The ghost in the radiation: Robust encodings of the black hole interior

Isaac H. Kim[1], Eugene Tang[2], John Preskill[2]

[1] *The University of Sydney, Camperdown, NSW 2006, Australia*

[2] *California Institute of Technology, Pasadena, CA 91125, USA*

February 6, 2021

In 1975, Hawking showed that black holes emit thermal radiation [1], thereby uncovering a deep tension between quantum mechanics and general relativity. Hawking's calculation suggests that black holes eventually evaporate away completely, leaving behind — independently of their initial conditions — perfectly thermal radiation. On the other hand, if black hole time evolution is to obey the usual unitary dynamics of quantum mechanics, pure initial states must be taken to pure final states. This is in apparent contradiction with Hawking's calculation. This problem, which has since been known as the *black hole information paradox*, has remained as one of the most perplexing paradoxes in physics.

More recently, the black hole information paradox was sharpened by using the monogamy of the entanglement [2]. This paradox, which is often referred to as the *firewall paradox*, led to various proposals to construct the interior of the black hole. One of the leading proposals involves the idea that the interior of the black hole is somehow encoded within the early radiation [3, 4]. While this proposal resolves the original firewall paradox, it is not without its own issues. Since the black hole interior is embedded within the early radiation, an exterior observer may ostensibly perturb the black hole interior by acting on the exterior radiation. If this were true, then it would appear that the conventional notions of spacetime locality and causality appear to break down [5].

In this work, we formalize the embedding of the interior into the exterior radiation as a quantum error-correcting code. In particular, we propose a complexity-theoretic mechanism by which locality and causality can remain intact even in the presence of such non-local encodings. Concretely, we consider a black hole that begins as a pure matter state $|\phi_{\text{matter}}\rangle$ which undergoes subsequent gravitational collapse to become a black hole state $|\Psi_{EBH}\rangle$ at time t. We model the gravitation collapse as unitary time evolution induced by some black hole unitary U_{bh}. The labels E,B, and H in the black hole state refer to the early radiation (E), an outgoing Hawking mode (B), and the remaining black hole (H), respectively; see Figure 1.

We assume that $|H| < |E|$ so that we work explicitly with late time black holes for which the firewall paradox applies in full force.[1] We also assume that the outgoing Hawking mode B consists of a single qubit, i.e., $|B| = 1$. This is simply for ease of exposition, our results do not change significantly for any B of constant size. To model the interactions of an exterior observer with the early Hawking radiation, we introduce an ancillary subsystem O. The subsystem O plays the role of a physical observer which is able to interact with the early radiation through some unitary process $U_{\mathcal{E}}$. We will assume that $|O| \ll |H|$, so that the observer is physically small compared to the remaining black hole.

Hawking's calculation famously shows the thermality of Hawking radiation. From a modern information theoretic point of view, this is analogous to saying that the radiation state appears maximally mixed to external observers. We also understand however, that such a statement must be made with additional qualifications; the purity of a quantum state may appear very different to observers with access to different computational resources. Our work recasts the thermality of Hawking radiation as a more precise statement about the computational pseudorandomness [6] of the radiation state.

Definition 1. *Let $|\Psi\rangle_{EBH}$ be the state of the black hole and the radiation. Let $\sigma_{EB} = I_{EB}/d_{EB}$ be the maximally mixed state of EB, and let $\rho_{EB} = \mathrm{Tr}_H(|\Psi\rangle\langle\Psi|)$. We say that the state $|\Psi\rangle_{EBH}$ is pseudorandom on the radiation EB, if there exists some $\alpha > 0$ such that*

$$|\Pr(\mathcal{M}(\rho_{EB}) = 1) - \Pr(\mathcal{M}(\sigma_{EB}) = 1)| \leq 2^{-\alpha|H|}, \tag{1}$$

[1] We denote the size, i.e., the number of qubits, of a system E by $|E|$. Thus E has dimension $2^{|E|}$.

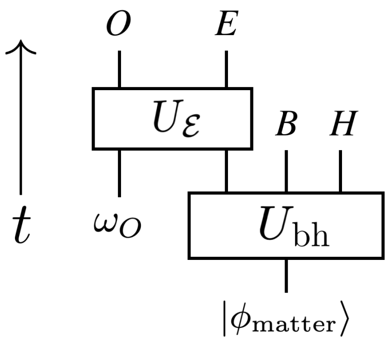

Figure 1: An external observer O interacts with the early radiation E. The joint time evolution of the black hole and the observer is assumed to be a unitary process.

for any two-outcome measurement $\mathcal{M}$ with quantum complexity polynomial in $|H|$, the size of the remaining black hole.

Our starting point and fundamental working assumption is that the radiation state associated with a late time black hole is computationally pseudorandom in the sense of definition 1. This pseudorandomness hypothesis has physical consequences for a realistic, i.e., computationally limited, observer O who is only able to interact with the early radiation through a low-complexity unitary $U_{\mathcal{E}}$. For such observers, the pseudorandomness hypothesis translates into a fundamental decoupling bound:

Lemma 2 (Decoupling Bound)**.** *Let ρ_{OEBH} denote any state obtained from $\omega_O \otimes |\Psi\rangle_{EBH}$ by acting with some polynomial-size circuit $U_{\mathcal{E}}$. See Figure 1. Then*

$$\|\rho_{OB} - \rho_O \otimes \rho_B\|_1 \leq 6 \cdot 2^{-(\alpha|H|-|O|)}. \tag{2}$$

The decoupling bound allows us to define a natural code subspace associated with the black hole. Letting $|\Psi\rangle_{EBH}$ denote the state of the black hole and radiation, we define an operator $V_\Psi : \mathcal{H}_{\tilde{B}} \to \mathcal{H}_{EH}$ by

$$V_\Psi|i\rangle_{\tilde{B}} = 2\left(I_{EH} \otimes \langle\omega|_{B\tilde{B}}\right)\left(|\Psi\rangle_{EHB} \otimes |i\rangle_{\tilde{B}}\right), \tag{3}$$

where $|\omega\rangle = 2^{-1/2}(|00\rangle + |11\rangle)$ denotes the maximally entangled state. We can use the decoupling bound to show that V_Ψ is ϵ-close to an isometric embedding, where $\epsilon = 2 \cdot 2^{-\alpha|H|}$. It follows that V_Ψ defines an approximate quantum error-correcting code. The code subspace defined by V_Ψ will be called the *black hole encoding*, which defines the desired embedding of the partnered interior modes $\tilde{B}$ into the exterior radiation (more precisely, into EH).

Having identified a suitable error-correcting code, we proceed to characterize its properties. Through the information-disturbance trade-off for approximate quantum error-correction [7, 8], the decoupling bound can be shown to be equivalent the correctability of error channels $\mathcal{E} : S(E) \to S(E)$ with low Kraus rank and low complexity. The conditions of the error channel having low rank and low complexity corresponds precisely to the physical assumption of the observer O (which serves as a purifying space of the channel) having small size (relative to H) and bounded computational power (so that $U_{\mathcal{E}}$ is polynomial complexity in $|H|$).

Theorem 3. *Let $\mathcal{E}$ be an error channel on E with purification $U_{\mathcal{E}}$. Suppose that the decoupling bound holds. Then $\mathcal{E}$ is ϵ-correctable for V_Ψ, where*

$$\epsilon = \sqrt{\frac{3}{2}} \cdot 2^{-(\alpha|H|-|O|)/2}. \tag{4}$$

The black hole code makes precise what it means to embed a part of the black hole interior into the early Hawking radiation. However, such an embedding is manifestly non-local, and so we must ensure that the embedding is consistent with the locality and causality of the ambient spacetime. More precisely, we must address the following problem: Since the interior of the black hole is encoded within the early radiation E, there should exist operators acting exclusively on E which can perturb the interior. By applying such an operator, an observer outside the black hole can ostensibly signal into the black hole interior from far away, in direct violation of Einstein causality.

The resolution of the locality problem above involves casting the error-correcting conditions of Theorem 3 into algebraic form. In the process, we establish a novel characterization for the correctability of a code which might find applications to more more general settings.

Definition 4. *Let $V : \tilde{\mathcal{H}} \to \mathcal{C} \subseteq \mathcal{H}$ be a (approximate) quantum error-correcting code. Let $\mathcal{E} : S(\mathcal{H}) \to S(\mathcal{H})$ be an error channel with Kraus operators $K = \{E_a\}$. Given an operator $\tilde{T} : \tilde{\mathcal{H}} \to \tilde{\mathcal{H}}$, we say that $T : \mathcal{H} \to \mathcal{H}$ is a δ-approximate* ghost logical operator *for $\tilde{T}$ (with respect to $\mathcal{E}$) if*

$$\|TE_aV - E_aV\tilde{T}\| \leq \|\tilde{T}\|\delta \tag{5}$$

for all $E_a \in K \cup \{I\}$, so that E_a above is either a Kraus operator or the identity. We say that a code admits a complete *set of δ-approximate ghost logical operators if there exists a δ-approximate ghost logical operator for every operator $\tilde{T} : \tilde{\mathcal{H}} \to \tilde{\mathcal{H}}$.*

Definition 4 formalizes the concept of logical operators which commutes with a set of error operators. It turns out that the existence of a complete set of ghost logical operators for an error channel $\mathcal{E}$ is essentially equivalent to the correctability of the channel.

Theorem 5. *Let $V : \tilde{\mathcal{H}} \to \mathcal{C} \subseteq \mathcal{H}$ be an approximate quantum error-correcting code. Given an error channel $\mathcal{E} : S(\mathcal{H}) \to S(\mathcal{H})$, define the associated channel $\mathcal{E}_{\mathcal{I}} = (\mathcal{E} + \mathcal{I})/2$, where $\mathcal{I}$ is the identity channel. If $\mathcal{E}_{\mathcal{I}}$ is ϵ-correctable for $\mathcal{C}$, then there exists a complete set of δ-approximate ghost logical operators, where*

$$\delta = 2^{5/4}(\dim \mathcal{C})^2\sqrt{\epsilon}. \tag{6}$$

Conversely, suppose that there exists a complete set of δ-approximate ghost logical operators with respect to $\mathcal{E}_{\mathcal{I}}$. Then $\mathcal{E}_{\mathcal{I}}$ is an ϵ-correctable channel for $\mathcal{C}$, where

$$\epsilon = r\sqrt{2(\dim \mathcal{C})\delta}, \tag{7}$$

where r is the Kraus rank of $\mathcal{E}_{\mathcal{I}}$.

The existence of ghost logical operators presents a resolution to the non-locality of the black hole code embedding. Together, Theorems 3 and 5 ensures that we can always find a complete set of ghost logical operators which commutes with all operators applicable by a computationally bounded observer. More precisely, let T_B be any operator acting on the outgoing Hawking mode B. Then there exists a corresponding ghost logical operator T_{EH} acting on EH such that

$$T_{EH}|\Psi\rangle_{EBH} \approx T_B|\Psi\rangle_{EBH}, \quad \text{and} \quad [T_{EH}, E_a]|\Psi\rangle_{EBH} \approx 0, \tag{8}$$

where $\{E_a\}$ is any set of operations that the observer can apply onto the radiation E, subject to their computational constraints. All of the approximate equalities above hold with exponentially small error in $|H|$. Roughly speaking, the first equality above certifies the fact that B and its interior mode encoded in EH remain "maximally entangled". The second equality above says that any physically reasonable operation performed by a computationally limited observer necessarily commutes with a complete set of logical operators acting on the encoded mode. This ensures that such an observer cannot signal into the black hole interior by acting on the exterior radiation, showing that our black hole code embedding retains the usual notions of locality and causality.

Our paper establishes that, albeit in our simplified model, one can find a definition of the interior mode that is consistent with both Hawking's calculation and spacetime locality. The crucial ideas that led us to this conclusion were (i) the construction of quantum error-correcting codes can correct *low-complexity errors*, and (ii) the existence of ghost logical operator, which are both novel concepts to the best of our knowledge. These concepts illuminate aspects of quantum error-correction that does not seem to have any classical counterpart. Perhaps these ideas can be used to protect quantum information in other setups as well.

References

[1] S. W. Hawking, "Particle creation by black holes," *Comm. Math. Phys.*, vol. 43, no. 3, pp. 199–220, 1975.

[2] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, "Black holes: complementarity or firewalls?," *JHEP*, vol. 2013, p. 62, Feb 2013.

[3] J. Maldacena and L. Susskind, "Cool horizons for entangled black holes," *Fortschritte der Physik*, vol. 61, no. 9, pp. 781–811, 2013.

[4] K. Papadodimas and S. Raju, "An infalling observer in ads/cft," *JHEP*, vol. 2013, p. 212, Oct 2013.

[5] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, "An apologia for firewalls," *JHEP*, vol. 2013, p. 18, Sep 2013.

[6] Z. Ji, Y.-K. Liu, and F. Song, "Pseudorandom quantum states," in *Advances in Cryptology – CRYPTO 2018*, (Cham), pp. 126–152, Springer International Publishing, 2018.

[7] C. Bény and O. Oreshkov, "General conditions for approximate quantum error correction and near-optimal recovery channels," *Phys. Rev. Lett.*, vol. 104, p. 120501, Mar 2010.

[8] P. Hayden and A. Winter, "Weak decoupling duality and quantum identification," *IEEE Transactions on Information Theory*, vol. 58, p. 4914–4929, Jul 2012.