Phi\rangle_{\rm in\otimes out}$, and the final-state boundary condition projects onto ${}_{M\otimes {\rm in}}\langle\Theta| = {}_{M\otimes {\rm in}}\langle\Phi|\, (U^{-1}S \otimes I)$,

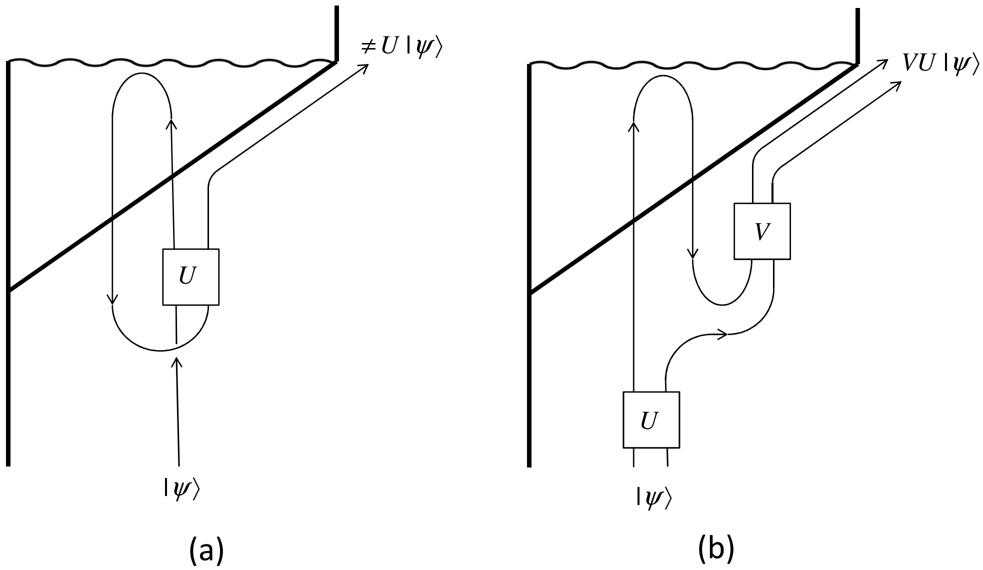

Figure 6: (a) Signaling from the outgoing Hawking radiation to the infalling matter can violate unitarity. (b) Signaling from the infalling matter to the outgoing radiation is consistent with unitarity.

where S is unitary, then the black hole S-matrix will be S. Neither the initial state nor the postselected state is maximally entangled, but the non-maximal entanglement of the $\langle\Theta|$ compensates perfectly for the non-maximally entanglement of $|\Psi\rangle$, resulting in overall unitarity. We see that the state at the apparent horizon need not be maximally entangled to ensure the unitarity of postselected teleportation, as long as the final-state condition is adjusted appropriately.

However, as we will discuss in Sec. 4.3 below, it seems natural to conjecture that the postselected state is in some sense generic, which means that ${}_{M\otimes\mathrm{in}}\langle\Theta|$ is likely to be very close to maximally entangled. In that case, unitarity demands that $|\Psi\rangle_{\mathrm{in}\otimes\mathrm{out}}$ be very nearly maximally entangled as well.

We have another reason to demand a high degree of entanglement for the state $|\Psi\rangle_{\mathrm{in}\otimes\mathrm{out}}$: a freely falling observer crossing the apparent horizon should see a smooth vacuum state rather than a seething firewall. Smoothness at the horizon requires the state to closely resemble the Unruh state, in which a mode localized outside the horizon which has sharply defined frequency with respect to Schwarzschild time is entangled with its Hawking partner behind the horizon, such that the reduced density operator of either mode is thermal when its partner is traced out. If the state that collapses to form a black hole has nearly definite energy, then we presume that the reduced density operator on $\mathcal{H}_{\mathrm{out}}$ for the global state of this Unruh vacuum is nearly maximally mixed — it is essentially the microcanonical ensemble in a narrow energy band, whose purification is a nearly maximally entangled state on $\mathcal{H}_{\mathrm{in}}\otimes\mathcal{H}_{\mathrm{out}}$. Actually, the compatibility of the mode-by-mode thermal entanglement (required for smoothness of the horizon) with the near maximal entanglement of the global state (required for unitarity) is a delicate quantitative issue which we find hard to resolve decisively; related issues were discussed in [34]. For the rest of our discussion, we will just assume that the initial state of $\mathcal{H}_{\mathrm{in}} \otimes \mathcal{H}_{\mathrm{out}}$ is maximally entangled, though n Sec. 4.4 we will revisit the robustness of this assumption with respect to interactions between $\mathcal{H}_{\mathrm{in}}$ and $\mathcal{H}_{\mathrm{out}}$.

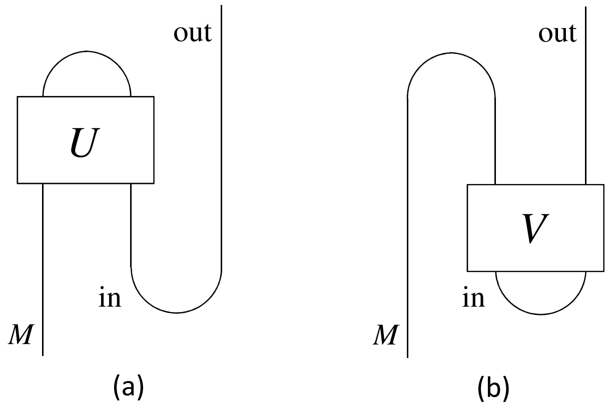

Figure 7: Entangling interactions between the infalling radiation and the collapsing matter (a), or between the infalling radiation and the outgoing radiation (b), may compromise the fidelity of postselected teleportation.

4.2 Signaling the infalling matter

Unitarity could be violated by interactions between $\mathcal{H}_M$ and $\mathcal{H}_{\text{out}}$, as shown in Fig. 6a. We can forbid such violations by demanding that the outgoing Hawking radiation is unable to send signals to the infalling matter encoding the same state, which are received either before or after the matter crosses the horizon. This is just the very natural requirement that information cannot be emitted in the radiation before it falls into the black hole; otherwise the information flow would admit effective closed timelike curves. Since the HM boundary condition at the singularity is highly nonlocal it could in principle violate this condition, but such violations are highly unlikely if the black hole S-matrix is a generic scrambling unitary transformation, as we discuss below in Sec. 4.3.

On the other hand, it is perfectly all right for the infalling matter to send signals to the outgoing Hawking radiation that encodes the same state, as in Fig. 6b. You can send a signal to your own clone as it emerges from the black hole without violating unitarity.

4.3 A generic final state

Barring such signaling from the outgoing radiation to the infalling matter, the only remaining threat to unitarity arises from interactions between $\mathcal{H}_{\text{in}}$ and the other systems. In the teleportation circuit, quantum information effectively flows backward in time in $\mathcal{H}_{\text{in}}$, and interactions of such chronology violating systems with chronology respecting systems can be dangerous, inducing closed timelike curves, and hence failure of unitarity [26, 27]. Put more prosaically, entangling interactions behind the horizon between $\mathcal{H}_{\text{in}}$ and $\mathcal{H}_M$, as in Fig. 7a, compromise the fidelity of teleportation, because in effect $\mathcal{H}_M \otimes \mathcal{H}_{\text{in}}$ will not be projected onto a maximally entangled state. Likewise, entangling interactions between $\mathcal{H}_{\text{in}}$ and $\mathcal{H}_{\text{out}}$, as in Fig. 7b, also cause trouble because in effect the state of $\mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{out}}$ used in the teleportation protocol will not be maximally entangled.

Let's first assume that the state of $\mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{out}}$ is exactly maximally entangled, and consider the consequences of entangling interactions between $\mathcal{H}_{\text{in}}$ and $\mathcal{H}_M$ behind the horizon, as in Fig. 7a. Intriguingly, if the final-state projection is chosen generically, or equivalently if the unitary transformation U in Fig. 7a acting on $\mathcal{H}_M \otimes \mathcal{H}_{\text{in}}$ is sampled uniformly with respect to the invariant Haar measure, then the evaporation process is very, very nearly, though not quite exactly, unitary.

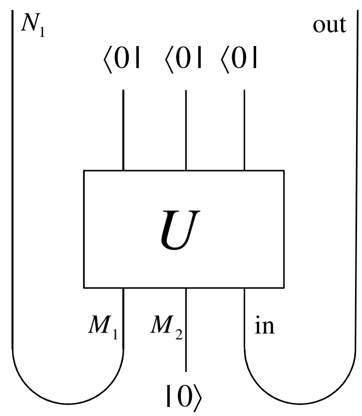

Figure 8: The HM model for a generic final-state boundary condition. Subsystem M_1 of the collapsing matter system M is maximally entangled with a reference system N_1, and $\mathcal{H}_M \otimes \mathcal{H}_{\rm in}$ is projected onto a Haar-random state determined by the unitary transformation U. In the resulting postselected state, N_1 is very nearly maximally entangled with a subsystem of the outgoing Hawking radiation.

A black hole with mass m has entropy $O(m^2)$ and evaporation time $O(m^3)$. The vast majority of ways of making a black hole look like the time-reversed evaporation process and require a time $O(m^3)$. Black holes created rapidly, in time $O(m)$, have entropy $O(m^{3/2})$, and hence have many fewer possible microstates than generic black holes. Analysis of the creation and evaporation of a generic black hole may be subtle, because substantial evaporation occurs while the black hole is still being assembled. Let's focus instead on the case where the black hole forms rapidly. We divide the Hilbert space of the infalling matter into two subsystems, $\mathcal{H}_M = \mathcal{H}_{M_1} \otimes \mathcal{H}_{M_2}$, where the states in $\mathcal{H}_{M_1}$ collapse rapidly; hence $|\mathcal{H}_{M_1}|/|\mathcal{H}_M| = \exp\left(-O(m^2)\right) \ll 1$, where $|\mathcal{H}|$ denotes the dimension of the Hilbert space $\mathcal{H}$.

For the purpose of analyzing whether quantum information initially carried by the rapidly collapsing matter system $\mathcal{H}_{M_1}$ can be decoded from the outgoing Hawking radiation $\mathcal{H}_{\rm out}$, it is convenient to ask what happens when M_1 is maximally entangled with a reference system N_1 as shown in Fig. 8. We assume that subsystem M_2 starts out in a fixed state, *e.g.*, its vacuum state. After the final-state projection, a random pure state $|\Psi(U)\rangle_{N_1 \otimes {\rm out}}$ on $\mathcal{H}_{N_1} \otimes \mathcal{H}_{\rm out}$ is obtained, which depends on the unitary transformation U that defines the postselected state of $\mathcal{H}_M \otimes \mathcal{H}_{\rm in}$. Tracing out the radiation system we obtain a mixed marginal state $\rho_{N_1}(U)$ on N_1, and by averaging over U we find [35]

$$\int dU \|\rho_{N_1}(U) - \rho_{N_1}^{\rm max}\|_1 \leq \sqrt{\frac{|\mathcal{H}_{M_1}|}{|\mathcal{H}_{\rm in}|}} \approx \exp\left(-S_{BH}/2 + O(m^{3/2})\right); \tag{3}$$

here $\|\cdot\|_1$ denotes the L_1-norm, dU is the normalized Haar measure on the unitary group, $\rho_{N_1}^{\rm max}$ is the maximally mixed state on N_1, and $S_{BH} = \ln |\mathcal{H}_{\rm in}|$ is the black hole entropy. Thus the typical state on N_1 is extremely close to maximally mixed.

Since the overall state of $\mathcal{H}_{N_1} \otimes \mathcal{H}_{\rm out}$ is pure, that ρ_{N_1} is almost maximally mixed means that the reference system N_1 is almost maximally entangled with a subsystem of the outgoing Hawking radiation, and correspondingly that a unitary decoding map acting on $\mathcal{H}_{\rm out}$ can isolate this subsystem which almost purifies ρ_{N_1}. It follows that for a Haar-typical final-state projection,

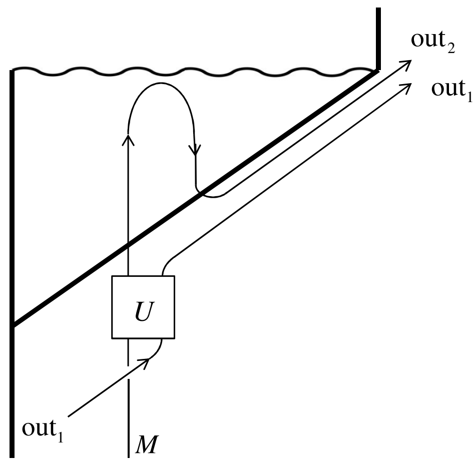

Figure 9: An agent crossing the event horizon of a black hole interacts first with the outgoing radiation outside the horizon and then with the infalling radiation inside the horizon (this latter interaction is not shown), thus applying an entangling transformation to the joint system. Unitarity of the black hole S-matrix is robust against such entangling interactions, provided that the final-state boundary condition has been tuned to ensure unitarity for arbitrary states of the infalling matter.

an arbitrary initial state of M_1 can be decoded in the outgoing Hawking radiation with a fidelity deviating from one by just $\exp\left(-O(m^2)\right)$. A similar conclusion would still apply if the unitary U were sampled from a unitary 2-design rather than the Haar measure, a sampling task which (unlike sampling from Haar measure) can be achieved exactly by a relatively small quantum circuit with size $O(m^4)$, or approximately with error ϵ by circuits with depth $O\left(\log m \log(1/\epsilon)\right)$ [36].

Nearly perfect unitarity is gratifying, but exact unitarity is what we yearn for. To ensure exact unitarity, we must restrict the form of the initial and final entangled states in the HM model, as well as the interactions of $\mathcal{H}_{\rm in}$ with infalling matter behind the horizon. This necessary fine-tuning in the model has been criticized [23, 24], but one might instead regard it as a tantalizing hint about quantum gravitational dynamics. Surely, that generic final-state projections come so close to achieving unitarity enhances the plausibility of the dynamical constraints we demand. Violations of unitarity scaling like $e^{-S_{BH}/2}$ could well be artifacts of the semiclassical framework used in the formulation of the HM model, as nonperturbative quantum gravitational corrections of that order are expected. Furthermore, information loss at such a tiny scale would be exceedingly difficult to detect "in practice," even if we disregard the complexity of decoding the highly scrambled Hawking radiation [37]. Indeed, the deviation from exact unitarity might be undetectable even in principle until the very last stage of the black hole evaporation process, when semiclassical methods no longer apply. Since assuming a generic final-state boundary condition is just a rather crude guess, finding such an excellent approximation to exact unitarity might be regarded as a success rather than a failure of the HM model.

4.4 Horizon-crossing agents

So far we have discussed only the effects of interactions between $\mathcal{H}_M$ and $\mathcal{H}_{\rm in}$. We should also worry about interactions between $\mathcal{H}_{\rm in}$ and $\mathcal{H}_{\rm out}$ as in Fig. 7b, which by degrading the maximal entangle-

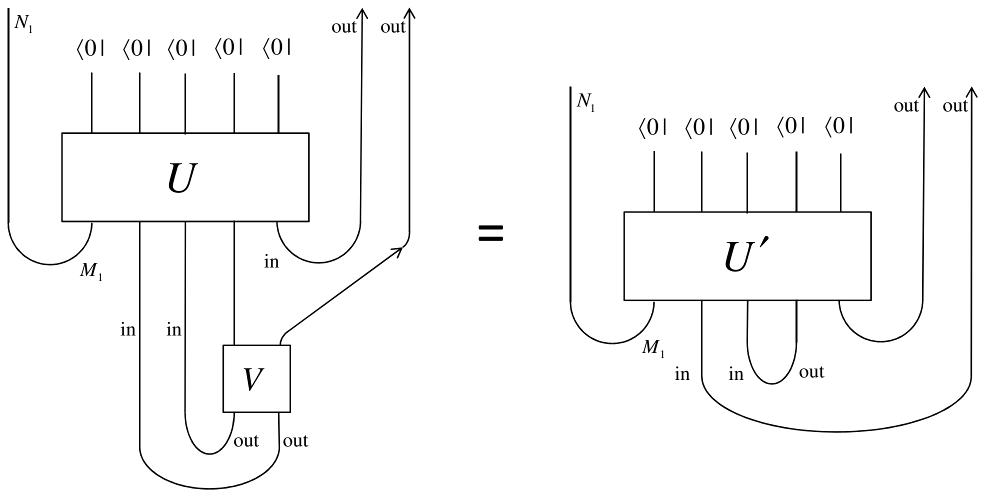

Figure 10: Some outgoing Hawking radiation is decoded and then redirected into the black hole. Here M_1 is a subsystem of the collapsing matter system, maximally entangled with a reference system N_1, and U is a unitary transformation specifying the final state at the singularity. The decoding map V can be replaced by its transpose acting on infalling radiation inside the black hole, and hence absorbed into U.

ment of the Unruh vacuum state might also compromise the fidelity of postselected teleportation. Since $\mathcal{H}_{\text{in}}$ is inside the black hole and $\mathcal{H}_{\text{out}}$ is outside, if we take the classical causal structure of spacetime seriously these interactions would have to be induced by an agent who first interacts with $\mathcal{H}_{\text{out}}$, then falls through the event horizon and interacts with $\mathcal{H}_{\text{in}}$, as shown in Fig. 9. Once inside the black hole, this falling agent would also be subject to the final-state boundary condition; it is really best to regard the agent as just another component of the infalling matter system $\mathcal{H}_M$. If the dynamics behind the horizon has been suitably tuned so that an arbitrary state of infalling matter becomes maximally entangled with the infalling radiation after the final-state projection, then an agent who interacts first with the outgoing radiation and then with the infalling radiation is merely a special case of this more general setting, and hence poses no additional threat to the unitarity of the black hole S-matrix.

Similarly, a malicious agent attempting to break unitarity by performing an entangled measurement straddling the horizon would be subject to the final-state condition and hence unable to carry out such mischief. The phenomenon captured in Fig. 3b, in which probabilities assigned to measurement outcomes depend on what measurements we will choose to performed in the future, can occur when the measurement outcome is stored in a memory which is not subject to the final-state boundary condition. The situation is different for a measuring agent who enters the black hole, as in that case the apparatus as well as (part of) the measured system is projected onto a particular final state. For the HM model with a unitarity-preserving boundary condition, the measurement as well as the agent herself are undone by the highly nonlocal scrambling transformation applied at the singularity.

4.5 Sending decoded radiation back into the black hole

There is another type of potential closed timelike curve we have not yet explicitly discussed, which has been considered previously in the context of the AMPS puzzle [8, 14]. Suppose that a subsystem

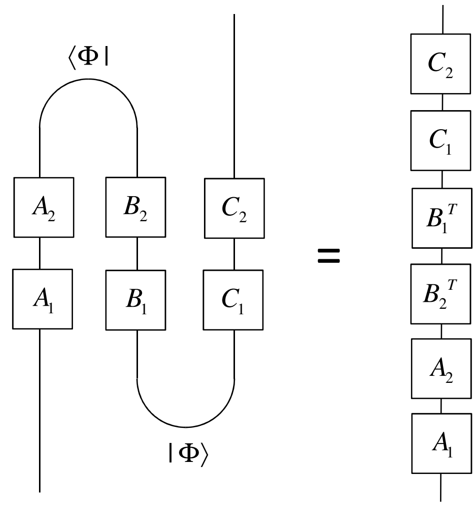

Figure 11: Correlation function in postselected teleportation, and the equivalent correlation function for the corresponding causally ordered process.

B of the outgoing Hawking radiation, recently emitted by an old black hole, is entangled with a subsystem A of the infalling radiation, and also with a subsystem R_B of the radiation emitted long ago. Disregarding the daunting computational complexity of the task [37], suppose that the R_B is then decoded into a compact quantum memory, and redirected back into the black hole, where it might reunite with its earlier self, encoded in A. What happens?

This situation is depicted in Fig. 10, where V is the unitary encoding map applied to the outgoing radiation, and U is the unitary transformation specifying the final state at the singularity. As in Fig. 8, we have introduced a reference system N_1, maximally entangled with a subsystem M_1 of the collapsing matter; in Fig. 10 we have suppressed the matter subsystem M_2, assumed to be initialized in a fixed state, which was shown in Fig. 8. The unitary V acting on $\mathcal{H}_{\rm out}$ is equivalent to its transpose V^T acting on $\mathcal{H}_{\rm in}$, which can therefore be absorbed into U as shown. With this revision, Fig. 10 is essentially the same as Fig. 8, except with M_2 replaced by an entangled state of outgoing and infalling radiation. By the same reasoning as previously, for a generic choice of U', the resulting map of M_1 to the outgoing radiation is very nearly unitary, and could become exactly unitary after a plausible dynamical adjustment.

5 Infalling observers

Up until now we have focused on the unitarity of the black hole S-matrix relating the asymptotic infalling matter and the asymptotic outgoing radiation. Even if this S-matrix is exactly unitary, though, infalling observers inside the black hole might still experience departures from conventional quantum theory in the HM model, arising from entangling interactions between $\mathcal{H}_M$ and $\mathcal{H}_{\rm in}$. What do infalling observers see?

5.1 Difficulty of detecting postselection when approaching the singularity

Postselected quantum mechanics provides no unambiguously defined state of a quantum system at intermediate times — evolving backward from the final-state boundary condition gives a different

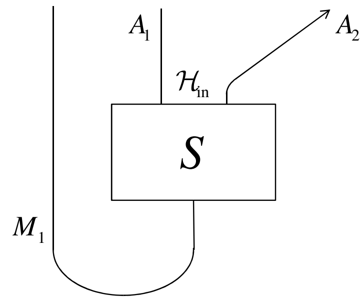

Figure 12: The collapsing matter subsystem M_1 is maximally entangled with a subsystem of the infalling radiation system $\mathcal{H}_{\rm in}$, specified by the scrambling unitary transformation S. If the subsystem A_2 of $\mathcal{H}_{\rm in}$ is discarded, then the complementary subsystem A_1 becomes nearly uncorrelated with M_1, if A_2 is larger than half the size of $\mathcal{H}_{M_1} \otimes \mathcal{H}_{\rm in}$.

answer in general than evolving forward from the initial-state boundary condition. There are unambiguous rules, however, for computing correlation functions of strings of operators inserted at various times [25]. For the HM model in the case where $\mathcal{H}_{\rm in}$ does not interact at all with $\mathcal{H}_{\rm M}$ or $\mathcal{H}_{\rm out}$, such correlation functions have a clear physical interpretation if each operator acts on only one of the three systems. As shown in Fig. 11 the correlation functions are exactly the same as for a system with a causally ordered information flow in which an infalling observer who reaches the singularity would subsequently experience (if still conscious after being scrambled at the singularity) the time-reversed evolution of the infalling Hawking radiation, followed by the forward time evolution of the outgoing Hawking radiation. That is what we really mean when we say the information flow can be "pulled tight." The physical interpretation of the correlators is less obvious for entangling operators that act on more than one of the three systems, just as one should expect for correlators in a causally ordered process in which a single operator acts at multiple times.

If $\mathcal{H}_M$ and $\mathcal{H}_{\rm in}$ interact yet the S-matrix is unitary, then entanglement between $\mathcal{H}_M$ and $\mathcal{H}_{\rm in}$ induced by the Hamiltonian dynamics between the horizon and singularity must in effect be undone at the singularity, then followed immediately by a projection onto a maximally entangled state of $\mathcal{H}_M \otimes \mathcal{H}_{\rm in}$. The resulting information flow behind the horizon does not have a well defined causal order, or in other words if we try to define a causal order we find that the quantum information encoded in the time-reversed infalling radiation could in principle interact with its earlier self encoded in the collapsing matter.

On the other hand, because the final-state projection throughly scrambles the quantum information encoded in the collapsing matter, the weird consequences of such closed timelike curves behind the horizon may be undetectable by observers with access to only a portion of the $\mathcal{H}_{\rm in}$ Hilbert space. To clarify this claim, consider Fig. 12, which depicts the time-reversed evolution from the singularity into the black hole interior. Here M_1 is a subsystem of the collapsing matter, which is maximally entangled with a subsystem, determined by the scrambling unitary S, of the infalling radiation system $\mathcal{H}_{\rm in}$. Suppose we discard the subsystem A_2 of $\mathcal{H}_{\rm in}$, presumed inaccessible to our infalling observer, and retain the complementary subsystem A_1, which the observer might be able to access. Averaging S over the normalized invariant Haar measure on the unitary group, and assuming that the overall state of $\mathcal{H}_{M_1} \otimes \mathcal{H}_{\rm in}$ is pure, we find that the density operator $\rho_{M_1 A_1}$

obeys the inequality [35]

$$\int dS \|\rho_{M_1A_1}(S) - \rho_{M_1}(S) \otimes \rho_{A_1}^{\max}\|_1 \leq \sqrt{\frac{|\mathcal{H}_{M_1}| \cdot |\mathcal{H}_{\rm in}|}{|A_2|^2}}, \tag{4}$$

where $\rho_{A_1}^{\max}$ denotes the maximally entangled state of A_1. The conclusion is that, for generic S, if the discarded system A_2 is larger than half the full system $\mathcal{H}_{M_1} \otimes \mathcal{H}_{\rm in}$, then M_1 is hardly entangled with A_1 at all; instead it is nearly maximally entangled with A_2. Specifically, if $\log_2 |M_1| = k$, $\log_2 |\mathcal{H}_{\rm in}| = n$, and $\log_2 |A_1| = \frac{1}{2}(n-k) - r$, we find that the state of M_1A_1 deviates in the L_1-norm from an uncorrelated product state by at most 2^{-r}. As in Sec. 4.3, we obtain the same result by averaging over a unitary 2-design rather than Haar measure.

Translated into the language of the HM model, this statement means that when quantum information encoded in a small subsystem of the collapsing matter Hilbert space is "reflected" at the singularity by a generic final-state boundary condition, the reflected information escapes the notice of an observer with access to much less than half of the infalling radiation. An infalling observer who crosses the event horizon of a black hole with mass m meets the singularity in proper time $O(m)$, and hence has very limited time to perform complex decoding operations on the infalling radiation. This observer may suffer horribly when subjected to the highly nonlocal scrambling transformation S at the singularity, but she might not have time to discern any other troubling violations of the rules of standard quantum mechanics.

5.2 Old black hole: entanglement verification and particle creation

Now, following AMPS [8, 14], we consider the case of a black hole H which is maximally entangled with an exterior system R, where R might for example be radiation previously emitted by a black hole older than the Page time [9]. For conceptual clarity and notational simplicity, suppose that H emits a single qubit B, which is required by unitarity to be maximally entangled with a qubit of R, as shown in Fig. 13a. This picture is vastly oversimplified, and in particular we are implicitly taking it for granted that the highly complex quantum decoding operation that distills from R a qubit entangled with B is possible in principle [37].

In the HM model, this transformation of HR entanglement into BR entanglement occurs as shown in Fig. 13b. Here AB denotes a pair of qubits in the maximally entangled Unruh vacuum state, whose state, after an appropriate basis choice, may be expressed as

$$|\phi^+\rangle_{AB} = \frac{1}{\sqrt{2}} \left(|00\rangle_{AB} + |11\rangle_{AB}\right). \tag{5}$$

H denotes a black hole qubit and R a radiation qubit, also assumed to be in the maximally entangled state $|\phi^+\rangle_{HR}$. We suppose for simplicity that the final-state boundary condition projects HA onto the maximally entangled state $|\phi^+\rangle_{HA}$ (corresponding to a trivial black hole S-matrix). This projection transforms the state $|\phi^+\rangle_{HR} \otimes |\phi^+\rangle_{AB}$ of the four qubits to $|\phi^+\rangle_{HA} \otimes |\phi^+\rangle_{BR}$, via the phenomenon known as *entanglement swapping* [38, 39]. In contrast to standard entanglement swapping, there is no need to convey the outcome of the entangled measurement of HA to the BR system to complete the protocol, because the HM boundary condition dictates that only one possible outcome can occur.

How might the entanglement structure indicated by Fig. 13b be verified? We would prefer not to consider entangling operations performed on AB, because such operations may alter the information flow, requiring us to think carefully about how the internal dynamics of the black hole must compensate to restore an information flow compatible with the unitarity of the black hole

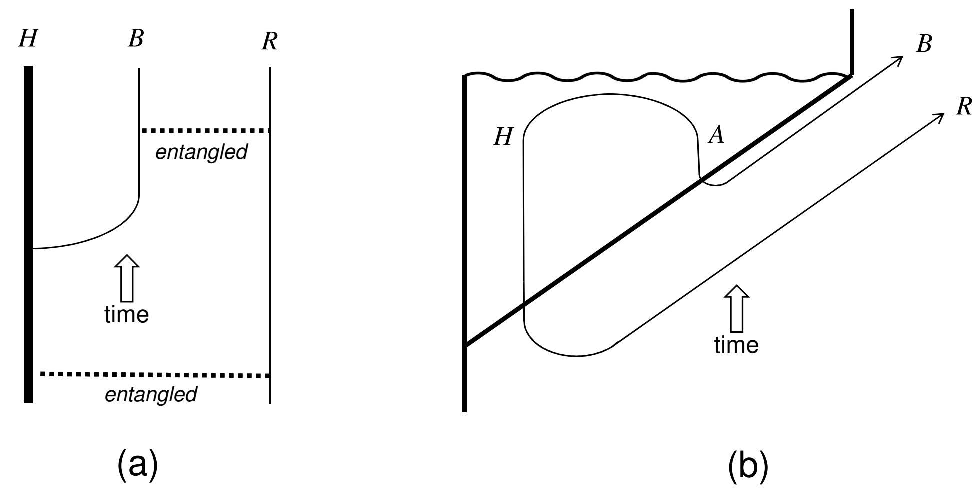

Figure 13: (a) Qubit H of an old black hole is maximally entangled with qubit R in the previously emitted Hawking radiation. When the black hole emits qubit B, the HR entanglement is transformed to BR entanglement. (b) Entanglement transfer in the HM model. AB is a maximally entangled qubit pair in the Unruh vacuum state. The final-state projection of HA onto a maximally entangled state creates maximal entanglement of BR via entanglement swapping.

S-matrix. It is much simpler, and sufficient, to think about product operators, as the maximally entangled state $|\phi^+\rangle$ can be completely characterized as the simultaneous eigenstate with eigenvalue 1 of the two commuting Pauli operators $X \otimes X$ and $Z \otimes Z$, where

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{6}$$

By checking both $X \otimes X = 1$ and $Z \otimes Z = 1$, we may verify that a two-qubit state is $|\phi^+\rangle$.

In Fig. 14, we consider the Pauli operator $\sigma \otimes \sigma$ applied to AB followed by $\tau \otimes \tau$ applied to BR, where $\sigma, \tau \in \{X, Z\}$. Since $\sigma^2 = \tau^2 = I$, it is clear, after "straightening" the information flow as shown, that these operators act trivially, as expected if both AB and BR are entangled pairs in the state $|\phi^+\rangle$, violating entanglement monogamy. But if we consider applying the Pauli operators in the opposite order, something strange happens if σ and τ anticommute, as shown in Fig. 15 — the product of $\sigma \otimes \sigma$ and $\tau \otimes \tau$ is -1 rather than 1. On the other hand, if we apply $X \otimes X$ or $Z \otimes Z$ to BR another time, we see that both still act trivially; hence the state of BR is still $|\phi^+\rangle$. Something seems to have altered the AB entanglement.

Such findings invite the following interpretation. (1) We note that an operator acting only on R has no effect on the verification of the AB entanglement. We conclude that a probe interacting with the early radiation R does not disturb the Unruh vacuum. (2) A Pauli operator acting only on B, because it anticommutes with $X \otimes X$ and/or $Z \otimes Z$, changes the entangled state of AB. Thus a probe acting on the recently emitted radiation can disturb the Unruh vacuum, creating a particle that can be seen by an infalling observable.

(3) A Pauli operator acting only on A, because it anticommutes with $X \otimes X$ and/or $Z \otimes Z$, changes the entangled state of BR. However, this change should not be interpreted as signaling from inside the black hole to outside. Rather, any operation acting on A that occurs behind the horizon is completely determined by the quantum state of infalling matter inside the black hole;

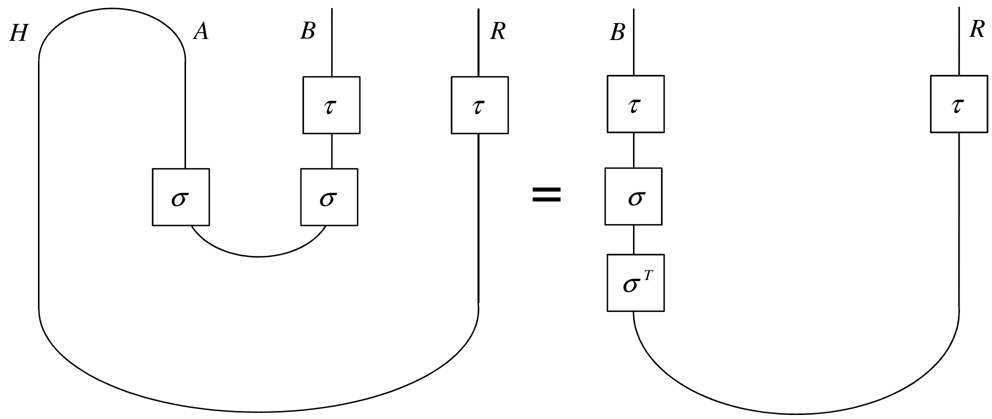

Figure 14: If we verify the AB entanglement first and the BR entanglement later, both verifications succeed.

hence the dependence of the outgoing radiation state on operations performed behind the horizon is just a special case of the black hole S-matrix relating the asymptotic incoming and outgoing states. For the black hole S-matrix to be causal as well as unitary we must not allow an action on A controlled by an infalling object to have an instantaneous effect on spacelike separated BR. In fact, we expect the emission in the Hawking radiation of information encoded in the infalling object to be delayed by at least the black hole's scrambling time [31, 32]. To enforce this constraint, the black hole's interior dynamics should be adjusted so that operations performed on A due to interactions with the infalling object are reversed before the the infalling radiation reaches the singularity, preventing any immediate response in BR.

Points (1), (2), and (3) are natural and as expected. But as noted above there is one surprise. (4) The action of $X\otimes X$ and $Z\otimes Z$ on BR is guaranteed to be trivial, consistent with our expectation that the state of BR is $|\phi^+\rangle$ due to entanglement swapping. But even though this entanglement verification on BR always succeeds, nevertheless it alters the entangled state of AB: *Successful verification of the BR entanglement excites the Unruh vacuum.*

This odd behavior clarifies how postselected teleportation resolves the AMPS puzzle. An agent who successfully verifies the BR entanglement before entering the black hole will fail if she attempts to verify the AB entanglement as well. This failure can be blamed on the agent's own activity prior to horizon crossing.

Were we to consider verifying the BR and AB entanglement by performing entangled measurements with outcomes $|\phi^+\rangle\langle\phi^+|$ and $I - |\phi^+\rangle\langle\phi^+|$, we would reach an even stranger conclusion as in Fig. 3 — that whether the BR entanglement verification fails or succeeds depends on whether or not we choose to perform the AB entanglement verification later on. We believe that this conclusion is incorrect because it relies on too naive a treatment of the entangled AB measurement which straddles the horizon. As we explained in Sec. 4.4, this entangled measurement must be performed by a horizon-crossing agent, and if the dynamics inside the horizon has been adjusted to ensure unitarity of the black hole S-matrix, no acausal effects outside the horizon should arise. Once we accept that unitarity ensures that the BR entanglement verification must succeed, our analysis here, which obviates the need for a delicate discussion of entangled horizon-straddling observables by focusing on product operators, supports the alternative conclusion that the BR entanglement verification interferes with the subsequent AB entanglement verification rather than vice versa.

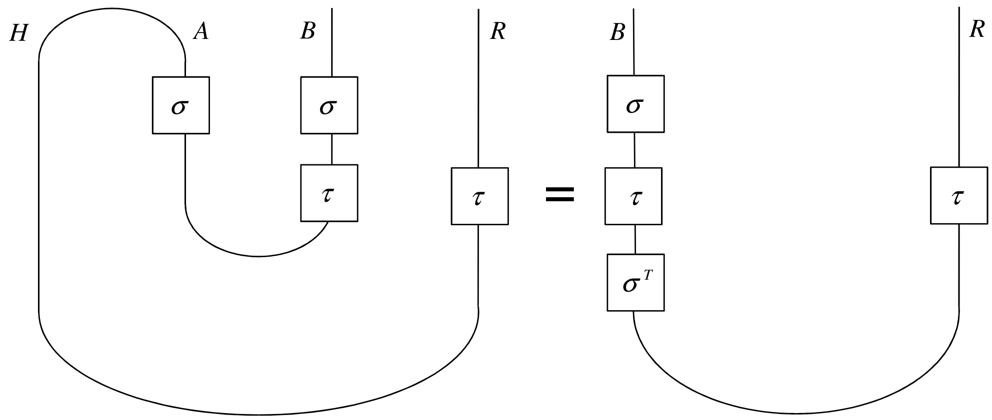

Figure 15: If we verify the BR entanglement first and the AB entanglement later, the first verification succeeds but the second one does not.

6 Discussion

The AMPS puzzle has deepened the mystery surrounding the fate of quantum information that falls into a black hole. AMPS investigated the compatibility of three reasonable assumptions: (1) unitarity of black hole evaporation, (2) smoothness of the black hole event horizon, and (3) validity of local effective field theory outside a black hole. They argued that these three assumptions are inconsistent, since together they imply that quantum correlations can be polygamous, contrary to standard quantum mechanics.

Our main point is that quantum correlations can be polygamous in the Horowitz-Maldacena final-state projection model, permitting these three assumptions to be reconciled. In the HM model, quantum information escapes from the black hole interior via postselected quantum teleportation, due to a boundary condition imposed at the spacelike singularity. Loosely speaking, quantum information flows forward in time from past infinity to the singularity, backward in time from the singularity to the horizon, then forward in time from the horizon to future infinity. If suitable dynamical constraints are satisfied, this flow of information is essentially equivalent to a manifestly unitary causally ordered flow moving only forward in time, at least for the purpose of describing the viewpoint of observers who stay outside the black hole. These constraints are nearly fulfilled by generic dynamical models, but as best we can tell they can be rigorously fulfilled only by fine tuning the model. On the other hand, since the HM model is formulated on a semiclassical spacetime background, achieving unitarity up to exponentially small corrections using a generic final-state boundary condition might be regarded as a success of the model.

In the HM model, observables inside the horizon fail to commute with observables outside the horizon acting on the same time slice, because in the corresponding causally ordered information flow, the outside observables act on the same system as the inside observables, but at a later "time." Other features of black hole complementarity are also realized; in particular, from the viewpoint of an observer who stays outside, the black hole behaves like a rapidly scrambling quantum system interacting with its surroundings. We see no reason why the physics outside the horizon could not be accurately captured by a dual boundary field theory as in AdS/CFT duality. The novel physics of the HM model occurs inside the black hole, particularly at the singularity; the model may provide helpful hints about how a dual description of the black hole interior should work, if such a description exists.

In postselected quantum mechanics, cloning of quantum states is possible, and because monogamy of quantum entanglement can be relaxed, we know no logically compelling argument for the existence of a firewall at the black hole horizon within the context of the HM model; conceivably, though, the horizon could nevertheless fail to be smooth for reasons other than those originally promulgated by AMPS. (See [34, 40, 41] for other arguments supporting the existence of firewalls.)

Like all other resolutions of the AMPS puzzle proposed so far, the HM model will need to be developed further before it can be conclusively assessed. In particular, we should strive to expunge the dynamical fine tuning the model seems to require, or to explain persuasively why the fine tuning is somehow natural.

Even if the HM model turns out to be wrong in detail, we believe that the picture of information flow in black hole spacetimes provided by the model is interesting and valuable. This picture reminds us that the global physics of the black hole interior could be subtle, and in particular that fundamental properties of standard quantum mechanics such as the no-cloning principle and monogamy of entanglement might be relaxed in a complete theory of quantum gravity. And if nature really indulges in postselection at future spacelike singularities, we may anticipate deep consequences in quantum cosmology as well as black hole physics.

Acknowledgments

We gratefully acknowledge very valuable discussions with Alexei Kitaev. SL thanks Max Tegmark for helpful discussions, and JP thanks Raphael Bousso, Juan Maldacena, and Douglas Stanford for inspiring discussions and correspondence regarding final-state projection models. JP appreciates many helpful interactions with other participants at the August 2013 KITP workshop "Black Holes: Complementarity, Fuzz, or Fire," and we also benefited from comments on the manuscript from Don Marolf and Douglas Stanford. The research of SL was supported in part by DARPA, by the ARO under a MURI program, and by Jeffrey Epstein. The research of JP was supported in part by NSF grant PHY-0803371 and DOE grant DE-FG03-92-ER40701. The Institute for Quantum Information and Matter (IQIM) is an NSF Physics Frontiers Center with support from the Gordon and Betty Moore Foundation.

References

[1] S. W. Hawking, Black hole explosions, Nature 248, 30 (1974).

[2] S. W. Hawking, Breakdown of predictability in gravitational collapse, Phys. Rev. D 14, 2460 (1976).

[3] L. Susskind, L. Thorlacius, and J. Uglum, The stretched horizon and black hole complementarity, Phys. Rev. 48, 3743 (1993), arXiv:hep-th/9306069.

[4] L. Susskind and L. Thorlacius, Gedanken experiments involving black holes, Phys. Rev. D 49, 966-974 (1994), arXiv:hep-th/9308100.

[5] G. 't Hooft, Dimensional reduction in quantum gravity, arXiv:gr-qc/9310026 (1993).

[6] L. Susskind, The world as a hologram, J. Math. Phys. 36, 6377-6396 (1995), arXiv:hep-th/9409089.

[7] J. Maldacena, The large-N limit of superconformal field theories and supergravity, Adv. Theor. Math. Phys. 2, 231 (1998), arXiv:hep-th/9711200.

[8] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, Black holes: complementarity or firewalls?, JHEP 1302, 062 (2013), arXiv:1207.3123.

[9] D. N Page, Average entropy of a subsystem, Phys. Rev. Lett. 71, 1291 (1993), arXiv:gr-qc/9305007; D. N. Page, Black hole information, arXiv:hep-th/9305040 (1993).

[10] B. M. Terhal, Is entanglement monogamous?, arXiv:quant-ph/0307120 (2003).

[11] M. Koashi and A. Winter, Monogamy of quantum entanglement and other correlations, Phys. Rev. A 69, 022309 (2004), arXiv:quant-ph/0310037.

[12] S. D. Mathur, The information paradox: A pedagogical introduction, Class. Quant. Grav. 26, 224001 (2009), arXiv:0909.1038.

[13] S. L. Braunstein, Black hole entropy as entropy of entanglement, or it's curtains for the equivalence principle, arXiv:0907.1190v1; S. L. Braunstein, S. Pirandola and K. Zyczkowski, Better Late than Never: Information Retrieval from Black Holes, Phys, Rev. Lett. 110, 101301 (2013).

[14] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, An apologia for firewalls, arXiv:1304.6483 (2013).

[15] S. B. Giddings, Nonviolent nonlocality, arXiv:1211.7070 (2012).

[16] K. Papadodimas and S. Raju, An infalling observer in AdS/CFT, arXiv:1211.6767 (2012).

[17] E. Verlinde and H. Verlinde, Black hole entanglement and quantum error correction, arXiv:1211.6913 (2012).

[18] J. Maldacena and L. Susskind, Cool horizons for entangled black holes, arXiv:1306.0533.

[19] G. T. Horowitz and J. Maldacena, The black hole final state, JHEP 0402:008 (2004), arXiv:hep-th/0310281 (2003).

[20] W. K. Wootters and W. H. Zurek, A single quantum cannot be cloned, Nature 299, 802-803 (1982).

[21] D. Dieks, Communication by EPR devices, Physics Letters A 92, 271-272 (1982).

[22] J. B. Hartle and S. W. Hawking, Wave function of the universe, Phys. Rev. D 28, 29602975 (1983).

[23] D. Gottesman and J. Preskill, Comment on "The black hole final state," JHEP 0403:026 (2004), arXiv:hep-th/0311269.

[24] S. Lloyd, Almost certain escape from black holes in final state projection models, Phys. Rev. Lett. 96, 061302 (2006), arXiv:quant-ph/0406205.

[25] Y. Aharonov, D. Z. Albert, and L. Vaidman, How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100, Phys. Rev. Lett. 60, 1351 (1988).

[26] S. Lloyd, L. Maccone, R. Garcia-Patron, V. Giovannetti, Y. Shikano, S. Pirandola, L. A. Rozema, A. Darabi, Y. Soudagar, L. K. Shalm, and A.M. Steinberg, Closed timelike curves via postselection: theory and experimental test of consistency, Phys. Rev. Lett. 106, 040403 (2011), arXiv:1005.2219.

[27] S. Lloyd, L. Maccone, R. Garcia-Patron, V. Giovannetti, and Y. Shikano, Quantum mechanics of time travel through post-selected telportation, Phys. Rev. D 84, 025007 (2011), arXiv:1007.2615.

[28] R. Bousso, Complementarity is not enough, Phys. Rev. D 87, 124023 (2013), arXiv:1207.5192.

[29] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels, Phys. Rev. Lett. 70, 18951899 (1993).

[30] E. Poisson and W. Israel, Inner structure of black holes, Phys. Rev. D41, 1796-1809 (1990).

[31] Y. Sekino and L. Susskind, Fast scramblers, JHEP0810:065 (2008), arXiv:0808.2096.

[32] P. Hayden and J. Preskill, Black holes as mirrors: quantum information in random subsystems, JHEP 0709:120 (2007), arXiv:0708.4025.

[33] S. Aaronson, "Quantum computing, postselection, and probabilistic polynomial-time," Proc. Roy. Soc. A 461, 3473 (2005), arXiv:quant-ph/0412187.

[34] D. Marolf and J. Polchinski, Gauge/gravity duality and the black hole interior, arXiv:1307.4706 (2013).

[35] P. Hayden, M. Horodecki, A. Winter, and J. Yard, A decoupling approach to the quantum capacity, Open Syst. Inf. Dyn. 15, 7 (2008), arXiv:quant-ph/0702005.

[36] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs: constructions and applications, Phys. Rev. A 80, 012304 (2009), arXiv:quant-ph/0606161.

[37] D. Harlow and P. Hayden, Quantum computation vs. firewalls, arXiv:1301.4504 (2013).

[38] M. Zukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, "Event-ready-detectors" Bell experiment via entanglement swapping, Phys. Rev. Lett. 71, 42874290 (1993).

[39] S. Bose, V. Vedral, and P. L. Knight, Multiparticle generalization of entanglement swapping, Phys. Rev. A 57, 822829 (1998), arXiv:quant-ph/9708004.

[40] R. Bousso, Firewalls from double purity, arXiv:1308.2665 (2013).

[41] R. Bousso, Frozen vacuum, arXiv:1308.3697 (2013).