Black Holes as Mirrors of Quantum Information - The Hayden-Preskill Thought Experiment (PHYS 364 Final Paper)

Rio Weil

This document was typeset on March 13, 2026

Abstract:
We discuss the seminal analysis of Hayden and Preskill [1], who presents an answer to the question: "how quickly can information can be recovered after being thrown into a black hole?". We discuss the surprising answer that - upon viewing black holes as a systems which randomize/scramble information - information discarded into a (sufficiently old) black hole re-emerges rapidly. To make the presentation self-contained, as background we provide a brief review of black hole thermodynamics and quantum information theory. The presentation style is that of a hypothetical lecture.

Contents

1 Introduction

Suppose you had a secret diary that you had to keep from an all-powerful adversary, who was only bound by the laws of physics; how would you best ensure that they had no access to your information? Your senior undergraduate friend, Albert, having just finished his first course in general relativity, may recommend that you throw your diary into a black hole - in lecture he learned that black holes are dense astronomical objects, and beyond the Schwarzschild radius $r < r_s$:

$$r_s = \frac{2GM}{c^2} \tag{1.1}$$

the escape velocity of the diary would surpass the speed of light[1]. This should make your diary impossible to retrieve!

Your first-year PhD student friend, Stephen, takes a small objection to this argument; he syas that contrary to initial belief, black holes are not quite such simple objects. Indeed, they have nonzero entropy [2], and emit thermal/blackbody radiation [3]. So, it's not guaranteed that the diary you throw in won't be re-emitted from the black hole as Hawking radiation. But, he's heard that black holes are conjectured to be fast scramblers of information [4], so your information is probably still safe.

Listening into your conversation, a pair of senior grad students, Patrick and John, chips in their opinion - indeed, assuming that black holes are fast scramblers, they say that (unfortunate for you), the information you throw in re-emerges and can be recovered quite rapidly [1]! And indeed, their argument rests on fairly simple ideas from quantum information - you decide that perhaps there is no keeping your diary safe, after all.

The rest of this article is dedicated to understanding Patrick (Hayden) and John (Preskill)'s argument and unintuitive conclusion. In Section 2, we give a high-level overview of black hole thermodynamics in order to sufficiently understand the setting of their argument, introducing Hawking radiation and black hole entropy. In Section 3, we review basic notions of quantum information theory, introducing density operators, distance measures on quantum states, the no-cloning theorem, and notions of Haar randomness. In Section 4, we present the central argument of [1]. In Section 5 we (not comprehensively) review some salient objections/assumptions of the presented argument. In Section 6 we conclude.

2 A Brief Review of Black Hole Thermodynamics

In this section, we give two derivations for the Hawking temperature (one heuristic, requiring only undergraduate physics knowledge - the other uses more formal tools from general relativity and quantum field theory), and use this to derive black hole entropy and evaporation.

2.1 "Intuitive" Derivation of Hawking Temperature

The heuristic picture/explanation of Hawking radiation is that a pair of virtual photons is emitted at the Schwarzschild radius, one of which falls into the black hole, and the other which is emitted. We can use this heuristic picture to derive the Hawking temperature.

First, if a particle is produced near the (Schwarzschild) black hole horizon, then its uncertainty in position can be approximated as the black hole circumference, given in terms of the Schwarzschild radius:

$$\Delta x = \frac{2\pi r_s}{2} = 2\pi r_s \tag{2.1}$$

[1] A nice way to derive the Schwarzschild radius comes from setting $v_{esc} = c$ in the context of Newtonian gravity, wherein:

$$0 = T + U = \frac{1}{2}mc^2 - \frac{GMm}{r_s} \implies r_s = \frac{2GM}{c^2}. \tag{1.2}$$

We then suppose that we work in the limit where the Heisenberg uncertainty principle is saturated; indeed many discussions of virtual particles in QFT take the uncertainty principle as a starting point to show that for small enough Δx, the momentum/energy uncertainty must be large and hence particle number becomes indeterminate. With this assumption, we have:

$$\Delta x \Delta p = \frac{\hbar}{2} \implies \Delta p = \frac{\hbar}{2\Delta x} = \frac{\hbar}{4\pi r_s} \tag{2.2}$$

The energy of a photon is given by:

$$E = pc \tag{2.3}$$

and so the uncertainty in the energy of the emitted photon is:

$$\Delta E = c\Delta p \tag{2.4}$$

Because for a relativistic photon gas described by a Planckian distribution we have $\langle E \rangle \propto \Delta E \propto k_B T$, we can write:

$$k_B T = \langle E \rangle = c\Delta p = \frac{\hbar c}{4\pi r_s} \implies T = \frac{\hbar c}{4\pi k_B r_s} \tag{2.5}$$

Finally substituting our expression for the Schwarzschild radius from Eq. (1.1), we obtain:

$$T = \frac{\hbar c^3}{8\pi G M k_B} \tag{2.6}$$

which is the Hawking temperature of the black hole; of course, we egregiously cheated our way to this result, but it is nevertheless instructive that many ideas from undergraduate physics gets us to the correct answer.

2.2 QFT Derivation of Hawking Temperature

Let us also derive the Hawking temperature slightly more formally, using tools from general relativity and quantum field theory. We follow the arguments presented in [5], wherein we consider quantum field theory in the presence of a Schwarzschild black hole of mass M.

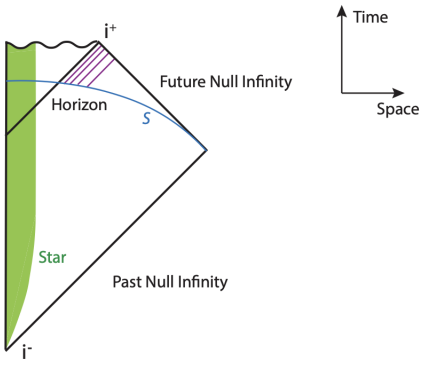

Figure 2.1: Penrose diagram of a Schwarzschild black hole formed by the collapse of a star. We derive the Hawking temperature by considering a hypersurface of $\mathcal{S}$ crossing the black hole horizon outside of the star radius, and consider outgoing radiation/massless signals (purple) arising from this hypersurface. Figure taken from [5].

We consider a black hole formed by a collapsing star, and choose a hypersurface $\mathcal{S}$ crossing the horizon $r = r_S$ outside of the star. On this hypersurface, we wish to find a coordinate u such that u vanishes on the black hole horizon, is positive outside the black hole, and has nonzero normal derivative along the horizon. We take time t to be a time in which a distant observer receives a signal/massless particle from the black hole (which must propagate through S at some time) - we wish to find how the coordinate u depends on the observer time t. To this end, we start with the Schwarzschild solution/metric for a black hole, as seen in class:

$$ds^2 = -\left(1 - \frac{r_s}{r}\right) dt^2 + \frac{dr^2}{1 - \frac{r_s}{r}} + r^2 d\Omega^2 \tag{2.7}$$

We then introduce (U, V) coordinates to introduce coordinates that are applicable both above/below r_s:

$$U = -\left(\frac{r}{r_s} - 1\right)^{1/2} e^{r/2r_s} e^{-t/2r_s} \tag{2.8}$$

$$V = \left(\frac{r}{r_s} - 1\right)^{1/2} e^{r/2r_s} e^{t/2r_s} \tag{2.9}$$

where the metric becomesL:

$$ds^2 = -\frac{4r_s^3}{r} e^{-r/r_s} dU dV + r^2 d\Omega \tag{2.10}$$

Consider a radially null geodesic with $d\Omega = 0$ (polar/azimuthal angles are fixed) and $ds = 0$ (null condition). From the above form of the metric, this implies one of $dU = 0$ or $dV = 0$ and so one of U, V is constant. Specifically, U is constant at an outgoing radial null geodesic, and notably vanishes on the horizon $r_s = 2GM$ and is negative for $r > 2GM$. Thus, our desiredata for u is satified taking:

$$u = -U = Ce^{-t/2r_s} \tag{2.11}$$

with a constant C dependent on observer position r and not time:

$$C = \left(\left(\frac{r}{2r_s} - 1\right)^{1/2} e^{r/2r_s}\right) \tag{2.12}$$

Now, we borrow some tools from quantum field theory. The distant observer measures the radiation emitted by the black hole by studying observables in quantum fields Ψ, which can be expanded in terms of partial waves with $1+1$ dimensional coefficients. Taking the simplest case, we consider the chiral free fermion ψ, which only depends on u. The two point function in the vacuum (vacuum expectation value) between two points u, u' looks like:

$$\left\langle \psi(u)\psi(u') \right\rangle = \frac{(dudu')}{u - u'}. \tag{2.13}$$

Using the vacuum expectation value is justified as for late times t, u, u' are both small and close to each other, and at sufficiently small distances $|u - u'|$ states can be replaced with the vacuum. Then using $u = C'e^{-t/2r_s}$, we obtain the two-point function as a function of observation times t, t':

$$\left\langle \psi(t)\psi(t') \right\rangle = \frac{1}{2r_s} \frac{(dtdt')^{1/2}}{e^{(t-t')/4r_s} - e^{-(t-t')/4r_s}} \tag{2.14}$$

This is antiperiodic in imaginary time, as under $t \to t + 4\pi r_s i$ we can see that:

$$\begin{aligned}\left\langle \psi(t+4\pi r_s i)\psi(t') \right\rangle &= \frac{1}{2r_s}\frac{(dtdt')^{1/2}}{e^{(t+4\pi r_s i-t')/4r_s} - e^{-(t+4\pi r_s i-t')/4r_s}} \\ &= \frac{1}{2r_s}\frac{(dtdt')^{1/2}}{e^{\pi i}e^{(t-t')/4r_s} - e^{-\pi i}e^{-(t+4\pi r_s i-t')/4r_s}} \\ &= \frac{1}{2r_s}\frac{(dtdt')^{1/2}}{(-1)e^{(t-t')/4r_s} - (-1)e^{-(t+4\pi r_s i-t')/4r_s}} \\ &= -\left\langle \psi(t)\psi(t') \right\rangle\end{aligned}$$

And using the Wick correspondence between imaginary time and (inverse) temperature $\frac{it}{\hbar} \leftrightarrow \beta = \frac{1}{k_B T}$, the corresponding thermal correlation function is that with:

$$T = \frac{\hbar}{4\pi r_s k_B} \tag{2.15}$$

and using the definition of r_s and recovering factors of $\hbar, c$, we conclude that the black hole has Hawking temperature:

$$\boxed{T = \frac{\hbar c^3}{8\pi G M k_B}} \tag{2.16}$$

2.3 Black Hole Entropy and Dissipation

We now derive two consequences of black holes having a nonzero temperature - their entropy and that they dissipate.

Bekenstein [2] originally derived the entropy of a black hole via the following line of reasoning. First, Hawking had proved the area theorem [6] :

Theorem

In Classical General Relativity, the area of a black hole can never decrease.

A full proof of the area theorem would be the subject of its own final paper, but the argument for a stationary Schwarzschild black hole is intuitive. As the black hole is stationary, its mass can only increase as objects fall in. Then, observing the Schwarzschild radius is $r_s \propto M$, the area of the black hole can only increase.

Given the area theorem, Beckenstein attempted to construct the entropy as k_B times a dimensionless quantity proportional to the black hole area A. Constructing the Planck length $l_P = (\hbar G/c^3)^{1/2}$ from fundamental units, his guess for the entropy was:

$$S \propto k_B \frac{A}{l_P^2} = k_B \frac{Ac^3}{\hbar G} \tag{2.17}$$

Armed with the Hawking temperature, we can re-derive this result and also get the correct prefactor. From the first law of thermodynamics, we have:

$$dE = TdS \implies dS = \frac{dE}{T} \tag{2.18}$$

where $E = Mc^2$ for a black hole and T is the Hawking temperature derived in the previous section, so:

$$dS = dMc^2 \cdot \frac{8\pi G M k_B}{\hbar c^3} = dM \frac{8\pi G M k_B}{\hbar c} \tag{2.19}$$

integrating from $M = 0$ (where $S = 0$ as no black hole exists) to M finite, we obtain:

$$S = \frac{4\pi G M^2 k_B}{\hbar c} \tag{2.20}$$

Now writing this in terms of $r_s = \frac{2GM}{c^2}$, we have:

$$S = \frac{\pi r_s^2 c^3 k_B}{\hbar G} \tag{2.21}$$

and recognizing $A = 4\pi r_s^2$ the above becomes:

$$\boxed{S = k_B \frac{A c^3}{4\hbar G}} \tag{2.22}$$

which is the Bekenstein-Hawking entropy, now with the correct prefactor.

The nonzero temperature of black holes also gives means they dissipate; their luminosity as a black body is determined by the Stefan-Boltzmann law:

$$\frac{dE}{dt} = L = \sigma A T^4 \tag{2.23}$$

With our expressions for the area in terms of the Schwarzschild radius and the Hawking temperature, this becomes:

$$\frac{dM}{dt} c^2 = -\sigma 4\pi \left(\frac{2GM}{c^2}\right)^2 \left(\frac{\hbar c^3}{8\pi G M k_B}\right)^4 \implies \frac{dM}{dt} = -\text{const.} \frac{1}{M^2} \tag{2.24}$$

which tells us that black holes evaporate via Hawking radiation (though very slowly, given their large mass).

3 A Brief Review of Quantum Information Theory

3.1 Density Operators

The standard presentation of states in quantum mechanics is as state vectors/kets $|\psi\rangle$ that live in Hilbert spaces $\mathcal{H} = \mathbb{C}^d$. The density operator ρ is a generalization of the state ket $|\psi\rangle$, which allows for the description of classical randomness/noise in quantum systems. It also is indispensable for local descriptions of quantum systems. More precisely, recall the notion of entanglement:

Definition: (Bipartite) Entanglement

Suppose $|\Psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$, i.e. a state in a composite Hilbert space. We say that $|\Psi\rangle$ is (bipartite) *entangled* if there are no $|\psi_1\rangle \in \mathcal{H}_1, |\psi_2\rangle \in \mathcal{H}_2$ for which:

$$|\Psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle. \tag{3.1}$$

Clearly, from the definition above there exists no local/subsystem description of an entangled $|\Psi\rangle$ in terms of a state kets $|\psi_1\rangle$. But density operators will allow for this.

Definition: Density operators

Consider an ensemble of quantum states $\mathcal{E} = (p_i, |\psi_i\rangle)$ with p_i the probabilities, satisfying $\sum_i p_i = 1$. Then, we can define a density operator:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{3.2}$$

Note that in the case where only one state is part of the ensemble, there is an exact correspondence to the state ket picture, wherein the states are *pure*:

Definition: Pure and Mixed states

If there exists some $|\psi\rangle$ for which $\rho = |\psi\rangle\langle\psi|$, then ρ is *pure* (and can be put into one-to-one correspondence with the state ket). Otherwise, it is *mixed*.

Note that the above condition can be shown to be equivalent to the fact that $\text{Tr}(\rho^2) = 1$ for pure states and $\text{Tr}(\rho^2) \leq 1$ for mixed states. An example of a state which is not pure is $\rho = \frac{I}{d}$, which is called the *maximally mixed state*.

Taking expectation values of observables O for states $|\psi\rangle$ as $\langle O\rangle = \langle\psi|O|\psi\rangle$ generalize as:

$$\langle O\rangle_\rho = \text{Tr}(\rho O). \tag{3.3}$$

And unitary evolution $|\psi\rangle \to U|\psi\rangle$ of pure states generalizes as:

$$\rho \to U\rho U^\dagger. \tag{3.4}$$

It is clear from the definition how density operators allow for classical randomness in quantum systems. Let us also follow up on our second claim that it allows for a local description of general (even entangled) quantum states. This is done via the notion of a reduced density operator:

Definition: Reduced Density Operators

Let ρ be a composite state in Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. The reduced density operator on $\mathcal{H}_1$ is obtained as:

$$\rho_1 = \text{Tr}_2(\rho) \tag{3.5}$$

where Tr_2 is the partial trace over subsystem 2. ρ_2 is defined analogously.

Note that given ρ_1, all local observables can then be probed via $\langle O_1\rangle = \text{Tr}(\rho_1 O_1)$. For a product state $|\Psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$, note that $\rho_1 = |\psi_1\rangle\langle\psi_1|$ is just the pure state, as we might expect. For $|\Psi\rangle$ entangled, ρ_1 will be generically a mixed state. For example, consider the maximally entangled state $|\Psi\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^d |i^1\rangle|i^2\rangle$. Then:

$$\rho_1 = \text{Tr}_2(|\Psi\rangle\langle\Psi|) = \frac{1}{d}\sum_{k=1}^d\sum_{i=1}^d\sum_{j=1}^d \langle k^2||i^1\rangle|i^2\rangle\langle j^1|j^2|k^2\rangle = \frac{1}{d}\sum_{k=1}^d\sum_{i=1}^d\sum_{j=1}^d \delta_{ki}\delta_{ji}|i^1\rangle\langle j^1| = \frac{I_1}{d} \tag{3.6}$$

so we see that the reduced density operator of a maximally entangled state is the maximally mixed state!

We end off this section by defining the notion of purification. If reduced density operators are a way in which we can take a global/composite quantum state and obtain a local description, a purification of a (generically, mixed) quantum state is a way in which we can promote a mixed state to a pure (composite) quantum state.

Definition: Purification

The purification of a quantum state $\rho^A = \sum_i p_i|i^A\rangle\langle i^A|$ in $\mathcal{H}_A$ is the following state in $\mathcal{H}_A \otimes \mathcal{H}_C$:

$$|AC\rangle = \sum_i \sqrt{p_i}|i^A\rangle|i^C\rangle. \tag{3.7}$$

By definition $|AC\rangle$ is pure, and we can check that the reduced density operator ρ_A for $|AC\rangle$ reduces to our original ρ:

$$\mathrm{Tr}_C(|AC\rangle\langle AC|) = \sum_{ij} \sqrt{p_i p_j}|i^A\rangle\langle j^A|\mathrm{Tr}(|i^C\rangle\langle j^C|) = \sum_{ij} \sqrt{p_i p_j}|i^A\rangle\langle j^A|\delta_{ij} = \sum_i p_i |i^A\rangle\langle i^A| = \rho^A. \tag{3.8}$$

3.2 Distances between quantum states

One question that arises in quantum information (and shall soon become relevant in our black hole discussion) is "how far apart are two quantum states ρ, σ"? There are two quantities that are commonly used to capture this notion of state distance. The first is the trace, or L_1 distance:

Definition: Trace Distance

The trace distance $D(\rho, \sigma)$ is defined as:

$$D(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1 \tag{3.9}$$

where $\|A\|_1 = \mathrm{Tr}(\sqrt{A^\dagger A})$.

This can be seen as the quantum analog of the L_1 distance of probability distributions, and thus embues the trace distance with an operational meaning - it measures the ability to distinguish two quantum states from the probability distributions arising from measurements.

Another notion of state distance is given by the fidelity:

Definition: Fidelity

The fidelity of two quantum states ρ, σ is given as:

$$F(\rho, \sigma) = \sqrt{\rho^{1/2}\sigma\rho^{1/2}}. \tag{3.10}$$

which is the quantum analog of the fidelity of two probability distributions. As will become useful in our later discussion, the fidelity between a pure state and a general state is given by:

$$F(|\psi\rangle, \sigma) = \sqrt{\langle\psi|\sigma|\psi\rangle}. \tag{3.11}$$

Note that both the trace distance and fidelity have nice properties as notions of distance, given that they both are metrics[2], are contractive under quantum channels, and are (strongly) convex [7]. They also provide bounds for each other [7]:

$$1 - F(\rho, \sigma) \leq D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)^2}. \tag{3.12}$$

And for the case of comparing a pure state and an arbitrary state, a stronger bound:

$$F^2(|\psi\rangle, \sigma) \geq 1 - D(|\psi\rangle, \sigma) \tag{3.13}$$

holds.

[2]Subtle note: the fidelity is not directly a metric, but the angle between states $A(\rho, \sigma) = \arccos F(\rho, \sigma)$ does satisfy the conditions of being a metric

3.3 The No-Cloning Theorem

A fundamental theorem of quantum information theory is the *No-Cloning theorem*; we introduce and prove it here as the Hayden-Preskill argument will later show that it may be called into question.

Theorem: No-Cloning

There is no quantum-mechanical process, i.e. a linear map $T : \mathcal{H} \otimes \mathcal{H} \to \mathcal{H} \otimes \mathcal{H}$, that can copy a quantum state $|\psi\rangle$, i.e

$$T(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle \tag{3.14}$$

for an arbitrary $|\psi\rangle \in \mathcal{H}$ and some reference state $|s\rangle \in \mathcal{H}$.

Proof. Suppose for the take of contradiction T exists. Now, consider $|\psi_1\rangle, |\psi_2\rangle$ orthogonal. By assumption:

$$T(|\psi_1\rangle \otimes |s\rangle) = |\psi_1\rangle \otimes |\psi_1\rangle \tag{3.15}$$

$$T(|\psi_2\rangle \otimes |s\rangle) = |\psi_2\rangle \otimes |\psi_2\rangle \tag{3.16}$$

Now, consider applying the cloning procedure to the state $|\psi_+\rangle = \frac{|\psi_1\rangle + |\psi_2\rangle}{\sqrt{2}}$. By the cloning assumption, we have:

$$T(|\psi_+\rangle \otimes |s\rangle) = |\psi_+\rangle \otimes |\psi_+\rangle = \frac{1}{2}(|\psi_1\rangle \otimes |\psi_1\rangle + |\psi_1\rangle \otimes |\psi_2\rangle + |\psi_2\rangle \otimes |\psi_1\rangle + |\psi_2\rangle \otimes |\psi_2\rangle) \tag{3.17}$$

but by the assumption that T is linear, we have:

$$T(|\psi_+\rangle \otimes |s\rangle) = \frac{1}{\sqrt{2}}T(|\psi_1\rangle \otimes |s\rangle) + \frac{1}{\sqrt{2}}T(|\psi_2\rangle \otimes |s\rangle) = \frac{1}{\sqrt{2}}(|\psi_1\rangle \otimes |\psi_1\rangle + |\psi_2\rangle \otimes |\psi_2\rangle) \tag{3.18}$$

But the RHS of Eqs. (3.17), (3.18) do not agree - contradiction. Indeed, such a cloning map can only exist for orthogonal quantum states, and the contradiction arises when attempting to clone states that are not orthogonal to each other. □

Note that while the above proof and statement used pure state vectors, the argument and result follow identically if one were to replace the state kets appearing above with density operators ρ.

3.4 Haar Randomness

What is a random operation in a quantum context? In the classical context, we may picture an operation where we apply a random permutation to our input (n-bitstring) - but in the quantum world the space of accessible operations is in some sense much larger (arbitrary unitary operations on a 2^n dimensional Hilbert space of n qubits, for example). To this end, we define the unitary group:

Definition: Unitary group

The unitary group $U(d)$ consists of the set of operators $U : \mathbb{C}^d \times \mathbb{C}^d \to \mathbb{C}^d \times \mathbb{C}^d$ such that $U^\dagger U = I$.

and on the unitary group, the Haar measure (the natural measure for a uniform matrix group) [8]:

Definition: Haar measure

The Haar measure on the unitary group $U(d)$ is the unique probability measure μ_H that is left/right invariant under $U(d)$, i.e:

$$\int_{U(d)} f(U)d\mu_H(U) = \int_{U(d)} f(UV)d\mu_H(U) = \int_{U(d)} f(VU)d\mu_H(U) \tag{3.19}$$

Note then that the integral of a matrix function $f(U)$ over the Haar measure is just the expectation value w.r.t. μ_H:

$$\mathbb{E}_{U\sim\mu_H}[f(U)] = \int_{U(d)} f(U)d\mu_H(U) \tag{3.20}$$

From this measure, we can define the notion of a Haar random operator:

Definition: Haar Random Unitaries

A Haar random unitary U_H is one sampled uniformly from $U(d)$ with respect to the Haar measure μ_H.

Note that generically, applying a Haar random unitary will be inefficient; this arises from the exponential size of the unitary group for generic quantum systems. We illustrate this with the following proposition:

Proposition: Hardness of approximating unitaries

Generic unitaries on n qubits require $m = \Omega(2^n \log(1/\epsilon))$ operations to approximate within error ϵ.

Proof. We follow the argument of [7]. Suppose we have n qubits, initialized in the computational basis state $|0\rangle^{\otimes n}$. If we have a quantum circuit consisting of m total gates, and have g gate types that work on a range of f qubits, there are $\binom{n}{f}^g$ choices per gate in the circuit and a total of $O(n^{fgm})$ final states of the circuit.

Our goal is to target a particular state $|\psi\rangle$ to distance ϵ. (this is equivalent to targeting a particular unitary U for which $U|0\rangle^{\otimes n} = |\psi\rangle$). We thus imagine covering the space of possible n-qubit states - the $2^{n+1}-1$ sphere- with spheres of radius ϵ, which are close to the volume of a $(2^{n+1}-2)$-sphere of radius ϵ. The number N of ϵ-patches necessary (using formulas of the surface area/volume of hyperspheres) goes like:

$$N = \frac{S_{2^{n+1}-1}(1)}{V_{2^{n+1}-2}(\epsilon)} = \frac{\sqrt{\pi}\Gamma(2^n - \frac{1}{2})(2^{n+1}-1)}{\Gamma(2^n)\epsilon^{2^{n+1}} - 1} \geq \frac{\sqrt{\pi}\Gamma(2^n)(2^{n+1}-1)}{\Gamma(2^n)2^n\epsilon^{2^{n+1}} - 1} \implies N = \Omega(\frac{1}{\epsilon^{2^{n+1}-2}}) \tag{3.21}$$

In order to reach all N ϵ-patches, we therefore have:

$$O(n^{fgm}) \geq \Omega(\frac{1}{\epsilon^{2^{n+1}-2}}) \implies m = \Omega\left(\frac{2^n \log(1/\epsilon)}{\log(n)}\right) \tag{3.22}$$

which proves the claim. □

Thus, although Haar random unitaries are very useful constructions in making arguments, the above argument seems to suggests that processes that apply Haar random unitaries are unphysical. One line of reasoning has been to consider unitaries that approximate Haar random unitaries, known as unitary t-designs [8]:

Definition: Unitary t-design

Let ν be a probability distribution over a set of unitaries $S \subseteq U(d)$. Then, ν is a unitary t-design if and only if:

$$\mathbb{E}_{U\sim\nu}[U^{\otimes t} O U^{\dagger\otimes t}] = \mathbb{E}_{V\sim\nu}[V^{\otimes t} O V^{\dagger\otimes t}] \tag{3.23}$$

Though the above definition is quite formal, the intuition/operational meaning of the definition is that a unitary t-design is a more restricted/smaller ensemble of unitary matrices, which mimics the Haar

measure in so far as the first t moments of the ensemble coincide. Any experiment with at most t copies of U cannot distinguish whether it arose from a t-design, or drawn randomly from Haar.

Though we do not describe specific arguments/constructions here, it turns out that unitary t-designs have efficient quantum circuit constructions (unlike their Haar counterparts) - the state of the art [9] is that an ϵ-approximate t-design can be realized in circuit depth:

$$d = O(\log(n/\epsilon)t\text{poly}(\log t)) \tag{3.24}$$

which is only logarithmic (and hence efficient!) in the number of qubits.

3.5 One-Shot Decoupling Theorem

There is a technical bound that appears in the argument of cite- it is central to the argument/conclusion - the proof of which we reproduce from [10].

Theorem: One-shot decoupling

Consider a density matrix ρ^{AE} and a (Haar) random pure state $|\psi\rangle^{RBE}$, with $R \subset A$. then, letting $\sigma^R_{\max}$ be the maximally mixed state on R, we have that:

$$\mathbb{E}\left\|\rho^{RE} - \sigma^R_{\max} \otimes \phi^E\right\|_1 \leq \sqrt{|R||E|\text{Tr}[(\rho^{AE})^2]} \tag{3.25}$$

with the expectation value $\mathbb{E}$ taken over the Haar measure.

Proof. Applying the Cauchy-Shwartz inequality and the convexity of the square root function to the quantity we want to bound, we find:

$$\mathbb{E}\left\|\rho^{RE} - \sigma^R_{\max} \otimes \rho^E\right\|_1 \leq \mathbb{E}\sqrt{|R||E|\left\|\rho^{RE} - \sigma^R_{\max} \otimes \rho^E\right\|_2^2} \leq \sqrt{|R||E|\text{Var}[\rho^{RE}]} \tag{3.26}$$

We thus want to bound the variance of ρ^{RE}; let us to this end rewrite:

$$\text{Var}[\psi^{RE}] \equiv \mathbb{E}\left\|\rho^{RE} - \mathbb{E}\rho^{RE}\right\|_2^2 = \mathbb{E}\left\|\rho^{RE} - \sigma^R_{\max} \otimes \rho^E\right\|_2^2 = \mathbb{E}\text{Tr}[(\rho^{RE})^2] - \frac{1}{d}\mathbb{E}\text{Tr}[(\rho^E)^2] \tag{3.27}$$

Now, let us look at the first piece, which we rewrite as:

$$\mathbb{E}\text{Tr}[(\rho^{RE})^2] = \frac{D^2}{d^2}\mathbb{E}\text{Tr}\left[\left((PU \otimes I^E)\rho^{AE}(U^\dagger P \otimes I^E)\right)^2\right] \tag{3.28}$$

where P is the projection operator onto a subspace $R \subset A$, and the respective dimensions $D = |A|$ and $d = |R|$.

Defining F_A to be the swap operator acting on two copies of A, F_E to be the swap operator on two copies of E, and $F_{AE} = F_A \otimes F_E$. Then we note the identity (used in the SWAP test to calculate the purity of a density operator using two copies):

$$\text{Tr}[F_{AE}(\rho^{AE} \otimes \rho^{AE})] = \text{Tr}[(\rho^{AE})^2] \tag{3.29}$$

And this implies that:

$$\mathbb{E}\text{Tr}[(\rho^{RE})^2] = \frac{D^2}{d^2}\text{Tr}[(\rho^{AE} \otimes \rho^{AE})(G \otimes I^{EE})] \tag{3.30}$$

with:

$$G = \mathbb{E}[(U^\dagger \otimes U^\dagger)F_R(U \otimes U)] \tag{3.31}$$

where F_R is the SWAP operator on the projected subspace:

$$F_R = (P \otimes P) F_A (P_{\otimes} P). \tag{3.32}$$

Now if we decompose:

$$G = \frac{1}{e}\left(\frac{d_+}{D_+} + \frac{d_-}{D_-}\right) F_A + \frac{1}{2}\left(\frac{d_+}{D_+} - \frac{d_-}{D_-}\right) I^{AA} \tag{3.33}$$

where $d_\pm \frac{d^2 \pm d}{2}, D_\pm = \frac{D^2 \pm D}{2}$, we obtain:

$$\mathbb{E}\mathrm{Tr}[(\rho^{RE})^2] = \frac{1}{2}\frac{D^2}{d^2}\left(\frac{d_+}{D_+} + \frac{d_-}{D_-}\right)\mathrm{Tr}[(\rho^{AE})^2] + \frac{1}{2}\frac{D^2}{d^2}\left(\frac{d_+}{D_+} - \frac{d_-}{D_-}\right)\mathrm{Tr}[(\rho^{E})^2] \tag{3.34}$$

Noting that:

$$\frac{1}{2}\frac{D^2}{d^2}\left(\frac{d_+}{D_+} + \frac{d_-}{D_-}\right) \leq 1 \tag{3.35}$$

$$\frac{1}{2}\frac{D^2}{d^2}\left(\frac{d_+}{D_+} - \frac{d_-}{D_-}\right) \leq \frac{1}{d} \tag{3.36}$$

we obtain:

$$\mathbb{E}\mathrm{Tr}[(\rho^{RE})^2] \leq \mathrm{Tr}[(\rho^{AE})^2] + \frac{1}{d}\mathrm{Tr}[(\rho^{E})^2] \tag{3.37}$$

And thus we can bound the variance of Eq. (3.27) via:

$$\mathrm{Var}[\rho^{RE}] \leq \mathrm{Tr}[(\rho^{AE})^2] + \frac{1}{d}\mathrm{Tr}[(\rho^{E})^2] - \frac{1}{d}\mathbb{E}\mathrm{Tr}[(\rho^{E})^2] \leq \mathrm{Tr}[(\rho^{AE})^2] \tag{3.38}$$

And finally putting this into Eq. (3.26) we obtain:

$$\mathbb{E}\left\|\rho^{RE} - \sigma^R_{\max} \otimes \phi^E\right\|_1 \leq \sqrt{|R||E|\mathrm{Tr}[(\rho^{AE})^2]} \tag{3.39}$$

as we wanted to show. □

4 Black Holes as Information Mirrors

Having now established the background notions of black hole thermodynamics and quantum information, we are ready to attack the central question! Throughout, the setting is that you, henceforce Alice, has some diary/information that they want to destroy and conceal from an external observer Bob (who is omniscient but bound to the laws of physics). Since we established the temperature and entropic qualities of black holes to be quantum mechanical in origin, we will promote our information source to be quantum mechanical - thus, we take Alice's information to be some quantum state on k qubits, i.e. a state in a Hilbert space $\mathcal{H} = \mathbb{C}^{2^k}$ with dimension $d = 2^k$.

Bob's Goal: To Bob, a successful "learning" of the information corresponds to obtaining a copy of the quantum state. This does not mean obtaining a full classical description of the state (Alice may not even have this!). More precisely, let us quantify what a success for Bob looks like:

1. If Alice's state is pure, i.e. some $|\Psi\rangle \in \mathbb{C}^{2^k}$ then success for Bob means obtaining a copy of the state $|\Psi\rangle$.

2. As the most general case/"hardest scenario", consider the situation where Alice's state is maximally mixed, i.e. $\rho^A = I/d$. Then, we can consider its purification with an auxilary system C ("Charlie"), which is the maximally entangled state of composite system AC:

$$|AC\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i^A\rangle |i^C\rangle \tag{4.1}$$

in this context, success for Bob looks like the ability to extract from the black hole a subsystem state of dimension d that is maximally entangled with C (this is exactly what Alice started with).

If Bob has the power to accomplish case (2), he indeed has the power to extract any pure/mixed state that Alice may have started with. Since this maximally mixed case is the easiest to analyze, it will be the object of our consideration from here on out.

Bob's prior information: How do we quantify Bob's "omniscience" in this scenario? We suppose that he has been studying the black hole for a long time, by observing and collecting all the Hawking radiation that it has emitted before Alice throws her state in. Again, it would not be physically reasonable to say that Bob has a classical description of the internal state of the black hole (this could be generically very, very large) but instead, let us quantify Bob's prior knowledge as Bob possessing a maximally entangled state with the black hole; in other words, modelling the black hole as a system of $n - k \gg k$ qubits, then Bob initially has access to:

$$|EB\rangle = \frac{1}{\sqrt{n-k}} \sum_{i=1}^{n-k} |i^E\rangle |i^B\rangle. \tag{4.2}$$

where E is the part of the system held by Bob (from the Hawking radiation "E"mitted), and B is the part of the system that currently remains inside the black hole.

Let us now analyze step-by-step what happens when Alice throws her qubits into the black hole; the protocol is described diagramatically in Fig. 4.1.

1. We start with two maximally entangled systems, AC (Alice's diary A and reference system Charlie C) of Eq. (4.1) and EB (Bob's system from the emitted Hawking radiation E, and the internal system of the black hole B) of Eq. (4.2).

2. Alice throws in her qubits into the black hole, making it a $n - k + k = n$ qubit system, with subsystems maximally entangled with E, C respectively.

3. The black hole, acting as a fast scrambler, applies a unitary scrambling operation V^B on its n qubits very quickly (we revisit this notion in Sec. TODO). The scrambling operation that the black hole applies is a Haar random unitary $V^B \in U(2^n)$, as discussed in Sec. 3.4.

4. Bob continues to monitor the black hole.and over time it emits s qubits as Hawking radiation (which Bob captures), which we identify as subsystem R. $n - s$ qubits stay inside the black hole, which we identify as subsystem B'.

It is clear that for sufficiently large s (i.e. for sufficiently large amount of collected radiation R by Bob), then the correlations of C and B' become negligible - the Black hole loses information about the correlations to the radiation. In this limit, Bob's radiation RE nearly purifies the auxilary system C, which is precisely Bob's goal - to obtain a system maximally entangled with C!

The question then becomes; how large does s need to be for this information leak to occur? The size of s quantifies the speed in which the quantum information leaks out of the black hole and is learned by Bob.

Let us thus deduce the necessary size of c. The state of the entire system is described by the pure state:

$$\Psi^{BCE} = |BCE\rangle\langle BCE| \tag{4.3}$$

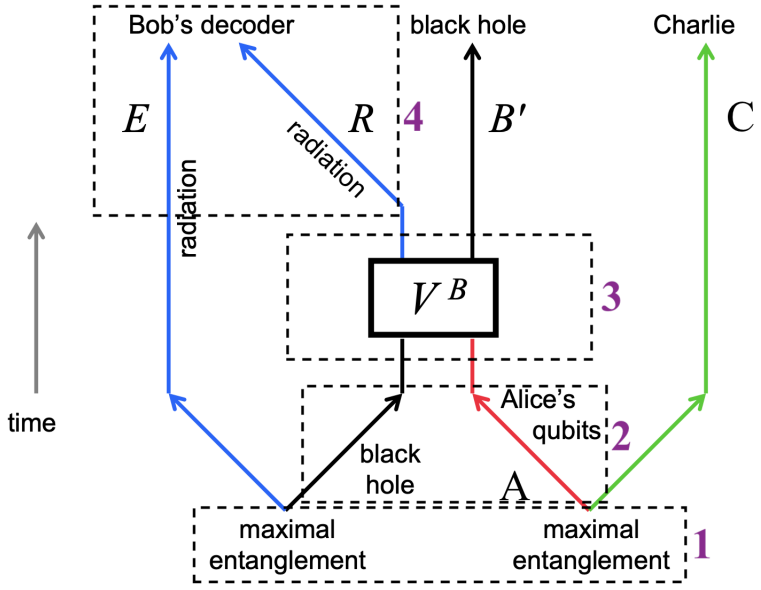

Figure 4.1: Diagram of the protocol, adapted from [1].

which is pure as it describes the entire system. The reduced density operator on the subsystem BC (the black hole and auxilary Charlie) is obtained by tracing out E:

$$\rho^{BC} = \mathrm{Tr}_E \Psi^{BCE} \tag{4.4}$$

Under the Haar random unitary V_B on the black hole subsystem, this density operator evolves as:

$$\rho^{BC} \to \rho^{BC}(V^B) = (V^B \otimes I^C)\rho^{BC}(V^{B\dagger} \otimes I^C) \tag{4.5}$$

and the reduced density operator on $B'C$ can be obtained by tracing out the radiation system R (from $B = B' \otimes R$):

$$\sigma^{B'C}(V^B) = \mathrm{Tr}_R(\rho^{BC}(V^B)) \tag{4.6}$$

and further the the reduced density operator on just C can be obtained by tracing out B':

$$\sigma^C(V^B) = \mathrm{Tr}_{B'}(\sigma^{B'C}). \tag{4.7}$$

We want to show that the state of $B'C$ is nearly maximally mixed. To this end, we consider the L^1 distance between the states $\sigma^{B'C}(V^B)$ and:

$$\sigma^{B'}_{\max} \otimes \sigma^C(V^B) \tag{4.8}$$

where:

$$\sigma^{B'} = \frac{I^{B'}}{|B'|} \tag{4.9}$$

is the maximally mixed state on B'. We consider this distance Haar averaged over the scrambling operator V^B, and find the bound, applying the one-shot decoupling theorem of section 3.5:

$$\int dV^B \left\| \sigma^{B'C}(V^B) - \sigma^{B'}_{\max} \otimes \sigma^C(V^B) \right\|_1^2 \leq \frac{|BC|}{|R|^2} \mathrm{Tr}[(\rho^{BC})^2] \tag{4.10}$$

Recall that the size $|A| = 2^k$ subsystem of B is maximally entangled with C, and the size $|E| = 2^{n-k}$ subsystem of B is maximally entangled with E. Thus, tracing out E, ρ^{BC} is maximally mixed on a subsystem of dimension $|E| = \frac{|B|}{|C|} (= 2^{n-k})$ (and pure on the $|C|$ subsystem/part corresponding to Alice's qubits). Thus:

$$\mathrm{Tr}[(\rho^{BC})^2] = \mathrm{Tr}(\left(\frac{I_{|B|}/|C|}{|B|/|C|}\right)^2) = \frac{|C|}{|B|} \tag{4.11}$$

Thus putting this into Eq. (4.10) we find the distance bound:

$$\int dV^B \left\| \sigma^{B'C}(V^B) - \sigma^{B'}_{\max} \otimes \sigma^C(V^B) \right\|_1^2 \leq \frac{|C|^2}{|R|^2} \tag{4.12}$$

and since C consists of k qubits and R of s qubits, this yields:

$$\int dV^B \left\| \sigma^{B'C}(V^B) - \sigma^{B'}_{\max} \otimes \sigma^C(V^B) \right\|_1^2 \leq \frac{2^{2k}}{2^{2s}} = 2^{-2c} \tag{4.13}$$

with $s = k + c$. In particular, the state $B'C$ is exponentially mixed in the number additional bits c after emitting the k qubits worth of Alice's quantum information. Hence, the black hole has rapidly forgotten and released Alice's quantum information as Hawking radiation to Bob. Although Alice's information is distributed amongst the Hawking radiation system ER of Bob, the argument of [10] guarantees that for a given mixing unitary V^B there exists a decoding map Bob can apply to recover a state $\rho^{\hat{A}C}$ (of which he holds the $\hat{A}$ part) to yield the following bound on the fidelity[3]:

$$F^2(V^B) = \langle AC|\rho^{\hat{A}C}|AC\rangle \geq 1 - \left\| \sigma^{B'C}(V^B) - \sigma^{B'}_{\max} \otimes \sigma^C(V^B) \right\|_1 . \tag{4.14}$$

If we integrate Eq. (4.14) over the Haar measure V^B and apply Eq. (4.13), we obtain:

$$\int V^B F^2(V^B) \geq 1 - \int dV^B \left\| \sigma^{B'C}(V^B) - \sigma^{B'}_{\max} \otimes \sigma^C(V^B) \right\|_1 \geq 1 - 2^{-c} \tag{4.15}$$

and thus Bob accomplishes his goal of extracting from the black hole a subsystem of size $d = 2^k$ that is maximally entangled with Charlie's state (of very high fidelity, with exponentially small error in c) - Alice's information has been learned!

5 Questioning Assumptions/Objections

In this section, we discuss/challenge some of the assumptions of the above argument.

5.1 How Old Is the Black Hole?

We assumed that the black hole Alice throws her qubits into is sufficiently old, such that Bob observing the prior Hawking radiation E held a maximally entangled state with the black hole interior B. This follows from an argument of the relative sizes of the given systems. If the Bekenstein-Hawking entropy (Eq. (2.22)) is $S \sim \log|B|$, at late times (after a lot of Hawking radiation has been released), at some point $\log|B| < \log|B_0|$ (with $|B_0|$ the initial size of the black hole interior Hilbert space) and we have $|B|/|E| \ll 1$. In this regime, we expect that the black hole internal state B is nearly maximally entangled with a subsystem of E (and hence our notion that "Bob knows the internal state of the black hole").

[3] To give a loose sketch of where this bound comes from, it is obtained by combining the fidelity inequality of Eq. (3.13), the contractivity property of the fidelity, and Uhlmann's theorem [7] on purifications. For this article, we wished to avoid going into detail about the machinery of encoding/decoding, but the full argument is given in [10].

However, we could instead ask what happens if Alice throws her qubits into a young black hole (that has yet to release much Hawking radiation) and so $|E|/|B| \ll 1$, wherein E is maximally entangled with some subsystem of B. In this limit, Bob has not been able to observe the Black hole for a sufficiently long time to "know" its internal state. The Hawking radiation released afterwards begins as appearing featureless, until the black hole has released enough radiation/dissipated enough to be maximally entangled with its surroundings. This occurs when $|B'| = |CER|$, and only then can we apply the arguments of the previous section. So, Alice is able to keep her information more secure for a longer period of time if she is careful about the age of the black hole she considers!

5.2 Aren't Haar Random Unitaries Inefficient?

As discussed in Section 3.4, applying a generic unitary requires an exponential number of operations - this can call into question about the physical efficiency of the scrambling operation V^B that the black hole applies. If we assume that (a) the black hole dynamics are unitary and (b) that the black hole indeed applies a Haar random scrambling operation, then we are led to conclude that the black hole dynamics must be exponentially fast (in the sense of circuit time/efficiency). Fortunately, our discussion of unitary t-designs comes to the rescue; these can indeed be realized efficiently. If we replace the integral over the Haar measure in Eq. (4.13) with a sum over ϵ-approximate unitary 2-designs $\mathcal{K}$, we instead obtain the bound:

$$\frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\left\|\sigma^{B'C}(V_k^B) - \sigma^{B'}_{\max}\sigma^C(V_k^B)\right\|_1^2 \leq 2^{-2c} + \epsilon + O(2^{-n}) \tag{5.1}$$

so up to a negligible exponential correction + error in the 2-design approximation, we obtain the same result and hence the conclusions of the argument remain unchanged. Thus, black holes can act as information mirrors when applying a more physically reasonable/efficient "random" scrambling operation.

5.3 What About the No-Cloning Theorem?

The conclusion of the argument seems to present the following tension - we can find surfaces crossed both by the quantum information that Alice throws in and by the Hawking radiation that Bob collects. Since the state that Alice throws in the black hole is in principle arbitrary, the argument then leads us to the conclusion that Bob is able to clone Alice's quantum state! However, this violates the no cloning theorem (see 3.3), a fundamental consequence of the linear nature of quantum mechanics.

We thus consider the thought experiment - is it possible for Bob to wait outside the black hole until he has reconstructed Alice's state, and then enter the black hole to confirm that he has an Alice's copy? We consider this experiment as from a physical perspective, we could postulate that we are only bothered by a violation of no-cloning if the cloning procedure could actually be verified - known as the the *Black hole complimentary hypothesis*. We will use it deduce a bound on what time the black hole needs to hold onto Alice's quantum information before re-emitting it as Hawking radiation so that the no-cloning violation cannot be verified.

For this we again consider the coordinates U, V given in our derivation of Hawing radiation:

$$U = -\left(\frac{r}{r_s} - 1\right)^{1/2} e^{r/2r_s} e^{-t/2r_s} \tag{5.2}$$

$$V = \left(\frac{r}{r_s} - 1\right)^{1/2} e^{r/2r_s} e^{t/2r_s} \tag{5.3}$$

where the black hole event horizon at $r = r_s$ is at $U = 0$ and the singularity at $r = 0$ is at $UV = 1$.

If Bob falls into the black hole at $V = V_{\text{Bob}}$, then he reaches the singularity at $U \leq V_{\text{Bob}}^{-1}$. For Alice,the proper time between $V = V_{\text{Alice}}$ crossing the horizon and reaching $U = V_{\text{Bob}}^{-1}$ to pass on the message is given by:

$$\tau_{\text{Alice}} = Cr_s(V_{\text{Alice}}/V_{Bob}) \tag{5.4}$$

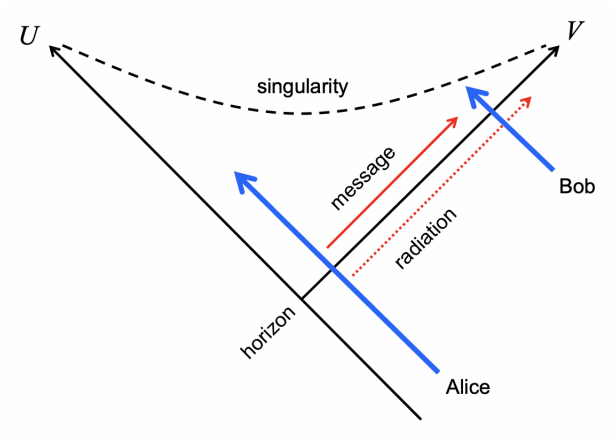

Figure 5.1: Penrose diagram of the thought experiment. Alice throws her qubits (and herself) into the black hole, which are then eventually emitted as Hawking radiation. Bob, having reconstructed Alice's quantum state, jumps into the black hole. Alice then sends Bob a message to verify that the cloning procedure succeeded. Figure taken from [1].

If Bob falls into the black hole Δt after alice, with $V_{\text{Bob}}/V_{\text{Alice}} = e^{\Delta t/2r_s}$, then:

$$\tau_{\text{Alice}} = Cr_s \exp(-\Delta t/2r_s). \tag{5.5}$$

In order for it to be impossible for Alice/Bob to verify the successful cloning, we require $\tau_{\text{Alice}} \leq t_{\text{Planck}}$ and thus require:

$$\Delta t = \Omega(r_s \log r_s) \tag{5.6}$$

and thus the no-cloning theorem gives us a lower bound on the time in which the Black hole can re-emit Alice's information!

6 Conclusions

It is worth noting that [1] goes into further discussion about assumptions of the argument that we will not discuss in full detail here, including modelling the thermalization/mixing time of the Black Hole and the hardness/complexity of Bob's decoding problem. However, through this article we were nevertheless able to give a high-level background to understand their argument and discuss some of the salient physical implications and objections. In particular, we are left with the counterintuitive result that - black holes being efficient scramblers also gives them the property of being "information mirrors" that quickly emit quantum information that was tossed in. That such a counterintuitive result arises from a unitary picture of the black hole interior also places an interesting datapoint towards the black-hole information paradox [11], and whether black holes preserve or destroy quantum information. Either way, it seems as attempting to conceal your secrets from the universe appears to be a more difficult and subtle task than you first expected!

References

[1] Patrick Hayden and John Preskill. Black holes as mirrors: quantum information in random subsystems. *Journal of High Energy Physics*, 2007(09):120–120, September 2007.

[2] Jacob D. Bekenstein. Black Holes and Entropy. *Physical Review D*, 7(8):2333–2346, April 1973.

[3] Stephen Hawking. Particle creation by black holes. *Communications in Mathematical Physics*, 43:199–220, 1975.

[4] Yasuhiro Sekino and Leonard Susskind. Fast Scramblers. *Journal of High Energy Physics*, 2008(10):065–065, October 2008. arXiv:0808.2096 [hep-th].

[5] Edward Witten. Introduction to Black Hole Thermodynamics, May 2025. arXiv:2412.16795 [hep-th].

[6] Stephen Hawking. Black holes in general relativity. *Communications in Mathematical Physics*, 25:152–166, 1972.

[7] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge ; New York, 10th anniversary ed edition, 2010.

[8] Antonio Anna Mele. Introduction to Haar Measure Tools in Quantum Information: A Beginner's Tutorial. *Quantum*, 8:1340, May 2024. arXiv:2307.08956 [quant-ph].

[9] Thomas Schuster, Jonas Haferkamp, and Hsin-Yuan Huang. Random unitaries in extremely low depth. *Science*, 389(6755):92–96, 2025.

[10] Patrick Hayden, Michal Horodecki, Andreas Winter, and Jon Yard. A decoupling approach to the quantum capacity. *Open Systems & Information Dynamics*, 15(01):7–19, March 2008. arXiv:quant-ph/0702005.

[11] Samir D. Mathur. The information paradox: A pedagogical introduction. *Classical and Quantum Gravity*, 26(22):224001, November 2009. arXiv:0909.1038 [hep-th].