DESY Workshop Seminar 2024: The Information Paradox

JML

January 2024

Abstract

In this lecture, we review the discussion in [1]. The goal is to make precise the arguments made by S. Hawking that simple assumptions lead one to a violation of unitarity or the existence of remnants during the evaporation of a black hole.

Contents

1 Statement

Let us review a bit of the discussion in [2]

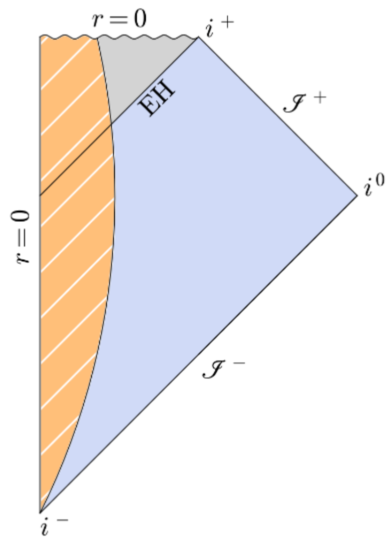

Figure 1: Realistic black hole Penrose diagram. Stolen from [2].

We consider a spherical shell of matter that collapses into a black hole. The Penrose diagram of this process is depicted in Fig. 1. As discussed in the previous lecture, the initial state is some pure matter state. However, one finds the outgoing radiation has the total number of particles in a given mode

$$n_\omega \simeq \frac{1}{e^{2\pi\omega/\kappa} \pm 1} \tag{1}$$

This is a thermal distribution with temperature $T = \frac{\kappa}{2\pi} = \frac{1}{8\pi GM}$ that arises from a system being in a mixed state. Why is this a problem? Let us review some basic quantum mechanics.

1.1 States of Quantum Systems

Let us consider a quantum system with Hilbert space $\mathcal{H}$ and an operator $\mathcal{A}$ with a complete set of eigenkets $\{|n\rangle\}$ such that

$$\mathcal{A}\,|n\rangle = a_n\,|n\rangle \tag{2}$$

and a state $|\Psi\rangle \in \mathcal{H}$ can be expanded as

$$|\Psi\rangle = \sum_n c_n \tag{3}$$

such that $\sum_n |c_n|^2 = 1$ (giving the usual $\langle\Psi|\Psi\rangle = 1$).

We now want to classify some states of quantum systems. We take the approach that a system is defined by a **density matrix** ρ such that the expectation value of an operator $\mathcal{A}$ is given by

$$\begin{aligned}\langle\mathcal{A}\rangle :&= \mathrm{Tr}(\rho\mathcal{A}) \\ &= \sum_n \langle n|\,\rho\mathcal{A}\,|n\rangle\end{aligned} \tag{4}$$

For a non-relativistic system, the time evolution is governed by the von Neumann equation.

Definition 1.1 (von Neumann Equation)**.** *For a quantum system governed by the Hamiltonian operator H, the density matrix satisfies the **von Neumann equation***

$$i\hbar\frac{\partial}{\partial t}\rho = [H,\rho] \tag{5}$$

As usual, we can define a **time evolution operator** $\mathcal{U}(t,t_0)$ as

$$\mathcal{U}(t,t_0) := \exp\left[-\frac{i}{\hbar}H(t-t_0)\right] \tag{6}$$

so that the density matrix of the system at time t, $\rho(t)$, can be built from the density matrix at time t_0,$\rho(t_0)$, as

$$\rho(t) = \mathcal{U}(t,t_0)\rho(t_0)\mathcal{U}^\dagger(t,t_0) \tag{7}$$

This is also a solution to Eq. (5).

With this technology, we can differentiate between two types of states for our system

Definition 1.2 (Pure States)**.** *A system is in a **pure state** if its density matrix can be written as*

$$\rho = |\Psi\rangle\langle\Psi| \tag{8}$$

for some $|\Psi\rangle \in \mathcal{H}$. The density matrix of pure states satisfies several properties, including

$$\begin{aligned}\rho^2 &= \rho \\ Tr(\rho) &= Tr(\rho^2) = 1\end{aligned} \tag{9}$$

Pure states are the usual states we study in introductory quantum mechanics. However, a system can also be a a mixed state.

Definition 1.3 (Mixed States)**.** *A system is in a **mixed state** if its density matrix can be written in terms of pure state density matrices as*

$$\rho = \sum_i p_i\,|\Psi_i\rangle\langle\Psi_i| \tag{10}$$

where $|\Psi_i\rangle \in \mathcal{H}$. This is a ***convex sum****, so $p_i > 0$ and $\sum_i p_i = 1$. Note that pure states correspond to the special case of only one non-zero coefficient p_i. Apart from this degenerate case, mixed states satisfy*

$$Tr(\rho^2) = \sum_i p_i^2 < \sum_i p_i = 1 \tag{11}$$

How do mixed states come about? Consider a system with two subsystems (which we call A and B) and let the state of the whole system be ρ_{sys}. Then we can define the state of the B subsystem by **tracing out** the A subsystem:

$$\rho_B := \mathrm{Tr}_A(\rho_{sys}) \tag{12}$$

The trace can be carried out by summing over a complete set of states spanning $\mathcal{H}_A$, the Hilbert space of the subsystem A.

A simple example illustrating the above arises from the entanglement of spins – let A and B each consist of a single spin-$\frac{1}{2}$ particle. We can consider the entangled state $\rho_{sys} = |\Psi\rangle \langle\Psi|$, where[1]

$$|\Psi\rangle = \frac{1}{\sqrt{2}} |0\rangle_A |0\rangle_B + \frac{1}{\sqrt{2}} |1\rangle_A |1\rangle_B \tag{13}$$

Note that the total system is in a pure state. However, we find

$$\rho_B = \frac{1}{2} |0\rangle \langle 0|_B + \frac{1}{2} |1\rangle \langle 1|_B \tag{14}$$

Thus if we only have access to the B subsystem, we would find that our system is in a mixed state.

For our purposes, an essential feature of the above is the conservation of the trace of the density matrix. Using Eq. (7), one finds that

$$\mathrm{Tr}(\rho^2(t)) = \mathrm{Tr}(\rho^2(t_0)) \tag{15}$$

which implies

Remark 1. *In (unitary) quantum mechanics, it is impossible for a pure state to evolve into a mixed state.*

This is the root of the information paradox.

For later use, we also define the **entanglement entropy** of a quantum system.

Definition 1.4 (Entanglement Entropy)**.** *For a system in a state defined by ρ, the entanglement entropy is*

$$S = Tr(\rho \ln(\rho)) \tag{16}$$

[1]We use the shorthand $|\cdot\rangle_A |\cdot\rangle_B := |\cdot\rangle_A \otimes |\cdot\rangle_B$, where the dots are 0 (1) for spin down (up).

2 Assumptions

One could try to poke many holes in the previous section to "resolve" the information paradox. After all, the discussion was based in non-relativistic quantum mechanics. However, relativistic quantum mechanics/quantum field theory (QFT) also enjoy unitary evolution, so the general principle outlawing pure-to-mixed evolution holds true.

One can then go on to claim that there is some funny business happening with QFT in curved spacetimes. This motivates the careful discussion in [1] to clearly define the assumptions and physics going into Hawking's calculation. It is to this task we now turn.

2.1 The Power of Niceness

First, we somewhat formalize the notion of "usual physics". This is described as the **Solar System Limit** (SSL)

Definition 2.1 (Solar System Limit)**.** *The SSL is an assumption about configurations in quantum gravity and is defined by the set* $\{\mathcal{N}_i, \epsilon_i\}$*. This set consists of* ***niceness conditions*** $\{\mathcal{N}_i\}$ *and small parameters* $\{\epsilon_i\}$*. The assumption states that if a configuration i). satisfies the conditions* $\{\mathcal{N}_i\}$ *and ii). has the* $\{\epsilon_i\}$ *arbitrarily small, then physical processes can be described to arbitrarily high accuracy by known, local, evolution equations.*

Another way to state the SSL is that when it holds, we can specify a quantum state on an initial spacelike slice $\mathcal{S}$ and a Hamiltonian operator gives the state on later slices.

What are these niceness conditions? We can adopt the list from [1]. Let us consider a state $|\Psi\rangle$ on a spacelike slice $\mathcal{S}$. Then a tentative set of niceness conditions $\{\mathcal{N}_i\}$ is

- $\mathcal{N}_1$: ${}^{(3)}R_{\mathcal{S}} \ll \ell_p^{-2}$
- $\mathcal{N}_2$[2]: $K_{\mathcal{S}} \ll \ell_p^{-2}$
- $\mathcal{N}_3$: ${}^{(4)}R \ll \ell_p^{-2}$ in the neighborhood of $\mathcal{S}$
- $\mathcal{N}_4$: All matter on $\mathcal{S}$ is "good" in the sense that

$$\begin{aligned} \lambda &\gg \ell_p \\ U, P &\ll \ell_p^{-4} \end{aligned} \tag{17}$$

  where λ is the wavelength of any matter particle and U and P are the energy & momentum densities, respectively. [1] also demands all matter satisfy the usual energy conditions.

[2] This arises from demanding that the slice sit nicely in 4D spacetime.

Figure 2: Pair creation after evolving from one time slice to the next. Stolen from [1].

- $\mathcal{N}_5$: In evolving the state from the initial slice $\mathcal{S}$, the matter state should be "good" on subsequent slices. We also demand that the lapse & shift vectors used to specify the evolution should vary smoothly with position.

Here ${}^{(3)}R_{\mathcal{S}}$ and ${}^{(3)}R_{\mathcal{S}}$ are the intrinsic and extrinsic curvature of $\mathcal{S}$, respectively. ${}^{(4)}R$ is the curvature of the full spacetime. Note that the author specifies that this list is not necessarily minimal.

2.2 Locality

We would also like to impose a notion of **locality**. An intuitive description is that if we have a state on a spacelike slice $\mathcal{S}$, then the influence of one region on another vanishes as the distance between the regions becomes infinitely large.

For our purposes, we will have to be a bit more precise. Our goal is to examine pair creation at the event horizon of a black hole. With this in mind, we consider a scenario like that depicted below The continuous lines are two slices of spacetime. The square dots denote matter in the state $|\psi\rangle_M$ and the circles indicate pair creation due to the deformation of the slice. We assume that the matter and created particles are extremely far apart.

Naively, locality dictates that the state on the upper slice should have the form

$$|\Psi\rangle_1 \simeq |\psi\rangle_M \otimes \frac{1}{\sqrt{2}}(|0\rangle_c |0\rangle_b + |1\rangle_c |1\rangle_b) \tag{18}$$

However, this is too strict - locality allows 'small' departures from this state. What do we mean by small? If we let the matter state be of the form $|\psi\rangle_M = \frac{1}{\sqrt{2}} |0\rangle_M + \frac{1}{\sqrt{2}} |1\rangle_M$, then we could have

$$|\Psi\rangle_2 \simeq \left(\frac{1}{\sqrt{2}} |0\rangle_M + \frac{1}{\sqrt{2}} |1\rangle_M\right) \otimes \left((\frac{1}{\sqrt{2}} + \epsilon) |0\rangle_c |0\rangle_b + (\frac{1}{\sqrt{2}} - \epsilon) |1\rangle_c |1\rangle_b\right) \tag{19}$$

with $\epsilon \ll 1$ but obviously not

$$|\Psi\rangle_3 \simeq \left(\frac{1}{\sqrt{2}} |1\rangle_M |0\rangle_c + \frac{1}{\sqrt{2}} |0\rangle_M |1\rangle_C\right) \otimes \left(\frac{1}{\sqrt{2}} |0\rangle_b + \frac{1}{\sqrt{2}} |1\rangle_b\right) \tag{20}$$

How can we quantify the closeness of Eq. (19) to Eq. (18)? We can turn to our earlier definition of the entanglement entropy. We can define the density matrix of the B subsystem for each of the three states above by $\rho_B = \mathrm{Tr}_{MC}(|\Psi\rangle \langle\Psi|)$. Then the entanglement entropies of this subsystem for the states above are

$$\begin{aligned} S_1 &= \ln(2) \\ S_2 &= \ln(2) - \epsilon^2(6 - 2\ln(2)) \approx \ln(2) \\ S_3 &= 0 \end{aligned} \tag{21}$$

This motivates [1] to define locality for pair creation as follows

Definition 2.2 (Locality)**.** *For a pair-creation process occurring in a configuration satisfying the SSL, as the matter state goes further and further from the pair-creation location, the entanglement entropy of either of the pair-created particle satisfies*

$$\frac{S}{\ln(2)} - 1 \ll 1 \tag{22}$$

2.3 Existence of traditional black holes

In addition to the above, we also assume the existence of "traditional" black holes. This is essentially the statement that there is no 'information' about the black hole in the vicinity of the horizon. We encapsulate this notion with a definition of a **information-free horizon**

Definition 2.3 (Information-Free Horizon)**.** *A point on the event horizon of a black hole is information-free if a neighborhood around this point is the 'vacuum' in the sense that the evolution of modes with wavelengths $\ell_p \ll \lambda < r_s/2$ is given by the semiclassical evolution of quantum fields on empty space up to $\mathcal{O}(m_p/M)$ corrections.*

And we also define

Definition 2.4 (Traditional Black Hole)**.** *A black hole with an information-free horizon is a traditional black hole.*

Note that in any curved space there is no unique definition of particles, but for particles of wavelength $\lambda \lesssim R$, with R the curvature radius, we can have a definition for "empty space". This is to say that one definition of particle will vary with another by $\mathcal{O}(1)$ quanta with $\lambda \sim R$. In the case of a black hole, on would expect $\mathcal{O}(1)$ quanta with $\lambda \sim M$ for any definition of particles – Hawking radiation quanta also fall in this category. "Empty space" in this case means that modes with $\lambda < M$ are not populated with $\mathcal{O}(1)$ quanta each.

3 The Hawking "Theorem"

Roughly, the statement we now turn it is to show that for a configuration with a black hole, the SSL and locality gives an unacceptable evolution (pure→mixed). One must either i). accept the unacceptable or ii). alter the coniditions of the SSL

3.1 The Slicing

As described in the sections above, the essence of the information paradox is the apparent evolution from a pure state to a mixed state. However, this argument simply looked at the initial and final states. To understand why the information paradox is so difficult to resolve, we must look at it step-by-step.

First, we must slice up the black hole spacetime in such a way as to satisfy the SSL. If we choose a spacelike slice that intersects with the spacelike singularity of the black hole, then the SSL would not hold everywhere on the slice. Then the task is to construct slices that satisfy the conditions in the SSL but capture the physics of Hawking radiation.

This is done as follows:

- Outside the horizon, $r > r_s$, we define the slice by t =constant
- Inside the horizon, $r < r_s$, the spacelike slices are r =constant. We pick this constant such that is lies between $r_s/4$ and $3r_S/4$ so that it is far from the horizon and singularity at $r = 0$.
- The above two segments are joined by a smooth connector piece $\mathcal{C}$.

Note that we are interested in physical black holes that were formed by collapse, whose Penrose diagram is displayed in Fig. 1. For such black holes, one can continue the interior $r < r_s$ part of the slices to early times and smooth extend all the way to $r = 0$, before the black hole formed. The slicing is depicted below.

Adjacent slices $\mathcal{S}_n$ and $\mathcal{S}_{n+1}$ have $t_{n+1} = t_n + \Delta$ and $r_{n+1} = r_n + \delta$, where $\delta \ll M$. The primary difference in adjacent slices is in the connector pieces $\mathcal{C}$. In particular, in going from $\mathcal{S}_n$ to $\mathcal{S}_{n+1}$, the connector piece stretches. This is where particle creation takes place.

With our slicing outlined, we move onto the evaporation process in detail.

3.2 The (Leading Order) Process

We now outline the Hawking evaporation process step-by-step utilizing our slicing of spacetime.

- The initial slice contains the matter state $|\psi\rangle_M$ that collapses into the black hole.
- After collapse, we evolve to the next slice. Pair creation happens in the connector segment at the horizon, as shown in Fig. 4. The state is now

$$|\Psi\rangle \approx |\psi\rangle_m \otimes |\phi_1\rangle \tag{23}$$

where we have defined an entangled state

$$|\phi_n\rangle = \frac{1}{\sqrt{2}} |0\rangle_{c_n} |0\rangle_{b_n} + \frac{1}{\sqrt{2}} |1\rangle_{c_n} |1\rangle_{b_n} \tag{24}$$

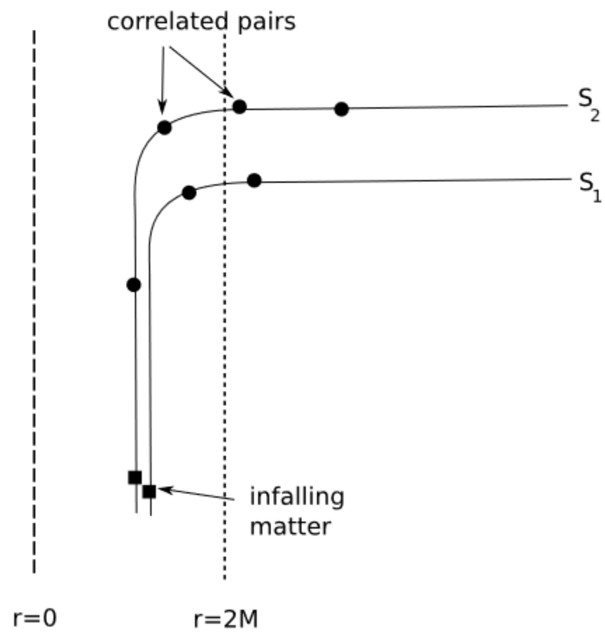

Figure 3: Slicing the black hole spacetime. Stolen from [1].

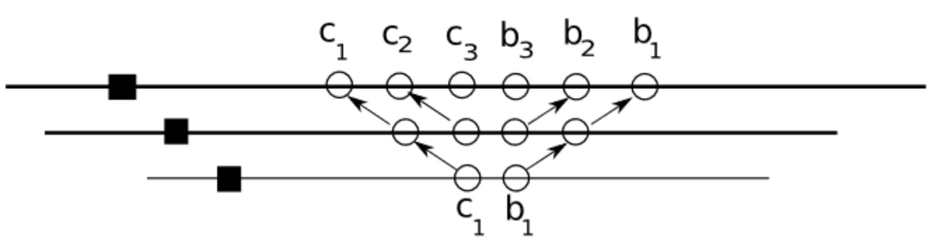

Figure 4: Artistic depiction of Hawking evaporation process, circa 2011. Stolen from [1].

The b_1 particle escapes to infinity and can be captured by an observer. The entanglement entropy of this system with the (M, c_1) system is

$$S_1 = \ln(2) \tag{25}$$

- Evolving to the next time slice, the state becomes

$$|\Psi\rangle \approx |\psi\rangle_M \otimes |\phi_1\rangle \otimes |\phi_2\rangle \tag{26}$$

  The $\{b_1, b_2\}$ quanta have entanglement entropy

$$S_2 = 2\ln(2) \tag{27}$$

  with the $(M, \{c_1, c_2\})$ system

- Repeating these steps a total of N times, we see that the resulting $\{b_i\}$ are in a state with entanglement entropy

$$S_N = N\ln(2) \tag{28}$$

  with the interior $(M, \{c_i\})$ particles

- As the black hole evaporates, its mass decreases. Eventually $M_{BH} \simeq m_{pl}$ and the conditions of the SSL are violated. We then stop the evolution.

When we stop the evolution, we have a collection of radiation quanta that are entangled with the interior state of the black hole. From here, there are two possible outcomes

- First, despite the SSL conditions being violated, the black hole could evaporate entirely away. This leaves the $\{b_i\}$ quanta in a mixed state that is not entangled with anything because the black hole has ceased to exist. This would represent non-unitary evolution

- The second possibility is that the evaporation has ceased and what remains is a remnant that is entangled with the radiation quanta. This would be a **remnant**.

Both of these possibilities are unsavory. The first for obvious reasons. As for the later, we note that the remnant must have at least N states to purify the Hawking radiation. This makes the remnant a compact object of finite size and energy but unbounded degeneracy. This would be a strange thing indeed.

For the above, one could argue that the Hawking process leads one to either pure state $\to$ mixed state evolution or the existence of remnants.

3.3 Small Corrections

As discussed above, in evolving from one slice to the next, the state evolves as

$$|\Psi_{M,c}, \psi_b; t_n\rangle \rightarrow |\Psi_{M,c}, \psi_b; t_n\rangle \otimes |\phi_{n+1}\rangle \tag{29}$$

At least, this is true to leading order. One may expect that interactions or non-perturbative effects alter this state. Thus we now consider deformations of the leading order state to see if they can resolve the paradox.

First, we want to define what we mean by small corrections. At the n-th step of the evolution process, we can choose a basis of orthonormal states $|\xi_n\rangle$ for the $(M, \{c_i\})$ quanta inside the black hole and an orthonormal basis $|\chi_m\rangle$ for the exterior $\{b_i\}$ quanta such that

$$\begin{aligned} |\Psi_{M,c}, \psi_b; t_n\rangle &= \sum_{m,n} C_{mn} |\xi_m\rangle |\chi_n\rangle \\ &\rightarrow \sum_i C_i |\xi_i\rangle |\chi_i\rangle \end{aligned} \tag{30}$$

Where we performed a unitary transformation on the states in going from the first to the second line. The density matrix of the exterior quanta is

$$\rho_B = |C_i|^2 \delta_{ij} \tag{31}$$

and the entanglement entropy is

$$S_B(t_n) = \sum_i |C_i|^2 \ln |C_i|^2 \tag{32}$$

We now proceed to the $(n+1)$-th step. A $\{c_{n+1}, b_{n+1}\}$ particle pair is created and we assume the space of the new particles is spanned by the states

$$\begin{aligned} |\gamma_1\rangle &= \frac{1}{\sqrt{2}} \Big(|0\rangle_{c_{n+1}} |0\rangle_{b_{n+1}} + |1\rangle_{c_{n+1}} |1\rangle_{b_{n+1}} \Big) \\ |\gamma_2\rangle &= \frac{1}{\sqrt{2}} \Big(|0\rangle_{c_{n+1}} |0\rangle_{b_{n+1}} - |1\rangle_{c_{n+1}} |1\rangle_{b_{n+1}} \Big) \end{aligned} \tag{33}$$

Rather than forcing the $(n+1)$-th state to have the form in Eq. (29), we relax the evolution somewhat. We **will** assume that the $\{b_i\}$ quanta already outside the black hole are unaffected by the creation of the new pair. This is reasonable since these quanta have completely escaped the vicinity of the black hole and cannot be affected except via some extremely non-local process. Then the most general evolution of the state from t_n to t_{n+1} is

$$\begin{aligned} |\chi_i\rangle &\rightarrow |\chi_i\rangle \\ |\xi_i\rangle &\rightarrow \left|\xi_i^{(1)}\right\rangle |\gamma_1\rangle + \left|\xi_i^{(2)}\right\rangle |\gamma_2\rangle \end{aligned} \tag{34}$$

where the $\{|\xi_i^a\rangle\}$ are states of the $(M, \{c_i\})$. By unitarity, $|\xi_i^{(1)}| + |\xi_i^{(2)}| = 1$. At leading order, we had

$$\begin{aligned}\left|\xi_i^{(1)}\right\rangle &= |\xi_i\rangle \\ \left|\xi_i^{(2)}\right\rangle &= 0\end{aligned} \tag{35}$$

Using all the above, we have the state at t_{n+1} as

$$\begin{aligned}|\Psi_{M,c}, \psi_b; t_{n+1}\rangle &= \sum_i C_i \bigg(\left|\xi_i^{(1)}\right\rangle |\gamma_1\rangle + \left|\xi_i^{(2)}\right\rangle |\gamma_2\rangle \bigg) \otimes |\chi_i\rangle \\ &= |\gamma_1\rangle \left|\Lambda^{(1)}\right\rangle + |\gamma_2\rangle \left|\Lambda^{(2)}\right\rangle\end{aligned} \tag{36}$$

where

$$\begin{aligned}\left|\Lambda^{(1)}\right\rangle &= \sum_i C_i \,|\chi_i\rangle \left|\xi_i^{(1)}\right\rangle \\ \left|\Lambda^{(2)}\right\rangle &= \sum_i C_i \,|\chi_i\rangle \left|\xi_i^{(2)}\right\rangle\end{aligned} \tag{37}$$

Now since the $|\gamma_i\rangle$ are orthonormal, normalization of the total state implies that

$$|\Lambda^{(1)}|^2 + |\Lambda^{(2)}|^2 = 1 \tag{38}$$

We can now define precisely what we mean by small corrections.

Definition 3.1 (Small Corrections)**.** *We call corrections to the leading-order Hawking process "small" if*

$$|\Lambda^{(2)}|^2 < \epsilon \ll 1 \tag{39}$$

If there is no such bound, then we say that the corrections are order unity.

3.4 Entropy Changes

We now show that at the step t_{n+1}, the entropy of the b quanta with the interior states **increases** if the corrections are small. First, let S_0 denote the entanglement entropy of the $\{b_i\}$ with the interior quanta at the time step t_n. That is,

$$S(\{b\}; t_n) := S_0 \tag{40}$$

Note that by the assumption that the $\{b_i\}$ are unaffected by the creation of the new quanta pair $p = (c_{n+1}, b_{n+1})$, we must have

$$S(\{b\}; t_{n+1}) = S_0 \tag{41}$$

We wish to argue that the entanglement entropy of the $\{b_i, b_{n+1}\}$ with interior quanta satisfies

$$S(\{b_i\}, b_{n+1}) > S_0 - 2\epsilon \tag{42}$$

Thus, despite small corrections, the entanglement entropy increases in the step from t_n to t_{n+1}.

This requires several lemmas

Lemma 1. *If Eq.* (39) *holds, then the entanglement of the pair* $p = (c_{n+1}, b_{n+1})$ *with the rest of the system is bounded as*

$$S_p := S(p) = Tr(\rho_p \ln(\rho_p)) < \epsilon \tag{43}$$

This shows that the new pair is weakly entangled with the rest of the system.

Lemma 2.

$$S(\{b_i\} + p) \geq S_0 - \epsilon \tag{44}$$

Lemma 3.

$$S(c_{n+1}) > \ln(2) - \epsilon \tag{45}$$

These all lead to the following theorem

Theorem 3.1. *Let* S_0 *be the entanglement entropy at time step* t_n *of the exterior quanta* $\{b_i\}$ *with the interior modes. At time step* t_{n+1}*, the pair* p *is created, and the new state deviates from the leading order one by a small amount less than* $\epsilon \ll 1$ *as defined in Eq.* (39)*. After this time step, the entropy of the emitted quanta* $\{b_i, b_{n+1}\}$ *will satisfy*

$$S(\{b_i\} + b_{n+1}) > S_0 + \ln(2) - 2\epsilon \tag{46}$$

Thus the entanglement entropy of exterior quanta necessarily increases with each emission if the corrections are small.

Proof. The proof follows using the **strong subadditivity theorem**

$$S(A + B) + S(B + C) \geq S(A) + S(C) \tag{47}$$

Setting $A = \{b_i\}$, $B = b_{n+1}$, and $C = c_{n+1}$, this gives

$$S(\{b_i\} + b_{n+1}) + S_p \geq S(\{b_i\}) + S(c_{n+1}) \tag{48}$$

Using the above lemmas, we obtain the result □

Why is the above important? Naively, one might have guessed that small departures from the leading order process cannot resolve the paradox. The reason why this might not hold is that the new pair p have many $\{c_i\}$ particles inside the black hole that they could get entangled with. This large number, paired with delicate entanglement structure, may be enough to overcome the smallness of the corrections and resolve the paradox.

A analogy can be drawn with burning a piece of paper or coal. When you burn paper, the radiation captures the information of the paper, but it is difficult to extract this information because it is delicately encoded in correlations between radiation quanta. Naively, this could have also been true for Hawking evaporation. The reason this is not true is because of the smallness of the corrections. In burning a piece of paper, let us assume the paper atoms are in a state $|1\rangle_a$. We then evolve this to a state with atoms plus radiation:

$$|1\rangle_a \to \frac{1}{\sqrt{2}}\Big(|1\rangle_a |1\rangle_{rad} + |2\rangle_a |0\rangle_{rad}\Big) \equiv |\gamma_1\rangle \tag{49}$$

$|2\rangle_a$ is an atom state orthogonal to the other. If we had started in this state, the evolution would have been

$$|2\rangle_a \to \frac{1}{\sqrt{2}}\Big(|1\rangle_a |1\rangle_{rad} - |2\rangle_a |0\rangle_{rad}\Big) \equiv |\gamma_2\rangle \tag{50}$$

This is orthogonal to $|\gamma_1\rangle$, as unitary evolution must map orthogonal states to orthogonal states. The point is that the t_{n+1} state depends on the t_n to *leading order*. While there are delicate correlations, there is no "smallness" here. In contrast, the black hole *always* produces roughly the same state during radiation – regardless of the state of the black hole, the pair-produced state is the same.

We are now ready for the Hawking "Theorem"

Theorem 3.2. *If we assume*

- *The SSL niceness conditions* $\{\mathcal{N}_i\}$ *and local Hamiltonian evolution*
- *A traditional black hole with information-free horizon exists*

then formation and evaporation of such a black hole will lead to mixed states/remnants.

Proof. We go one step at a time

- From the above, we know that there is a slicing of the black hole geometry that satisfies the SSL conditions are satisfied
- In a region where the SSL is valid, we can identify and follow the evolution of an outgoing mode with wavelength $\lambda \simeq \frac{M}{100}$. Since SSL is valid, we can expand this mode as

  $$|\psi\rangle_{mode} = \alpha_0 |0\rangle + \alpha_1 |1\rangle + \cdots \tag{51}$$

  Such an expansion can have one of two forms

  - $\sum_i |\alpha_i|^2 \simeq 1$
  - $\sum_i |\alpha_i|^2 < \epsilon' \ll 1$

In the former case, the form implies there are particles with wavelength $\simeq M/100$ at the horizon. Thus the horizon is not the "vacuum" and we do not have a traditional black hole, in violation of the second assumption of the theorem. In the latter case, the horizon is the vacuum and the SSL informs us that the evolution of the mode will have to agree with the leading order evolution of vacuum modes on this geometry up to deviations controlled by a parameter ϵ. For $\epsilon \ll 1$, Eq. (39) is satisfied as the state evolves and pair creation takes place.

- In the second scenario, the SSL implies that the created pair will be in a state close to $|\gamma_1\rangle$. By the preceeding theorem, the entanglement entropy of the radiation *increases* by at least $\ln(2) - 2\epsilon$ with each time step. Unlike a normal hot body, the entropy cannot start decreasinf after the evaporation halfway point.

- The evaporation process produces $N \sim (M/m_{pl})^2$ pairs before the black hole reaches Planckian size. At this point, the outgoing radiation has a large entanglement entropy of

$$S \simeq \frac{N}{2} \ln(2) \tag{52}$$

  From the previous sections, we return to the notion that we are forced into either mixed states or remnants.

□

A corollary follows:

Corollary 3.2.1. *If the state of Hawking radiation has to be a pure state with no entanglement with the rest of the black hole, then the evolution of low energy modes at the horizon has to be altered by order unity.*

A small change in the state at the horizon changes this entanglement by only a small fraction, and cannot reduce it to zero. Conversely, if we wish this entanglement to be zero then we have to change the state of the created pairs to a state that is close to being orthogonal to the semiclassically expected one A small change in the state at the horizon changes this entanglement by only a small fraction, and cannot reduce it to zero. Conversely, if we wish this entanglement to be zero then we have to change the state of the created pairs to a state that is close to being orthogonal to the semiclassically expected one

4 What does it take?

4.1 Misconceptions

- Misconception 1: AdS/CFT Solves the Information Paradox
  One can sometimes come across the logic that the information paradox is

solved because holography implies that that black hole system is dual to a CFT, which is inherently unitary. Thus the formation and evaporation of a black hole must also be unitary. However, this fact is besides the point. The information paradox is not just a question about fixing the issue of pure$\rightarrow$ mixed state evolution – one must also explain *how* the logic presented above is altered.

- Misconception 2: The Paradox is (only) about information
The problem discussed above is often referred to as the information paradox, but no where did we reference information carried by the initial matter state. This is because the paradox involves two distinct (but related) issues: namely, the evolution of pure states to mixed states *and* the loss of information of the initial matter state. These are separate issues because it is possible to have final radiation states that are mixed but carry the information of the initial state. It is also possible to have a final radiation state that is pure but carries no information on the initial state. Thus "information" is only one aspect of the problem. Presumably the solution to the paradox will address both issues simultaneously.

4.2 Implications

Here we consider the consequences of the Hawking theorem. [1] argues that we have three choices:

- New physics: If one wishes to keep the assumptions of the theorem, then one is forced to predict some new physics occurs as the black hole evaporates

- No traditional black holes: Rather than an information-free horizon, the black hole has some distortions (hair) that depend on the state of the black hole. Generally such hair produces divergent stress-energy at the horizon.

- Incompleteness of the niceness conditions: One could argue the above niceness conditions were not sufficient to guarentee Hamiltonian evolution from one slice to the next.

References

[1] S. D. Mathur. "The Information paradox: A Pedagogical introduction". In: *Class. Quant. Grav.* 26 (2009). Ed. by A. M. Uranga, p. 224001. DOI: 10.1088/0264-9381/26/22/224001. arXiv: 0909.1038 [hep-th].

[2] A. Miscioscia. "Notes on Hawking Radiation". In: *DESY Workshop Seminar: Black Holes*. Jan. 2024.