Department of Physics:

Physics C191

Univeristy of California - Berkeley

# Quantum Gravity – Information Recovery From a Black Hole


**Authors:**

Shrihan Agarwal shrihan@berkeley.edu
Matthew Massaro mdmassaro@berkeley.edu
Sam Paplanus paplanus8@berkeley.edu

**Instructors:**

Dr. Ashok Ajoy ashokaj@berkeley.edu
Dr. Geoff Penington geoffp@berkeley.edu
Dr. Umesh Vazirani vazirani@eecs.berkeley.edu


**Due: May 7th, 2023**

**Abstract**

The black hole information problem is one of the most important challenges in theoretical physics, with profound implications for our understanding of the nature of space and time. The problem remains a major challenge that arises due to conflicting predictions of general relativity and quantum mechanics regarding the preservation of information in a black hole. While general relativity suggests that information is lost when it falls into a black hole, Hawking proposed that black holes may produce radiation in a mixed quantum state, leading to a loss of information. Various proposed resolutions to the information loss problem have been put forward, including the argument for information recovery made by Hayden and Preskill. While this sparked a growing consensus that information is conserved, the correct method of information recovery remains elusive, making solving it an important challenge for the theory of quantum gravity. We discuss at a high level the details of some recent proposed solutions to the information paradox, and the future outlook on this problem.

# 1 Contributions

The three authors listed: Shrihan Agarwal, Matthew Massaro, and Sam Paplanus, all contributed to the research and production of the this report. The introduction, focusing on the history of the black hole information paradox, along with the report abstract was completed by Shrihan Agarwal. The primary result, consisting of Preskill's analysis of the paradox and the conclusion of black holes acting as mirrors, as well as discussions on the qubit gate model, as well as the conclusion, was written by Sam Paplanus. Additional analysis, including a possible solution being found through VQE simulation, was completed by Matthew Massaro. Matthew was also responsible for a discussion section. Any additional pieces of this report, in addition to the previous presentation, were done jointly in equal quantities amongst the listed authors.

# 2 Introduction

Quantum gravity seeks to reconcile the deterministic predictions of general relativity and probabilistic quantum mechanics as a coherent whole. Unfortunately, this is not trivial. The pair of theories predict two opposing results when considering the information preservation in a black hole, leading to one of the most notorious contradictions in physics. This is known as the black hole information problem, a longstanding challenge to the theory of quantum gravity [10].
When matter or light falls into the event horizon of a black hole, the matter is irretrievably lost according to general relativity, as the escape velocity required is greater than the speed of light. General relativity also predicts a mass singularity at the center of the black hole, since there is no known force capable of opposing the inward collapse of matter - a point of infinite density, and infinitesimal volume. At this point, the fundamentals of general relativity, the concepts of space and time, break down [7]. According to Hawking, this places a limitation on our ability to predict the future state of a system (in addition to the Heisenberg Uncertainty Principle).
Applying quantum mechanics to this problem, we find that the "information" describing the state of the wavefunction of the infalling particles and photons must be trapped inside (or on the event horizon of) the black hole. Hawking further proposed that black holes may create particles in pairs at their event horizon surface - one which falls into the black hole, and the other which escapes to infinity. These particles constitute a mixed entangled state. The escaping particle constitutes Hawking radiation, which causes the black hole to evaporate with time. Once the black hole evaporates in its entirety leaving only thermal Hawking radiation, the initial zero

entropy (maximum information) pure quantum state has evolved into a maximum entropy (zero information) mixed quantum state. This means information is lost [4].
We briefly note here that this information loss is different from the entropy increase from thermodynamic processes, i.e. thermodynamic irreversibility. In general, although thermodynamic processes generally increase the entropy of the universe on a macroscopic scale, the principles of quantum mechanics state that these changes at a microscopic level are completely reversible. Quantum mechanics relies on the preservation of information. The wavefunction containing the information of the state of a system, can neither be created nor destroyed, and evolves unitarily. For a typical particle, this is determined by the time-dependent Schrödinger equation,

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r},t) = \hat{H}\Psi(\mathbf{r},t). \tag{1}$$

Generalizing this to quantum systems, the new state of the system can be obtained from the old state of the system by applying a general unitary transformation that can be determined from the time-dependent Hamiltonian $H(t)$:

$$\Psi(\mathbf{r},t) = U(t)\Psi(\mathbf{r},0). \tag{2}$$

To evolve the same state backwards in time, the conjugate transpose of the unitary operator $U^\dagger$ may be used. In quantum mechanics, the information of the state of the wavefunction must be preserved, such that time-evolution can be represented with unitary operators, and past and future states can be determined from the current state. If there is information loss, and pure states are converted to mixed states, this reversibility is no longer possible, and evolution is non-unitary. This sets up the crux of the black hole information problem.
There is therefore two options: either information recovery is possible in a new theory of quantum gravity, or impossible. Either case is compelling, and numerous solutions have been proposed over the years since Hawking's original work, with more recent consensus tending towards belief that information is indeed conserved [1, 13]. However, it is still not clear how Hawking's calculations should be corrected, causing this problem to remain at the forefront of quantum gravity research.
Some resolutions that have been proposed claim that Hawking radiation is not purely thermal, and instead does contain encoded information and correlations that allow us to reconstruct information from the interior of a black hole [10, 13]. Hawking accepted this view in 2005, effectively losing a long standing bet with John Preskill that information could indeed be recovered from a black hole [8]. We delve into this in detail in Section 3.
Other resolutions to the problem argue that Hawking's approach fails to account for small scale corrections [12], that the problem may be resolved by a modelling of the black hole interior geometry as a fuzzball or firewall [11], that quantum effects become important near the end of a black hole's evaporation [2], or that the information about the initial black hole state is stored in "soft particles" [9]. Other mechanisms have been proposed which modify quantum mechanics to allow possibilities of non-unitary evolution, involving a truly complete loss of information in black holes. These solutions are largely out of the scope of this paper.
In the following sections, we present the argument for recovery of information made by Hayden and Preskill (Section 3), other techniques of information recovery (Section 4), a discussion regarding the current state of research and possible future directions (Section **??**), as well as our conclusions (Section 5).

# 3 Main Results

Recovery of information that has been deposited into a black hole has been debated for decades, with some postulating that such a recovery is possible without breaking physical laws of the universe. The primary notion of recovery stems from Hayden and Preskill, who proposed recovery of information from a black hole is possible both classically and quantum mechanically in a relatively short period of time [10].

## 3.1 Black Holes as Mirrors

The idea that black holes act as mirrors, returning information deposited in a black hole quickly, is a proposed solution to the black hole information paradox developed by Hayden and Preskill in 2007 [10]. In this paper, they postulate that any information that Alice deposits into a black hole which Bob knows the internal state of will be returned to Bob after the black hole has halfway evaporated. The exact methods are of this process are described more intricately in section 4.1.1, but the general idea revolving around the recovery process and timescale for information return are discussed here.

### 3.1.1 Recovery Process

The overall recovery process presented here is an excerpt from Hayden and Preskill's paper, specifically section 3: *A quantum randomizer* [10]. Initially, imagine that there are two events occurring at the same time but at different locations which are so far apart it can be assumed they are happening infinitely far apart (i.e. there is no initial interaction between these two events).

The first event involves Alice and her friend Charlie. Let Alice possess some quantum information in the form of a $k$ qubit memory. This memory is Alice's subsystem which will be noted as subsystem $M$. Her friend Charlie is in communication with Alice and holds a system that is maximally entangled with Alice's subsystem $M$. Let Charlie's subsystem be denoted as subsystem $N$ and have the same dimension as Alice's memory such that $|N| = |M|$. Because the two subsystems have the same dimension and are maximally entangled, it can be stated that the initial state that joins the quantum information Alice possess between subsystems $M$ and $N$ is a pure state.

The second event is the evaporation of a massive black hole. Let this black hole have been formed a long time ago (i.e. long enough that it has begun to evaporate) and be set in a fixed location. Note that the fixing of this location does not eliminate the possibility of black hole rotation or other dynamics of a black hole itself, only that it is fixed in a location sufficiently far from Alice and Charlie's interaction occurring at the same time. This black hole has an internal structure/system that is denoted as subsystem $B$ that is comprised of $n$-$k$ qubits. As noted previously, this black hole formed a long time ago, meaning that it has already begun evaporating through the emission of Hawking radiation [7]. The Hawking radiation that has already been emitted comprises the radiated subsystem $E$. As this process has occurred over a long period of time, it is assumed that the radiated subsystem $E$ is maximally entangled with the black hole's internal system $B$. Let Bob, a close friend of both Alice and Charlie, be an avid observe of the black hole since its initial formation. This means that Bob has access to not only all of the emitted Hawking radiation comprising subsystem $E$, but also knows the internal state of the black hole, i.e. the $n$-$k$ qubits making up subsystem $B$. This means that

Bob possess a quantum memory that is maximally entangled with the black hole's internal state and, therefore, is perfectly correlated with the internal subsystem $B$.
Now that the initial events have occurred, let the parties meet and the interaction commence. In an attempt to get rid of the information, Alice throws the $k$ qubit memory into the black hole Bob has been observing. These qubits interact with the qubits that make up the black hole's system, yielding a total of $n$ qubits making up the black hole's internals at this point. These qubits are thoroughly mixed by the internal dynamics of the black hole, which is modeled as a deterministic unitary transformation $V^B$. This unitary transformation then uniformly randomly selects $s$ qubits such that $s > k$ to be emitted from the black hole in the form of Hawking radiation. This newly emitted Hawking radiation forms the subsystem $R$ whereas the remaining $n$-$s$ qubits inside the black hole are now referred to as subsystem $B'$.
Once subsystem $R$ is sufficiently large, Charlie's reference system $N$ is said to be purified by the emitted radiation system $RE$, which Bob is in control of due to his observations of the emitted Hawking radiation. It is at this point that Bob has sufficient information from the Hawking radiation to recover Alice's initial quantum memory of length $k$ qubits. A diagram from the Preskill paper is presented in figure 1 in which the subsystem interactions are shown. Some mathematics to support this general notion are explained in section 4.1.1.

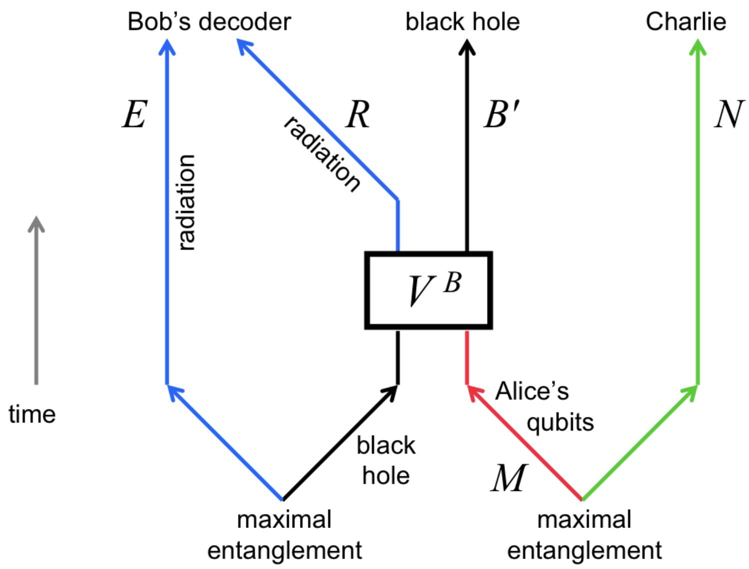


**Figure 1:** Diagram depicting the interactions, entanglement, path of the various subsystems for the recovery process. The diagram was originally presented in Preskill's paper [10].

### 3.1.2 Timescale

While knowing that Bob can recover Alice's information after she deposits her quantum information into the black hole Bob is observing, the timescale of said recovery is crucial to know if this recovery is practical or merely a thought exercise. The general conclusion from Hayden and Preskill is that Bob, with the assistance of Charlie's reference system, is able to recover Alice's quantum memory from the Hawking radiation emitted from the evaporating black hole

rapidly [10]. By rapidly, they posit that the quantum information is thermalised by the internal dynamics of the black hole on a time scale on the same order of magnitude, if not more similar, to the time between the emission of two successive radiation quanta [10].
There are two variations of timescales that can be considered in this scenario. The first occurs if Alice deposits her quantum information in the black hole after the black hole has halfway evaporated, i.e. the black hole has a total number of qubits of $\frac{n+s}{2}$ or fewer immediately following Alice placing her information in the black hole where $s = k + c$. In this situation, Bob only needs to wait for just over $k$ qubits, which we just mentioned would be $k + c = s$ qubits, to be emitted via Hawking radiation to recover the initial quantum memory [10]. However, if Alice inserts her quantum memory into Bob's black hole prior to the halfway point of evaporation, Bob needs the black hole to first reach the half-evaporated point, and then an additional $s$ qubits be emitted in the form of Hawking radiation to recover Alice's information. In either scenario, this timescale is small relative to the evaporation time of the black hole, which led Hayden and Preskill to come to the conclusion that black holes do not act as information vacuums, but are instead mirrors that quickly reflect information they take in back out.

### 3.2 Significance of Recovery

The notion that information can be recovered from a black hole is a critical breakthrough in the world of black holes and quantum information. One such significant impact is helping to show that black holes do not simply exist taking in celestial information never to be seen again, essentially erasing information from the cosmos. Instead, this helps solidify the notion of information unitary and provides insight into possible models for what occurs within a black hole, including the potential physics of the singularity. Information recovery could lead to further work on the dynamics of black holes and further verify the laws of physics that were proposed to be broken within the Schwarzschild radius, specifically at the singularity, of a black hole.

## 4 Techniques

Now that the general idea of black hole information recovery has been presented, the more minute details that help make the previous notion physically possible are discussed. Additional methods of black hole recovery are also presented.

### 4.1 Black Hole Dynamics

The primary idea behind recovery of information deposited into a black hole is to model the internal dynamics of the black hole as some form of unitary operation. This is the primary driver behind Hayden and Preskill's notion in 2007, but does not address the entirety of the problem. If it is assumed that the final state of the information after it has been emitted is a mixed state rather than pure, the internal dynamics of the black hole must possess some form of non-unitary transformation to make the change from pure to mixed state physically possible.

#### 4.1.1 Preskill's Quantum Model

The quantum model described in Hayden and Preskill's paper requires some intricate quantum details to be discussed to fully understand. For instance, the idea that Charlie's reference system

$N$ is maximally entangled with Alice's quantum memory $M$. This means that the joint state of these systems can be any pure state of the form of equation 3 [10].

$$|\Phi\rangle^{MN} = \frac{1}{\sqrt{|M|}} \sum_{a=1}^{|M|} |a\rangle^M \otimes |a\rangle^N \tag{3}$$

This is the initial pure state of the quantum information that sets the stage for unitarity of the system. In this case, $N$ provides a purification for state $M$. If Charlie and Alice hold onto the systems $N$ and $M$ respectively, $\rho_N$, which is the density operator for Charlie's system $N$ after $M$ has been traced out, is maximally mixed. Since it was determine that the initial state of Alice's memory was a pure state, which is noted here as $|\psi\rangle$, then Bob, upon extracting the information from the Hawking radiation, would have recovered $|\psi\rangle$ in a subsystem of his choice [10]. This procedure ensures unitarity, since a pure state which underwent a unitary transformation within the black hole, is recovered as a pure state, therefore not violating the rules of quantum mechanics.

For recovery to occur, it is necessary that the black hole's internal system $B$ is maximally entangled with the external system $NE$, which is possible once the radiation emitted after Alice tosses her information in $R$, is sufficiently large [10]. By sufficiently large, it is meant that there is no correlation between the reference frame $N$ and the remaining black hole's internal system $B\,'$ is essentially nonexistent such that the overall state $B\,'RNE$ is pure [10]. For a unitary transformation $V^B$, which is what was used to describe the internal dynamics of the black hole, the system $NB\,'$ is maximally mixed after $s = k + c$ qubits escaped according to equation 4 [10].

$$\int dV^B ||\sigma^{NB'}(V^B) - \sigma^N(V^B) \otimes \sigma^{B'}_{max}|| \frac{|N|^2}{|R|^2} = \frac{2^{2k}}{2^2 s} = 2^{-2c} \tag{4}$$

In equation 4, $\sigma$ represents the marginal density operator on a given subsystem described in its superscript. The mathematics behind this integral are out of scope of this paper and course, but more details can be illuminated by review Hayden and Preskill's paper [10]. The important outcome of this integration is the time at which $NB\,'$ becomes maximally entangled, which leads to Bob recovering these $s = k + c$ qubits while the black hole, for all intensive purposes, has forgotten that information.

This conclusion can be stated in other terms more directly correlated to Physics C191, in the form of noisy quantum channels. The following description of this method is presented in Preskill's paper at the end of section 3: *A quantum randomizer* [10]. For the case of the black hole being past half evaporated upon Alice's information being mixed, there exists a quantum erasure channel with probability $p$, then the entanglement-assisted quantum capacity $Q_E = 1 - p$. The black hole retains $n$ - $k$ - $c$ qubits after Bob is able to recover Alice's quantum memory, which are erased from existence as far as Bob is concerned. Despite losing $n$ - $k$ - $c$ qubits of information, Bob has been able to extract $k$ qubits of quantum information of high fidelity. The rate of communication is thus $R = \frac{k}{n}$, which is achieved by a random unitary encoder. If, however, the black hole was not half evaporated when Alice's memory was thrown in, the probabilities are slightly altered. If $p \geq \frac{1}{2}$, then $Q_E = 0$, but if $p < \frac{1}{2}$, $Q_E = 1 - 2p$. Once $\frac{n+k}{2}$ have escaped from the black hole, $\frac{n-k}{2}$ are still trapped within. But Bob now has gathered $k$ qubits of quantum information from the black hole, leading to a communication rate identical to the previous scenario with a probability of erasure of $p = \frac{1}{2} - \frac{R}{2}$.

The capacity of these erasure channels, as with all quantum erasure channels, is achieved by using stabilizer codes. Thus, if the internal dynamics of the black hole are modelled as a connected system of $n$ two qubit gates, resulting in a unitary transformation, then a stabilizer code of with size on the order of $n^2$ could be found to aid in the decoding of information. This makes Bob's recovery of Alice's information significantly easier: replace an erased qubit with a null state $|0\rangle$, and then measure the stabilizer code operators to determine if that error has occured. With high probability, there is a unique set of Pauli operations that act on the erased qubits that will allow Bob to restore his state to the original code space [10]. The model of describing the internal state of a black hole as a set of two qubit gates with decoding being achievable through the use of a stabilizer code of some form was also proposed by Halyo [6].

## 4.2 Simulating Hawking Radiation

As stated before, the premise of combining the classical ideas of gravity and the quantum realm have been a hurdle for scientists for decades. In Dhaulakhandi's paper [5], a simulation of the ground state wave equation of Hawking radiation emitted from black holes first uses a method to form the Hamiltonian for the Schwarzschild metric of spacetime, and then quantizes it in a Pauli basis.

$$g = -c^2 dr^3 = -(1 - \frac{r_s}{r})c^2 dt^2 + (1 - \frac{r_s}{r})^{-1} dr^2 + r^2(d\phi^2) \tag{5}$$

Next, using a VQE (Variational Quantum Eigensolver), an upper bound on the energy of the system is found based on its ground state. By inserting an ansatz into the algorithm and selecting key parameters the time evolution of two qubits of Hawking radiation are simulated utilizing unitary operators and CNOT gates.
Three ansatz are run through the VQE with varying amounts of operators and gates, each producing significantly different results. Then each run will simulate increasing distance from the surface of the black hole.

(a) Ansatz 1 (b) Ansatz 2 (c) Ansatz 3

**Figure 2:** Diagram of three ansatz used for VQE simulation [5].

The first has unitary operators act on each qubit individually with CNOT gates spaced between and after each operation. This results in large gaps between the expected energy values but maintain proportionality. Unlike energy, the temperature does not follow expected results and does not conform with the expected inverse relationship with the mass of the black hole. The second ansatz applies one unitary operator on the first qubit and controlled unitary operators on the second and then first qubit again. Similar to the first, there are large, but proportional gaps in the energy eigenvalues, but temperature still does not conform. It is only in the third simulation of a single qubit undergoing a unitary operation that matches the expected interpretations of both energy eigenstates and temperature profiles. Based on the assumptions of this paper, the result of this implies that in order to maintain proper information on eigenstates utilizing unitary operators, you cannot have entangled states between qubits of a system within the black hole and the Hawking radiation produced.

## 5 Conclusion

The idea of retrieving information from a black hole has long been thought to be impossible due to the properties that a black hole possesses. However, after Hawking proved that radiation can escape a black hole, the idea was revived. Using quantum models to simulate and explain the internal dynamics of a black hole, it has been theorised that information not only can be recovered after it has been deposited in a black hole, but can be done so relatively quickly. Preskill modeled the internal dynamics of a black hole as a unitary operator and showed that this, along with mixing between states and use of quantum principles, any information deposited within a black hole can be retrieved shortly after is was originally deposited. It was also stated that the internal dynamics of a black hole could be modeled as a series of two qubit gates, which Halyo also described, which can lead to recovery through the use of stabiliser codes.
Further research into the recovery of information deposited in a black hole is necessary to further verify these claims, along with potentially explain the actual internal dynamics of the mysterious celestial objects. Is information that falls into a black hole truly lost forever, or can it be recovered using simple models? Only further simulations, theories, testing, and experimentation can answer this question definitively.

## 6 Acknowledgements

We'd like to thank the Professors Ajoy, Vazirani and Penington for their comments and questions during the project and presentation. We'd also like to thank the GSIs for their support in office hours. We'd like to give a special thanks to Saul, Nicolas and Elna for interesting discussions regarding quantum gravity and clarifying the scope of the project. Lastly, we'd each like to thank our group partners for an excellent job on the project.

## References

[1] Ahmed Almheiri, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini. The entropy of hawking radiation. *Rev. Mod. Phys.*, 93:035002, Jul 2021.

[2] Abhay Ashtekar. Black Hole Evaporation: A Perspective from Loop Quantum Gravity. *Universe*, 6(2):21, January 2020.

[3] Erik Aurell, Michał Eckstein, and Paweł Horodecki. Hawking radiation and the quantum marginal problem. *jcap*, 2022(1):014, January 2022.

[4] Stephen Boughn. A Modest View of the Black Hole Information Paradox. *arXiv e-prints*, page arXiv:2207.09421, July 2022.

[5] Ritu Dhaulakhandi and Bikash K. Behera. Quantum Simulation of Hawking Radiation Using VQE Algorithm on IBM Quantum Computer. 12 2021.

[6] Edi Halyo. A model for unitary black hole evaporation, 2020.

[7] S. W. Hawking. Breakdown of predictability in gravitational collapse. *prd*, 14(10):2460–2473, November 1976.

[8] S. W. Hawking. Information loss in black holes. *prd*, 72(8):084013, October 2005.

[9] Stephen W. Hawking, Malcolm J. Perry, and Andrew Strominger. Soft hair on black holes. *Phys. Rev. Lett.*, 116:231301, Jun 2016.

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

[11] Samir D. Mathur. The information paradox: a pedagogical introduction. *Classical and Quantum Gravity*, 26(22):224001, November 2009.

[12] J. Preskill. Do Black Holes Destroy Information? In Sunny Kalara and D. V. Nanopoulos, editors, *Black Holes, Membranes, Wormholes and Superstrings*, page 22, January 1993.

[13] Suvrat Raju. Lessons from the information paradox. *Physics Reports*, 943:1–80, 2022. Lessons from the information paradox.