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

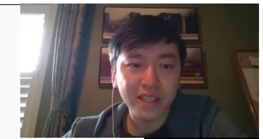
Speakers: Eugene Tang

Series: Perimeter Institute Quantum Discussions

Date: September 16, 2020 - 4:00 PM

URL: http://pirsa.org/20090016

Abstract: We reconsider the black hole firewall puzzle, emphasizing that quantum error-correction, computational complexity, and pseudorandomness are crucial concepts for understanding the black hole interior. We assume that the Hawking radiation emitted by an old black hole is pseudorandom, meaning that it cannot be distinguished from a perfectly thermal state by any efficient quantum computation acting on the radiation alone. We then infer the existence of a subspace of the radiation system which we interpret as an encoding of the black hole interior. This encoded interior is entangled with the late outgoing Hawking quanta emitted by the old black hole, and is inaccessible to computationally bounded observers who are outside the black hole. Specifically, efficient operations acting on the radiation, those with quantum computational complexity polynomial in the entropy of the remaining black hole, commute with a complete set of logical operators acting on the encoded interior, up to corrections which are exponentially small in the entropy. Thus, under our pseudorandomness assumption, the black hole interior is well protected from exterior observers as long as the remaining black hole is macroscopic. On the other hand, if the radiation is not pseudorandom, an exterior observer may be able to create a firewall by applying a polynomial-time quantum computation to the radiation.



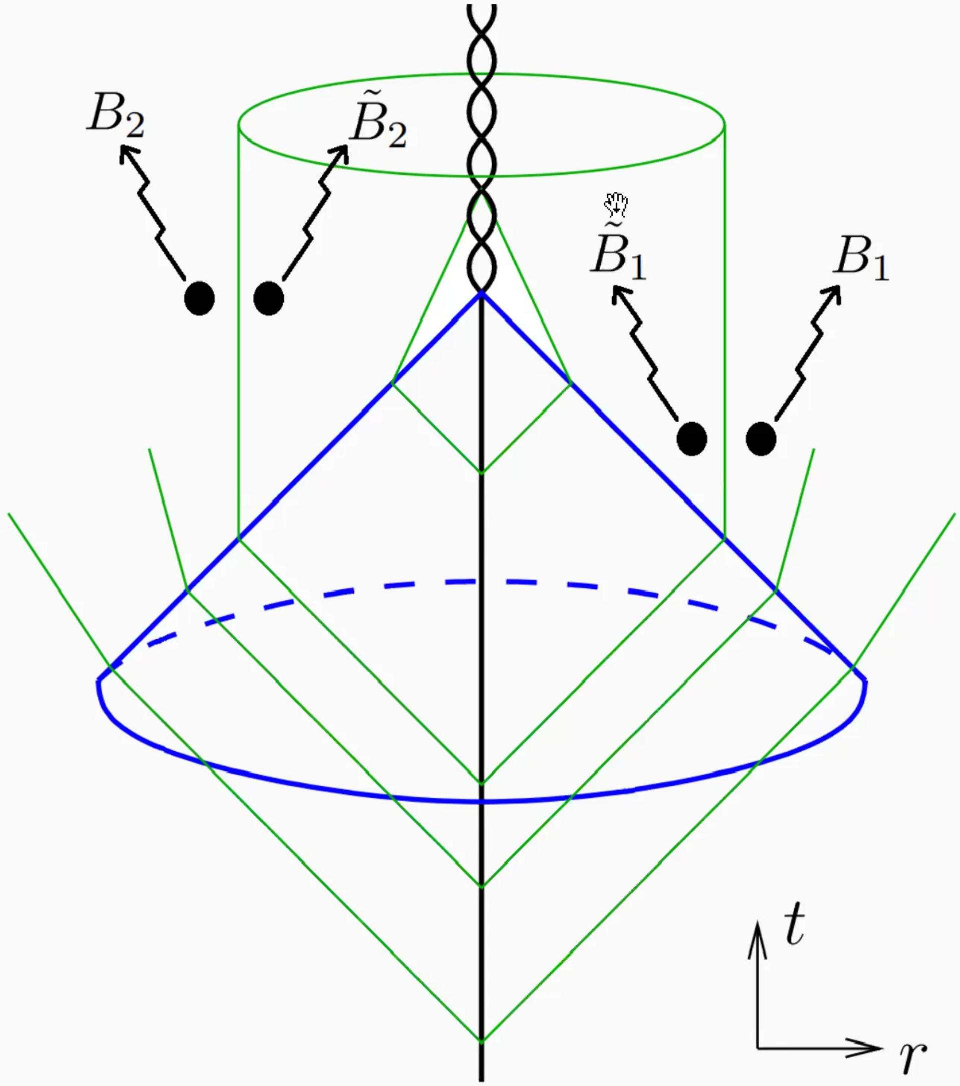
The ghost in the radiation: Robust encodings of the black hole interior

Isaac Kim, Eugene Tang, John Preskill

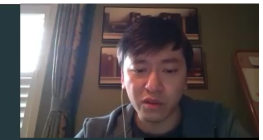
September 16, 2020 - Perimeter Quantum Information Seminar

arXiv: 2003.05451

The life of a black hole



The Entanglement Structure of Hawking Radiation



The pair creation picture suggests a simple toy-model for Hawking radiation:

1. We begin with a pure state $|\psi\rangle_M$ representing the state of the black hole on an initial time slice.

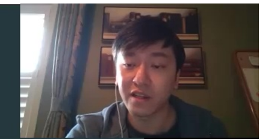
2. Hawking radiation is then induced from pair creation.

   The schematic evolution of our state is given by

$$|\psi\rangle_M \mapsto |\psi\rangle_M \otimes \frac{1}{\sqrt{2}}(|00\rangle_{\tilde{B}_1 B_1} + |11\rangle_{\tilde{B}_1 B_1}). \tag{1}$$

3. Continual time evolution repeatedly produce new correlated pairs, with the state at step N being

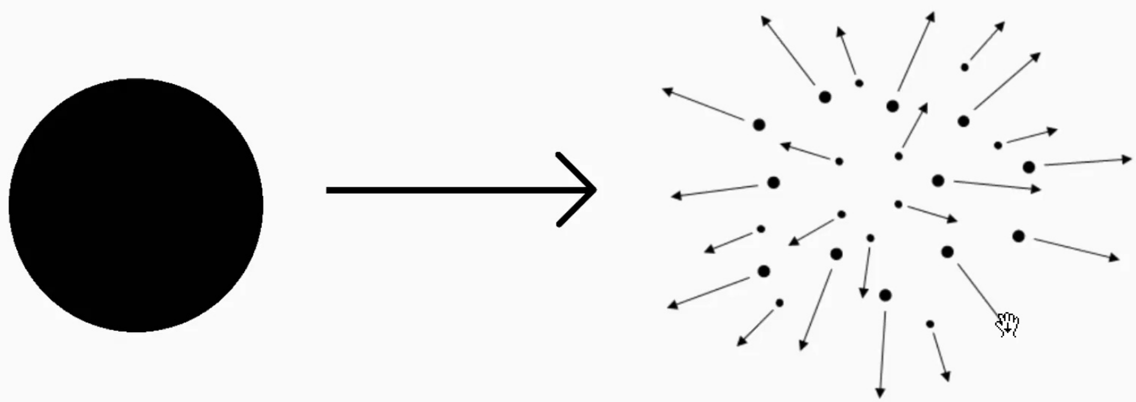
$$|\Psi_N\rangle = |\psi\rangle_M \otimes \frac{1}{\sqrt{2^N}} \bigotimes_{k=1}^{N} \left(|00\rangle_{\tilde{B}_k B_k} + |11\rangle_{\tilde{B}_k B_k}\right). \tag{2}$$



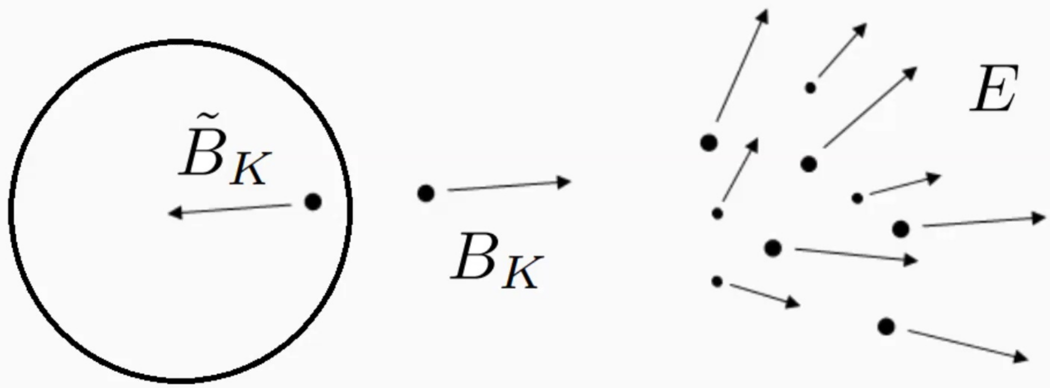
The Black Hole Firewall Problem

If we trace away the interior modes $\tilde{B}_k$, then the exterior modes B_k are left in a maximally mixed state: the Hawking radiation is *thermal*.

If evaporation continues until the black hole has completely radiated away, then we are left with a mixed thermal state, violating the unitarity of quantum mechanics.



The Black Hole Firewall Problem

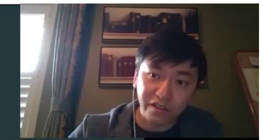


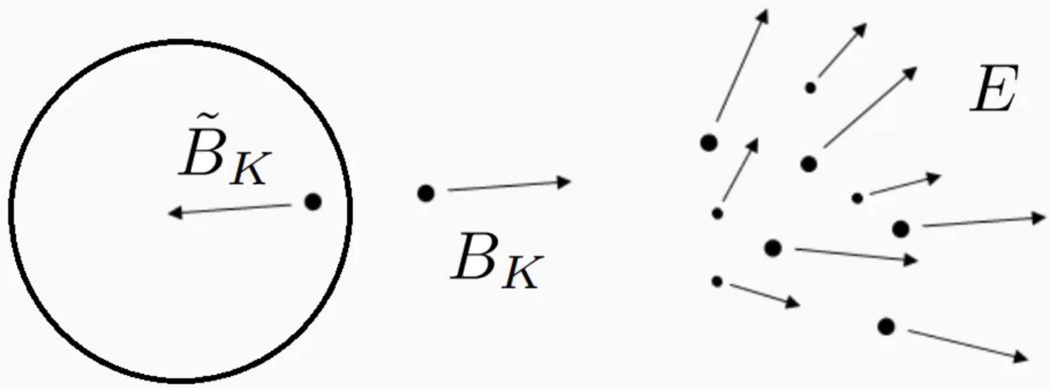
Consider some late outgoing mode B_K, and denote the collection of early modes $\{B_k \mid k < K\}$ as E.

From the toy-model above, B_K must be (maximally) entangled with its partner mode $\tilde{B}_K$ in the interior:

$$S(B_K \tilde{B}_K) = 0. \tag{3}$$

The Black Hole Firewall Problem





Consider some late outgoing mode B_K, and denote the collection of early modes $\{B_k \mid k < K\}$ as E.

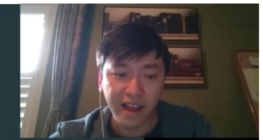
From the toy-model above, B_K must be (maximally) entangled with its partner mode $\tilde{B}_K$ in the interior:

$$S(B_K \tilde{B}_K) = 0. \tag{3}$$

If the final state of evaporation is to remain unitary, then B_K must purify the early radiation:

$$S(B_K E) < S(E). \tag{4}$$

The Black Hole Firewall Problem



We have the conditions:

$$S(ABC) + S(B) \leq S(AB) + S(BC), \tag{5}$$

$$S(B_K \tilde{B}_K) = 0, \tag{6}$$

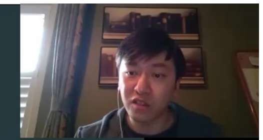
$$S(B_K E) < S(E). \tag{7}$$

Collectively, these conditions are contradictory:

$$\begin{aligned} S(B_K) + S(E) &= S(B_K) + S(E B_K \tilde{B}_K) && (9) \\ &\leq S(B_K \tilde{B}_K) + S(B_K E) && (10) \\ &= 0 + S(B_K E) && (11) \\ &< S(E) && (12) \end{aligned}$$

This formulation of the black hole information problem is called the *firewall paradox*.

The Black Hole Firewall Problem



There are several proposed solutions of the firewall paradox, each with its own merits and flaws. We will focus on the $\tilde{B} \subseteq E$ proposal.

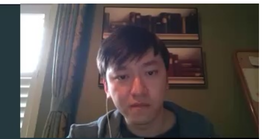
Problem: An outgoing mode B must be entangled with both an interior mode $\tilde{B}$ *and* the early radiation E. Violation of monogamy.

Solution: Identify the two problematic subsystems. Embed the interior partner mode $\tilde{B}$ *within* the exterior radiation E.

Appears to cause just as many problems as it solves:

- If the interior is actually embedded within the exterior, how can such an embedding respect the causal structure of the black hole?
- How do we protect the interior from outside observers?
- How can such an embedding be realized in practice?

Black Holes as Quantum Error-Correcting Codes



1. We postulate that black holes are efficient "scramblers" in a very precise way, namely that the exterior Hawking radiation emitted by a black hole is a *computationally pseudorandom* state.

2. Through the pseudorandomness hypothesis, we show that black holes define a natural encoding of each interior mode into the exterior Hawking radiation.

3. Such an encoding forms an error-correcting code which protects against all operations with sufficiently small complexity.

A sufficiently powerful observer can detect violations of causality and locality, but only provided that they are able to perform operations which are of exponential complexity in the entropy of the remaining black hole.

Pseudorandomness

A quantum state ρ is said to be *computationally pseudorandom* if for all polynomial time algorithms $\mathcal{A}$, we have

$$\left| \Pr[\mathcal{A}(\rho) = 1] - \Pr[\mathcal{A}(\sigma) = 1] \right| = \text{error}. \tag{13}$$

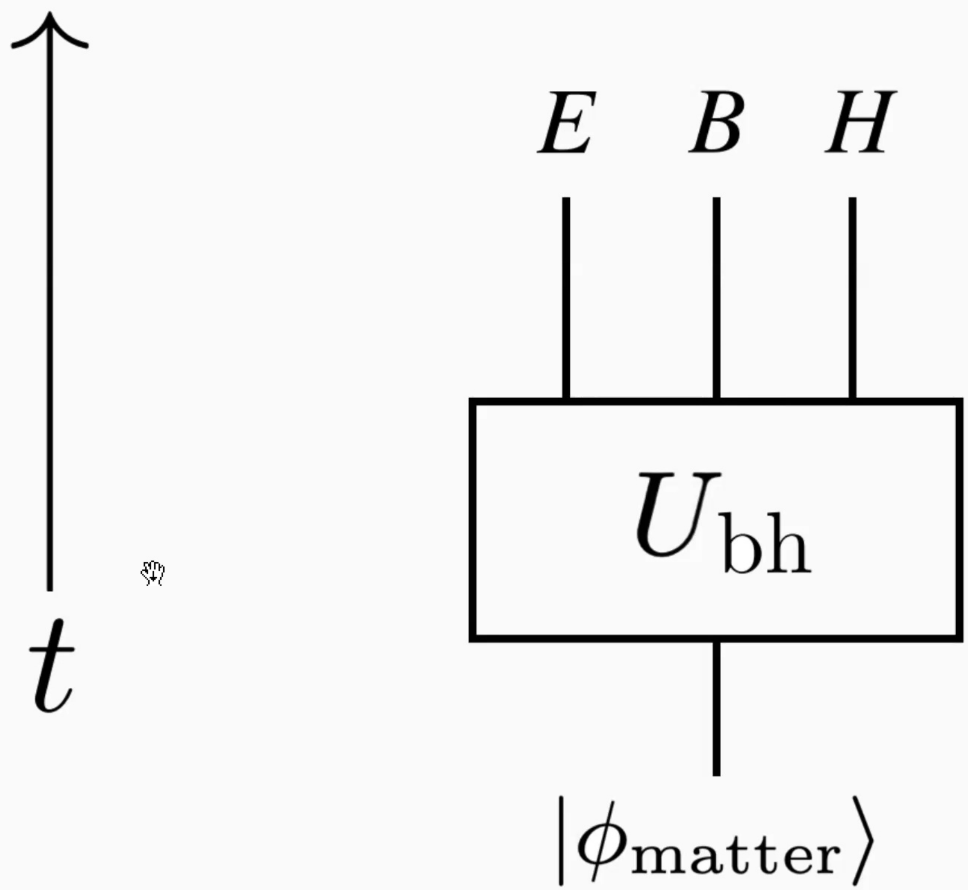
Not clear if pseudorandom states even exists. Simple counting argument: Too many states, not enough circuits.

Theorem [1208.0692]: Let $\mathcal{M} = \{I - M, M\}$ be a POVM of complexity less than n^k. Then almost all n-qubit states of circuit complexity $O\left(n^{11k+9}\right)$ will be indistinguishable from the maximally mixed state with respect to $\mathcal{M}$, with error $2^{-n/4}$.

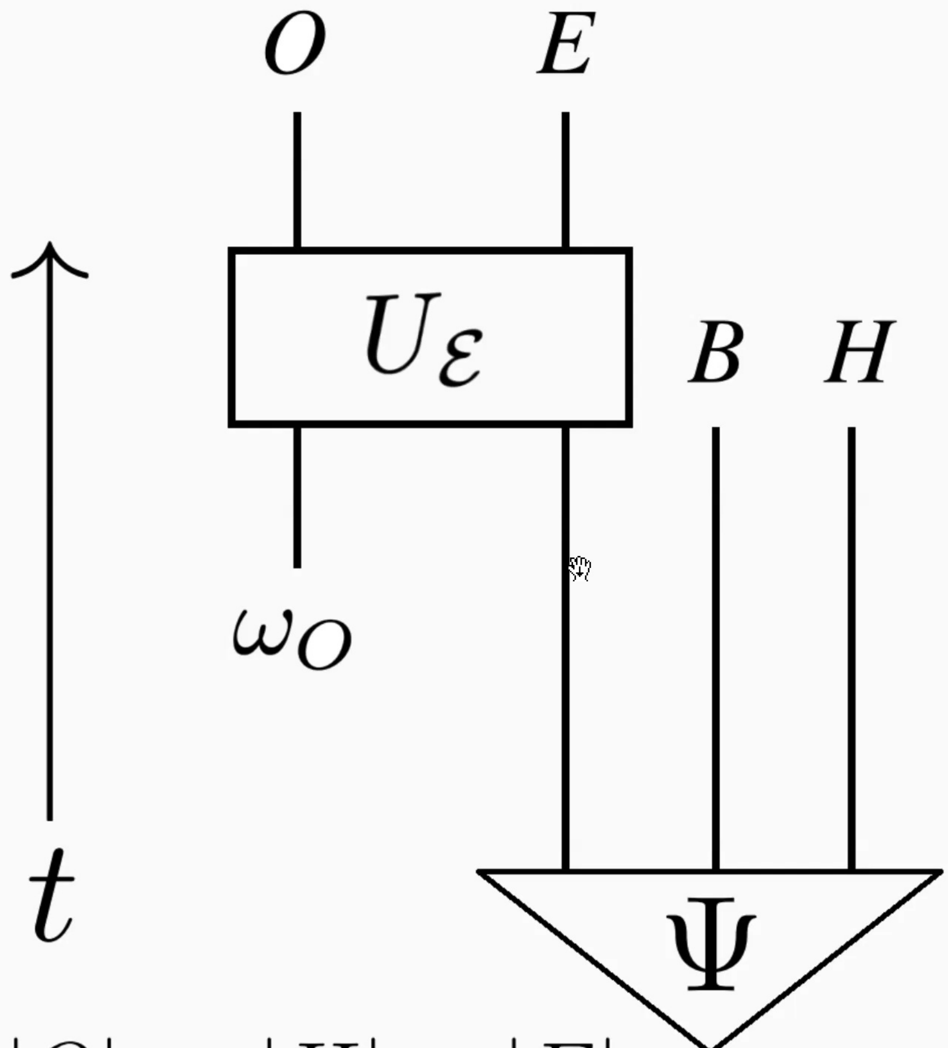
A black hole state cannot be pseudorandom by virtue of being Haar random; black hole formation is an *efficient* process. For pseudorandom states to be useful to us, they must also be *efficiently generated*.

Assuming standard cryptographic primitives, it can be shown that there exists efficiently computable pseudorandom quantum states [1711.00385].

Setup



Setup



$$|O| \ll |H| < |E|$$

12

Limitations of Effective Field Theory

Hayden and Harlow [1301.4504] argued that any experiment witnessing firewalls must be exponentially complex (in the entropy of the black hole).

Exponentially complex operations are unphysical. A physically reasonable effective field theory should come with restrictions not just to low energies, *but also low complexities*.

Hawking radiation appears thermal *to a computationally bounded observer*.

Let us explore the consequences for taking this modification seriously.

Let $|\Psi\rangle_{EBH}$ denote the state of the black hole and the exterior radiation. Let $\sigma_{EB} = I_{EB}/d_{EB}$ be the maximally mixed state of EB, and let $\rho_{EB} = \mathrm{Tr}_H(|\Psi\rangle\langle\Psi|)$.

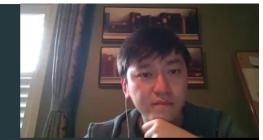
Definition: We say that the state $|\Psi\rangle_{EBH}$ is *pseudorandom* on the radiation EB if there exists some constant $\alpha > 0$ such that

$$\Big|\Pr[\mathcal{M}(\rho_{EB}) = 1] - \Pr[\mathcal{M}(\sigma_{EB}) = 1]\Big| \leq 2^{-\alpha|H|}, \tag{14}$$

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

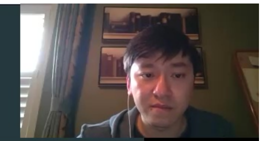
The pseudorandomness hypothesis can be seen as an axiomization of the thermality of Hawking radiation for Harlow-Hayden type effective field theories.

Implications of Pseudorandomness



We show the following:

1. There exists an encoding $V : \mathcal{H}_{\tilde{B}} \to \mathcal{H}_{EH}$ of each interior mode $\tilde{B}$ into the Hilbert space of the exterior radiation.
2. The encoding V defines a quantum error-correcting code, which we call the black hole code.
3. The black hole code protects against all operations performed by a computationally bounded observer. Specifically, the code corrects against all channels of sufficiently small complexity and Kraus rank.
4. Moreover, there exists a complete set of logical operators, the *ghost operators*, which commutes with all correctable errors of the code. The ghost operators serve as witness to the smoothness of the horizon and the preservation of causality.

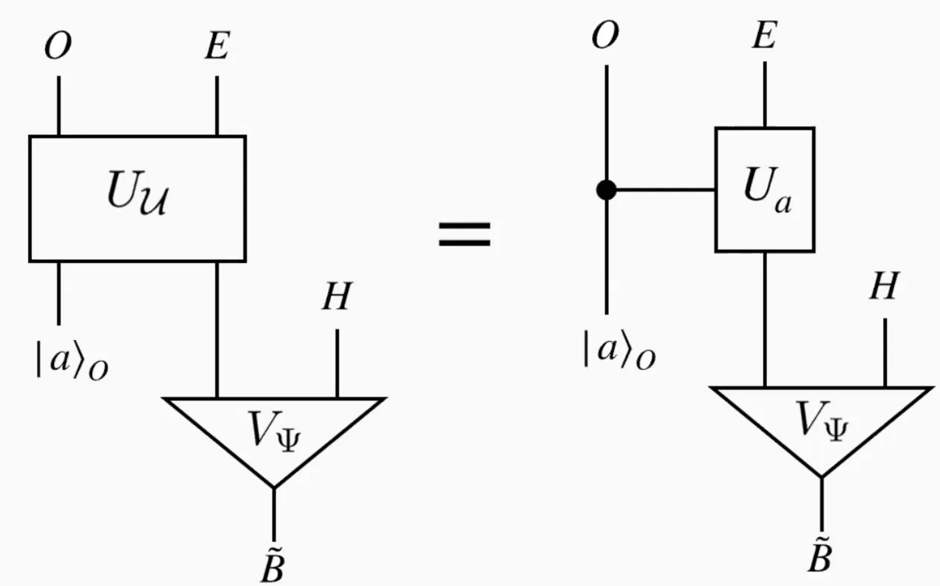


A Concrete Model

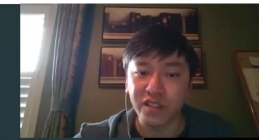
Imagine a scenario where an agent has access to a quantum device with limited memory. The agent is able to program the device to perform one of a long list of unitaries $\mathcal{U} = \{U_1, \cdots, U_N\}$ to act on the early radiation E.

The net action of the agent can be modeled as a controlled unitary

$$U_{\mathcal{U}} = \sum_{a=0}^{N} |a\rangle\langle a|_O \otimes (U_a)_E. \tag{15}$$



A Concrete Model



Theorem: Let $V : \mathcal{H}_{\tilde{B}} \to \mathcal{H}_{EH}$ denote the black hole code embedding. Let $\mathcal{U} = \{U_a\}_{a=1}^{N}$ denote an arbitrary set of $N = \mathrm{poly}(|H|)$ unitaries acting on the early radiation E, where each unitary has complexity $\mathrm{poly}(|H|)$. Suppose that the pseudorandomness hypothesis holds.

Then there exists a complete set of logical operators $\mathcal{L} \subseteq \mathcal{B}(\mathcal{H}_{EH})$ for the black hole code, such that for all $T \in \mathcal{L}$, and all $U_a \in \mathcal{U}$, we have

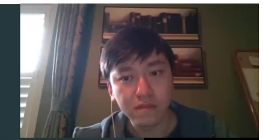
$$\|TV - V\tilde{T}\| \leq \delta \|\tilde{T}\|, \tag{16}$$

$$\|[U_a, T]V\| \leq 2\delta \|\tilde{T}\|, \tag{17}$$

where $\tilde{T}$ is the operator on $\tilde{B}$ corresponding to T, and where

$$\delta \approx N \cdot 2^{-\alpha|H|/4}. \tag{18}$$

A Concrete Model



Theorem: Let $V : \mathcal{H}_{\tilde{B}} \to \mathcal{H}_{EH}$ denote the black hole code embedding. Let $\mathcal{U} = \{U_a\}_{a=1}^{N}$ denote an arbitrary set of $N = \mathrm{poly}(|H|)$ unitaries acting on the early radiation E, where each unitary has complexity $\mathrm{poly}(|H|)$. Suppose that the pseudorandomness hypothesis holds.

Then there exists a complete set of logical operators $\mathcal{L} \subseteq \mathcal{B}(\mathcal{H}_{EH})$ for the black hole code, such that for all $T \in \mathcal{L}$, and all $U_a \in \mathcal{U}$, we have

$$\|TV - V\tilde{T}\| \leq \delta \|\tilde{T}\|, \tag{16}$$

$$\|[U_a, T]V\| \leq 2\delta \|\tilde{T}\|, \tag{17}$$

where $\tilde{T}$ is the operator on $\tilde{B}$ corresponding to T, and where

$$\delta \approx N \cdot 2^{-\alpha|H|/4}. \tag{18}$$

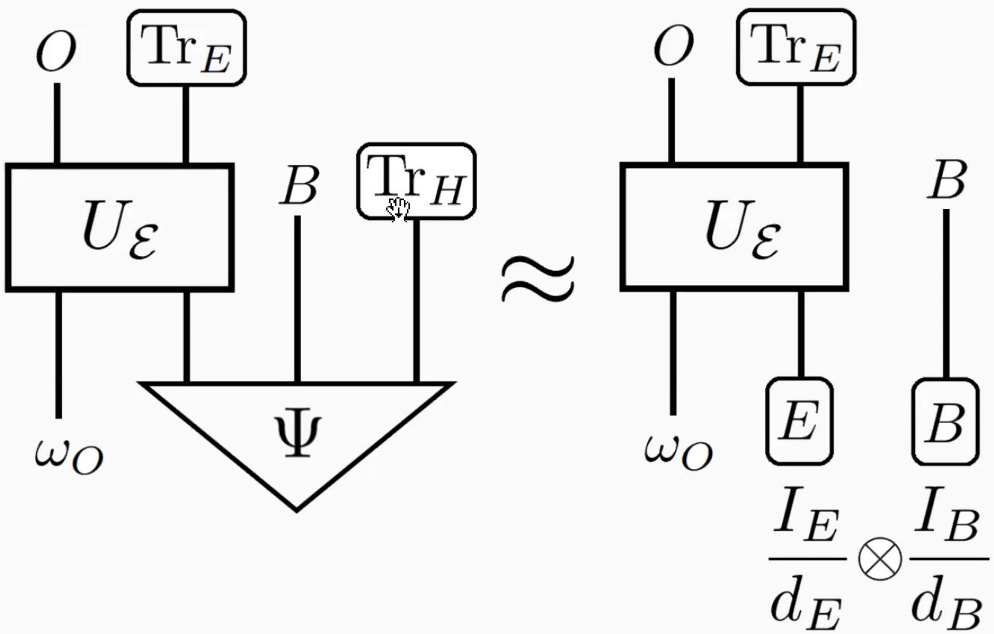
The black hole interior is protected from any action that an external agent can perform, so long as the actions are computationally limited.

The Decoupling Bound

The pseudorandomness hypothesis immediately allows us to prove a key decoupling bound.

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

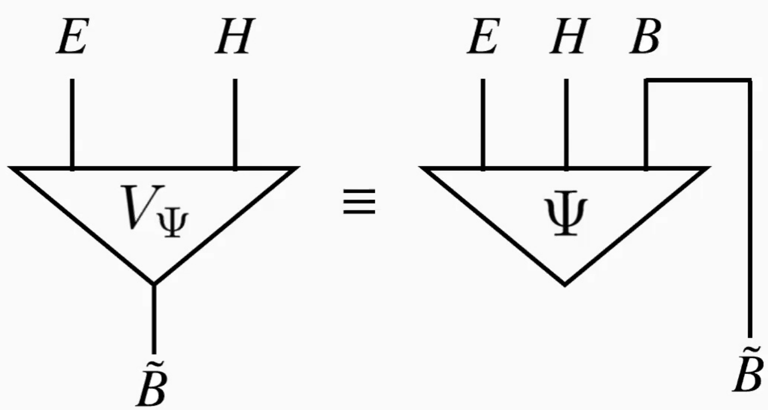
$$\|\rho_{OB} - \rho_O \otimes \rho_B\|_1 \leq 6 \cdot 2^{-(\alpha|H|-|O|)}. \quad (19)$$



The Black Hole Encoding

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

$$V_\Psi |i\rangle_{\tilde{B}} = 2\left(I_{EH} \otimes \langle\omega|_{B\tilde{B}}\right)\left(|\Psi\rangle_{EHB} \otimes |i\rangle_{\tilde{B}}\right). \tag{20}$$



One can use the decoupling bound to show that V_Ψ is ϵ-close to an isometric embedding, where $\epsilon = 2 \cdot 2^{-\alpha|H|}$.

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).

The Information-Disturbance Tradeoff

Having identified a suitable codespace, we now characterize its error-correcting capabilities.

The *information-disturbance relation* says that a code can protect quantum information from noise if and only if the "environment" of the noise channel $\mathcal{E}$ learns nothing about the logical information.

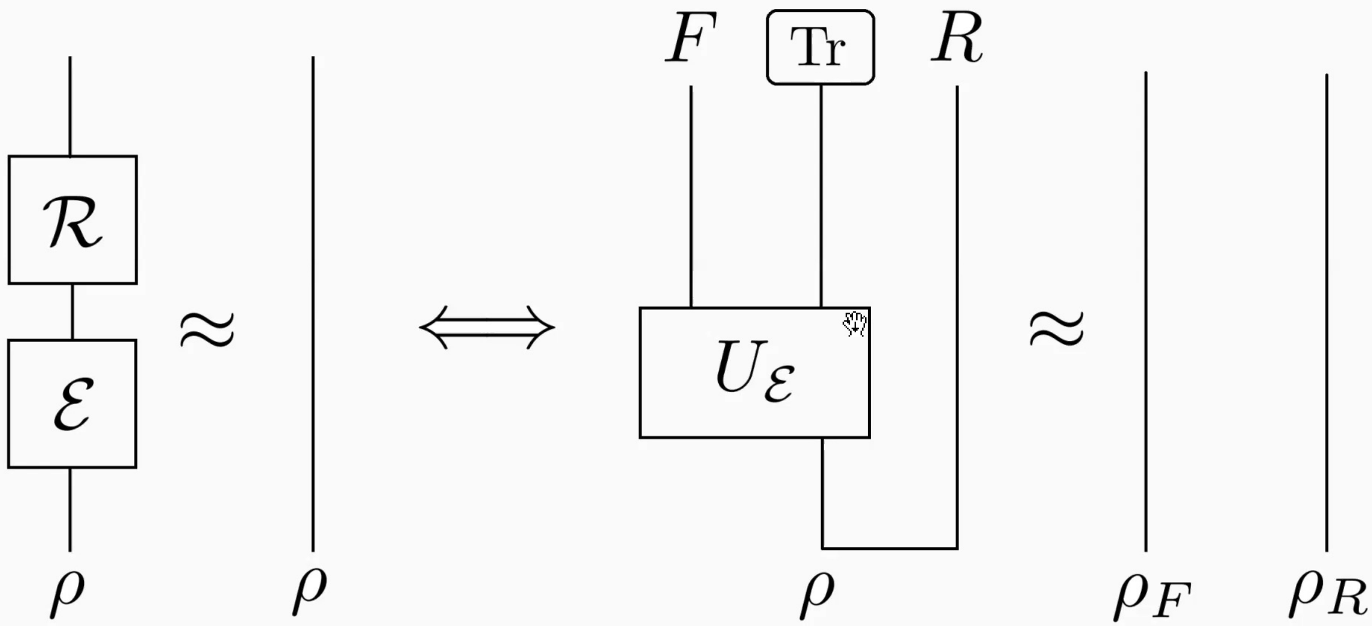
Specifically, there is a physical process $\mathcal{R}$, the recovery process, which reverses some error $\mathcal{E}$:

$$\mathcal{R} \circ \mathcal{E} \approx \mathcal{I}, \tag{21}$$

if and only if the reference system that purifies the quantum error-correcting code decouples from the environment of $\mathcal{E}$.

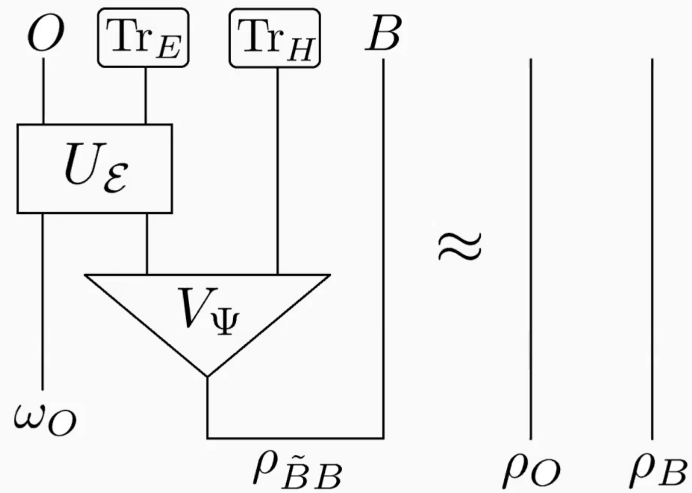
The Information-Disturbance Tradeoff

Diagrammatically, the information-disturbance relation can be depicted as:



The Black Hole Error-Correcting Code

Compare the information-disturbance tradeoff to the decoupling condition:

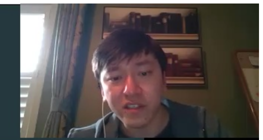


This shows that the black hole code corrects against all channels of polynomial complexity in $|H|$ which have sufficiently low-rank.

Theorem: 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

$$\epsilon = \sqrt{\frac{3}{2}} \cdot 2^{-(\alpha|H|-|O|)/2}. \tag{22}$$

Ghost Logical Operators



We can recast the error-correcting conditions of the black hole code into an algebraic form.

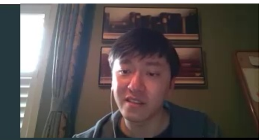
Let $T_{\tilde{B}}$ be an operator which acts on the interior mode $\tilde{B}$. By microcausality, any operator $\mathcal{O}$ that an exterior agent can apply onto the early radiation should commute with $T_{\tilde{B}}$, i.e.,

$$[\mathcal{O}, T_{\tilde{B}}] = 0. \tag{23}$$

The corresponding logical operator T_{EH} on the code subspace should satisfy the same relation when restricted to the code subspace:

$$[\mathcal{O}, T_{EH}]|\Psi\rangle_{EBH} = 0. \tag{24}$$

We will call operators which satisfy such a relation *ghost logical operators*.



Ghost Logical Operators

Theorem: Let T_B be any operator acting on an outgoing mode B. Then there exists a corresponding logical operator T_{EH} such that

$$T_B|\Psi\rangle_{EBH} \approx T_{EH}|\Psi\rangle_{EBH}, \tag{25}$$

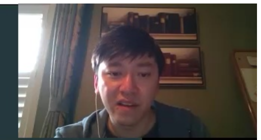
$$[E_a, T_{EH}]|\Psi\rangle_{EBH} \approx 0, \tag{26}$$

where $\{E_a\}$ is some set of operators that a computationally bounded external observer can apply on the radiation (such that the decoupling bound holds).

The first relation serves as a witness to the entanglement between the outgoing mode and the encoded interior mode, as required for the existence of a smooth horizon.

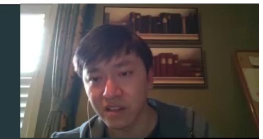
The second relation serves as a witness to microcausality between the encoded interior and an external observer, as required to maintain the correct causal structure.

Conclusion



- We find that the $\tilde{B} \subseteq E$ proposal for the firewall problem can be given reasonable meaning within the context of quantum information theory and quantum complexity.
- Under a reasonable pseudorandomness hypothesis, black holes form natural error-correcting codes protecting against low-complexity operations.
- Error-correcting properties of black holes lead to the existence of ghost operators, which certify the preservation of geometry and causality.
- Operations with large complexity or high rank can see violations of causality and create firewalls. Effective field theory should be restricted to low complexity, not simply low energy.

Addendum 1 - Efficient Pseudorandomness



Specifically, under the assumption that there exists a family of *quantum-secure pseudorandom functions* $\{\mathrm{PRF}_k\}_{k\in K}$, we may efficiently prepare states which are computationally pseudorandom.

Consider the family of states defined by

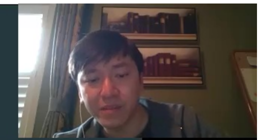
$$|\phi_k\rangle = \frac{1}{\sqrt{|X|}} \sum_{x\in X} \omega_N^{\mathrm{PRF_k}(x)} |x\rangle.$$

These states are efficiently computable via a Fourier transform, since the family of pseudorandom functions $\{\mathrm{PRF}\}_{k\in K}$ is efficiently computable.

Then it can be shown that sampling from $\{|\phi_k\rangle\}_{k\in K}$ is computationally indistinguishable from the maximally mixed state [1711.00385]:

$$\{|\phi_k\rangle\}_{k\in K} \sim \{|f\rangle\}_{f\in X^X} \sim \mathrm{Haar}. \tag{27}$$

Addendum 2 - Ghost Operators - Proof Sketch



Let $\tilde{\mathcal{H}}$ be an abstract logical Hilbert space, and consider an encoding $V : \tilde{\mathcal{H}} \to \mathcal{C} \subseteq \mathcal{H}$. Let $\mathcal{E}$ be a correctable error channel for $\mathcal{C}$, with Kraus representation given by

$$\mathcal{E}(\rho) = \sum_{a=1}^{|K|} E_a \rho E_a^\dagger, \tag{28}$$

where we denote the set of Kraus operators as $K = \{E_a\}$.

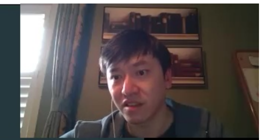
Let

$$\tilde{T} = \sum_k \lambda_k |\tilde{k}\rangle\langle\tilde{k}| \tag{29}$$

be a normal operator on $\tilde{\mathcal{H}}$, with eigenvalues $\{\lambda_k\}$, and eigenbasis $\{|\tilde{k}\rangle\}$.

A natural first attempt at defining ghost logical operators might proceed as follows....

Ghost Operators - Proof Sketch



Let $|k\rangle = V|\tilde{k}\rangle$ be the encoded eigenbasis. Consider the subspaces

$$S_k = \mathrm{span}\{E_a|k\rangle \mid a = 1, \cdots, |K|\}. \tag{30}$$

By the Knill-Laflamme conditions, the spaces S_k are mutually orthogonal for distinct k. Let P_k denote the projection onto S_k. Define the operator

$$T = \sum_k \lambda_k P_k. \tag{31}$$

Then these operators T satisfy

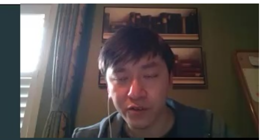
$$TE_a|j\rangle = \sum_k \lambda_k P_k E_a|j\rangle = \lambda_j E_a|j\rangle = E_a \hat{T}|j\rangle, \tag{32}$$

where $\hat{T}$ is any logical operator for $\tilde{T}$. In particular, if T itself is a logical operator for T, then it would satisfy

$$[E_a, T]V = 0 \tag{33}$$

for all $E_a \in K$.

Ghost Operators - Proof Sketch



We seek conditions where T, possibly after a suitable extension, becomes a logical operator $\tilde{T}$.

Theorem: An operator T constructed as before can be extended to be a ghost logical operator for $\tilde{T}$ if and only if

$$\langle i|E_a|j\rangle = 0, \tag{34}$$

for all $E_a \in K$ and all i, j where $\lambda_i \neq \lambda_j$.

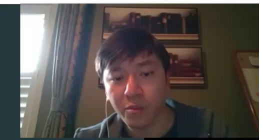
Proof Sketch: This condition ensures that the code subspace is in the orthogonal complement of the subspaces S_k defined before. We can therefore extend the action of T onto the code subspace.

Conversely, if T extends to a ghost logical operator, say T', then we must have $T'|j\rangle = \lambda_j|j\rangle$. Left multiplying by $\langle i|E_a^\dagger$, we get

$$\lambda_j\langle i|E_a^\dagger T'|j\rangle = \langle i|E_a^\dagger T'|j\rangle = \langle i|T'E_a^\dagger|j\rangle = \lambda_i\langle i|E^\dagger|j\rangle. \tag{35}$$

If $\lambda_i \neq \lambda_j$, then we must have $\langle i|E_a|j\rangle = 0$.

Complete Set of Ghost Operators



We see that the existence of ghost logical operators has to do not with just the correctability of the errors in K, but rather the correctability of $K \cup \{I\}$.

A succinct way to relate the ghost operators to the correctability of the code subspace is to require a *complete set* of ghost logical operators, i.e., to require the existence of a ghost logical operator corresponding to any operator on $\tilde{\mathcal{H}}$.

Theorem: Let $\mathcal{E}_{\mathcal{I}} = \mathcal{E}/2 + \mathcal{I}/2$. Then there exists a complete set of ghost logical operators for $\mathcal{E}$ if and only if $\mathcal{E}_{\mathcal{I}}$ is a correctable channel.

Proof Sketch: If $\mathcal{E}_{\mathcal{I}}$ is correctable, then the Knill-Laflamme conditions for $\mathcal{E}_{\mathcal{I}}$ means that equation (34) is automatically satisfied for any choice of $\tilde{T}$.