The Black Hole Information Paradox

in the Age of

Holographic Entanglement Entropy

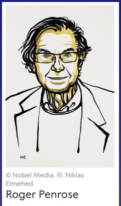
Netta Engelhardt

GGI Colloquium

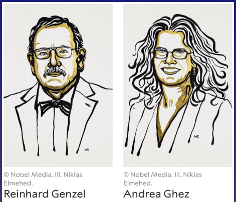
"Black holes are the most mysterious objects in the universe."
–France Córdova

image credt: LIGO collaboration

2020 Nobel Prize in Physics



"For the discovery that black hole formation is a robust prediction of the general theory of relativity."



"For the discovery of a supermassive compact object at the centre of our galaxy."

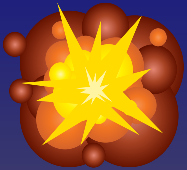
What we know and don't know

- Black holes also serve as a stage for gravitational phenomena that we don't understand.
- e.g. deep inside black holes (or in the pre-inflationary universe), general relativity breaks down.
- GR is *not* a complete theory of gravity: when spacetime is curved on quantum scales, the universe cannot be described in terms of weakly-interacting GR and QFT: **a quantum theory of gravity is needed**.
- Potential snag: we don't really have one*.
- This is a problem. Black holes **do exist** thanks, LIGO!; our description of the universe is incomplete without a theory that accurately describes what happens deep inside black holes.

The Information Paradox: the what

- Expectation: weakly interacting (decoupled) GR and QFT is valid near the event horizon of a (large) black hole.

- The information paradox is an apparent contradiction between the predictions of gravity and quantum mechanics in this regime: *at the event horizon of a black hole*.



The Information Paradox: the what

- If the event horizon were a regime where strong quantum gravity effects are always important, it wouldn't be a paradox: we would simply say that semiclassical gravity is **not** valid there. But this is not such a regime!
- We expect semiclassical gravity to be valid at the event horizon: the predictions of quantum mechanics should not have significant impact on the predictions of gravity, and vice versa.
- And yet they do.

The Information Paradox: a Caricature

Caricature: Alice and Bob are entangled, and Alice ends up in the black hole while Bob doesn't. If the black hole evaporates, we seem to find that information is lost.

In particular, a pure state evolves into a mixed state, in violation of unitarity of quantum mechanics.

The Information Paradox: the why

- We have never seen violations of unitarity. But then again, we have also never seen violations of GR.
- Information loss would be a catastrophic loss of determinism.
- Either way, since black holes exist, there is a definitive answer about whether information is lost or conserved.
- Ultimately, the information paradox has a lot to teach us about quantum gravity... provided we can get insight into it without having a direct formulation of quantum gravity.

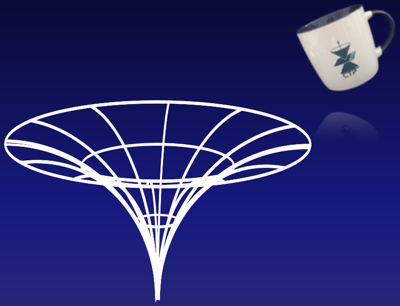
The Information Paradox: Then and Now

- Pre-2019: thought that nonperturbative quantum gravity is *necessary* to see evidence of unitarity (i.e. information conservation) in black hole evaporation
- The CW was that semiclassical gravity appears to be losing information – and somehow quantum gravity fixes it as nonperturbative corrections add up

- May 2019 Almheiri, Engelhardt, Marolf, Maxfield; Penington: a **semiclassical** analysis that is a smoking gun signal of information conservation. I think it's safe to say that most of us thought this was impossible.

Table of Contents

BH Entropy: Wheeler's Gedankenexperiment



"Teacup Experiment"

If a black hole has no microstates, then it has no entropy. If you take a hot teacup (which most certainly has entropy) and throw it into the black hole, then you have decreased the entropy of the universe.

Black Hole Entropy

If your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation.
– Sir Arthur Eddington

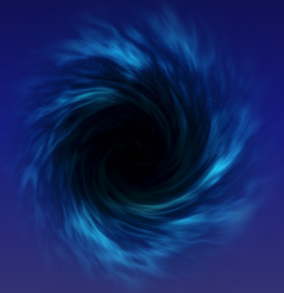
Bekenstein: Black holes are thermal objects: they have temperature and entropy. This thermodynamic entropy is due to quantum gravity microstates and is given by:

$$S = \frac{\text{Area[event horizon]}}{4G\hbar}.$$

For a black hole close to equilibrium, we expect a temperature $T \propto M^{-1}$.

The Need for Quantum Corrections

This doesn't actually make sense as anything more than analogy in classical gravity, because objects with a temperature have to emit radiation, and black holes in classical GR are perfect absorbers.



In fact, to actually calculate the temperature of a black hole, need to work in *semiclassical* gravity – where quantum fields are very weakly (perturbatively) interacting with gravity.

Evaporating Black Holes

So black holes radiate, and potentially evaporate if $T_{\mathrm{BH}} > T_{\mathrm{surroundings}}$. But

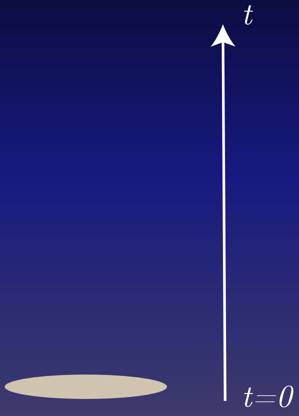
$$T \propto M^{-1}$$

As the black hole evaporates, it gets hotter, so it doesn't stop evaporating – until it is completely gone.

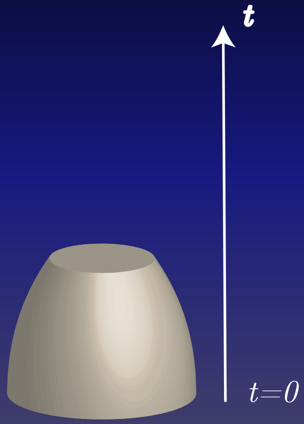
Some numbers:

- Sagittarius A*, $T \sim 10^{-14} K$
- Solar mass black hole, $T \sim 10^{-8} K$
- Lunar mass black hole, $T \sim T_{CMB}$
- Coronavirus-sized black hole, $T \sim$ room temperature.

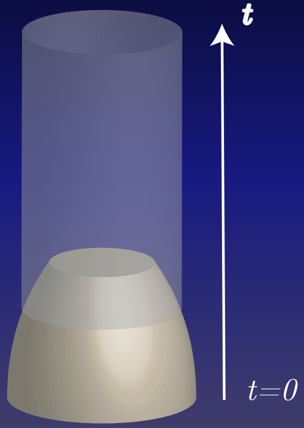
Collapsing Black Holes



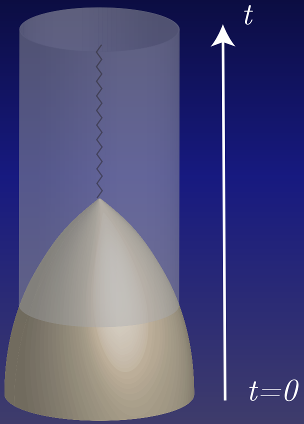
Collapsing Black Holes



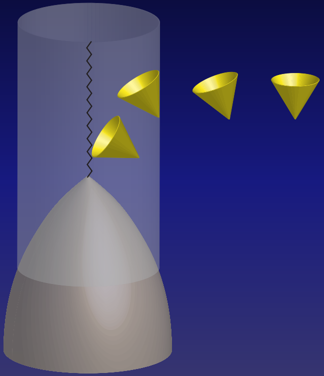
Collapsing Black Holes



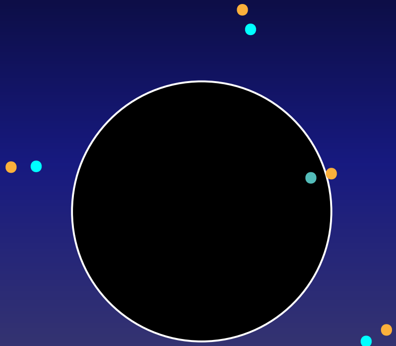
Collapsing Black Holes



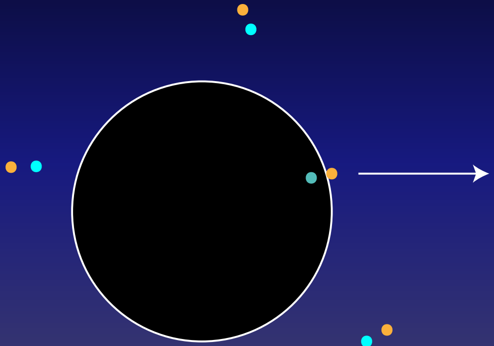
Collapsing Black Holes



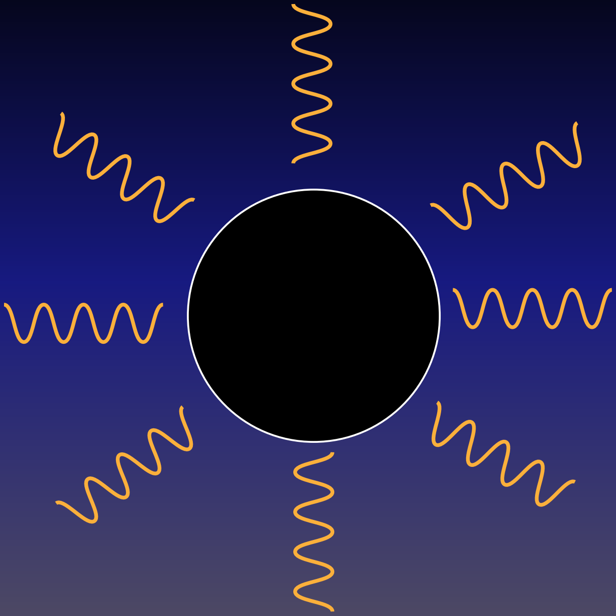
Evaporating Black Holes



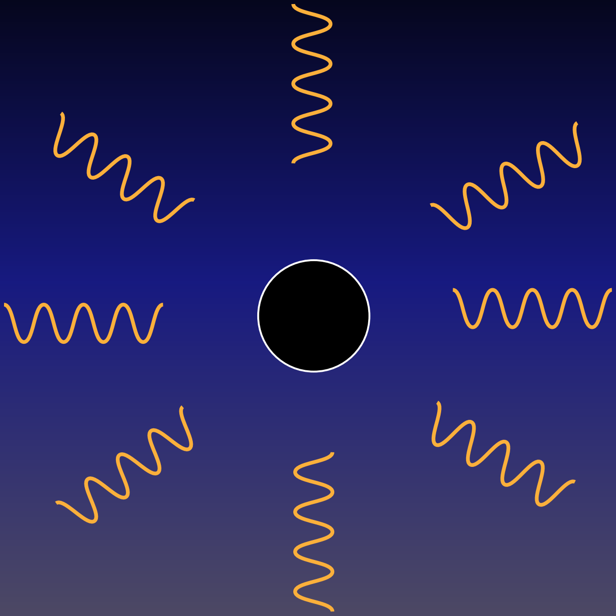
Evaporating Black Holes



Evaporating Black Holes

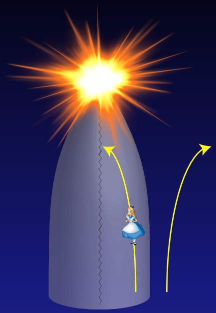


Evaporating Black Holes



Evaporating Black Holes

So Where is the Paradox?



In a dramatically oversimplified model not very accurate, but still gives a qualitative idea of what's going on, let's take an entangled pair (the Alice and Bob of before), where one – the "left" particle – falls across the horizon:

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|+\rangle|+\rangle + |-\rangle|-\rangle\right)$$

The right particle by itself can only be described by a density matrix:

$$\rho_R = \frac{1}{2}\left(|+\rangle\langle+| + |-\rangle\langle-|\right) = \frac{1}{2}\mathbb{1}$$

Entanglement Entropy of a Quantum System

For a density matrix ρ, the von Neumann (entanglement) or fine-grained entropy is defined:

$$S_{vN}[\rho] = -\mathrm{tr}\rho \log \rho$$

Properties of S_{vN}:

- It is invariant under unitary evolution.
- It vanishes for a pure state (and only for a pure state): $S_{vN}[|\psi\rangle\langle\psi|] = 0$
- It is bounded from above by the thermal entropy (also called the coarse-grained entropy)

Information Loss

Some very rough intuition: we can quantify in a rough way the information in a state in terms of the difference

$$S_{\text{thermal}} - S_{\text{vN}}.$$

The thermal state is at equilibrium and has essentially **no information content**. Thus, if a system evolves to progressively larger S_{vN}, we say that **information is lost**.

Alternatively, if we think of a pure state as the entire system (a mixed state can always be purified by adding an additional system), then S_{vN} serves as a rough guide for how short we are of having the “full system”.

This happens all the time in open systems.

But the entire universe is not an open system.

Information Loss in Hawking Radiation

In the very simplistic toy model of the EPR pair, the initial state of the pair is pure:

$$S_{\mathrm{initial}} = 0.$$

After the black hole has finished evaporating,

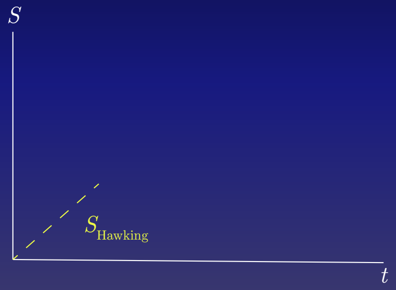
$$S_{\mathrm{final}} = S_{\mathrm{right}} = \log 2 \neq 0$$

The entropy has increased: the system was not evolving unitarily.

Hawking's Calculation

Create a black hole from a pure state, and compute the entropy of everything outside of the black hole.

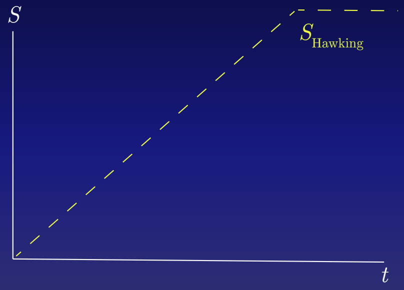
Initially, before the black hole starts evaporating, this is zero. It then increases as the black hole evaporates.



That's ok - there are two interacting subsystems – the radiation and the black hole – the entropy of a subsystem is allowed to increase.

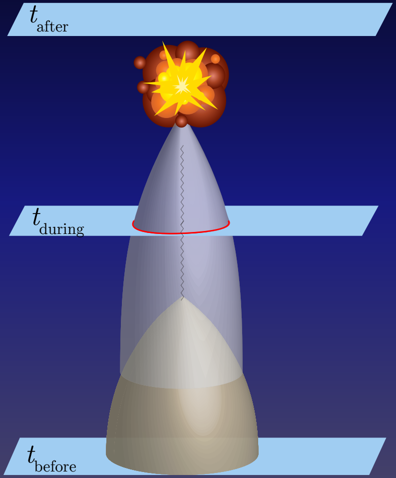
Hawking's Calculation

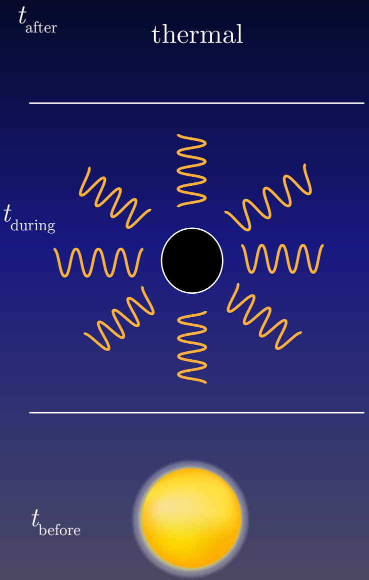
But once the black hole finishes evaporating, the entire system is *just* the radiation.



If the radiation doesn't have zero entropy after the black hole finishes evaporating, the end state of the entire universe is not pure: the universe evolves from a pure state to a mixed state, losing information.

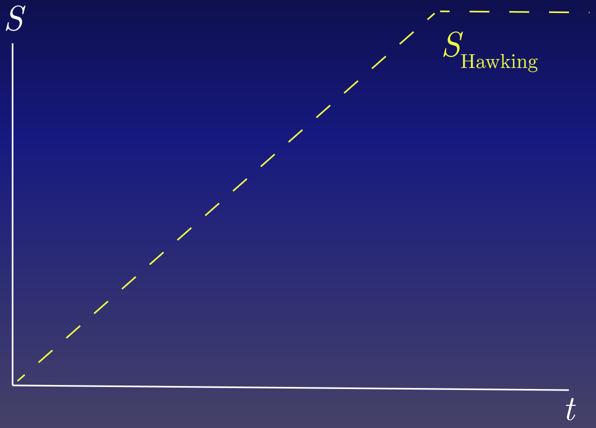
The Idea: Computing Entropy





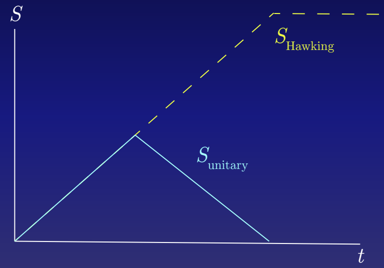
The Page Curve

A pure state in a closed system evolves to a mixed state. Information appears to be lost. Are there corrections that only come in nonperturbatively? What does a unitary entropy curve look like?



The Page Curve

The unitary curve – named after Page for proposing it:



Which is the correct curve for QG? How do we calculate it?

Table of Contents

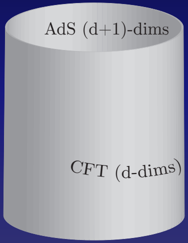
Preview: the New Developments

- From the perspective of semiclassical gravity, it's hard to see where Hawking's analysis goes wrong.
- It seems that we need an ingredient from quantum gravity to help us figure this out.
- New developments: a completely semiclassical calculation can give the unitary Page curve – but only if we use an interpretation inherited from quantum gravity.
- Specifically, inherited from holography.

Holography

AdS/CFT aka Holography

Quantum gravity "in a box" (with Anti-de Sitter boundary conditions) is dual to a lower-dimensional nongravitational QFT.



Colloquially, we call the AdS quantum gravity the bulk, and the QFT the boundary.

Holography: a Black Hole is just another Quantum System

This means that a black hole in AdS is just an ordinary, nongravitational quantum system.

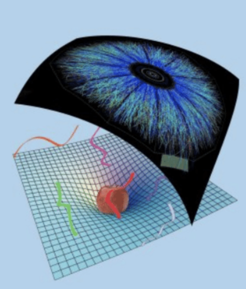


image credit: ESI Programme on AdS Holography and the Quark-Gluon Plasma

Unitarity: nongravitational isolated quantum systems don't run into an information loss problem: they evolve unitarily. So if black holes = nongrav quantum states, black holes must evolve unitarily!
Are we done?

The Idea

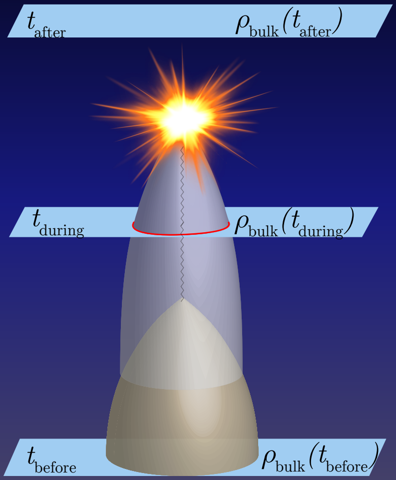
But this isn't satisfactory! How does quantum gravity preserve unitarity??

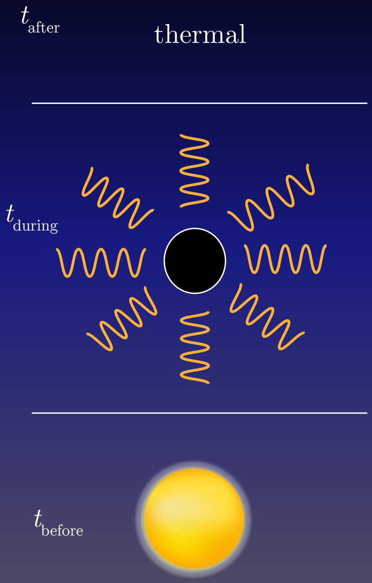
We already know that if we compute the entropy in the bulk via the usual formula:

$$S_{\mathrm{vN}}[\rho_{\mathrm{bulk}}] = -\mathrm{tr}\rho_{\mathrm{bulk}} \ln \rho_{\mathrm{bulk}}$$

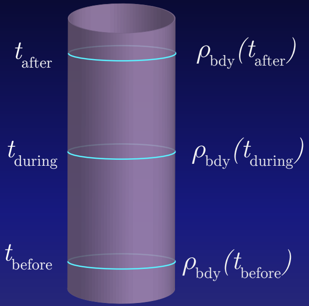
then we get the "Hawking curve": information loss (here ρ_{bulk} is the density matrix describing the quantum fields in the bulk portion outside of the black hole).

The Idea: Computing Entropy





Meanwhile, on the back of the soupcan...



If we compute the entropy in the CFT via the usual formula:

$$S_{\text{vN}}[\rho_{\text{bdy}}] = -\text{tr}\,\rho_{\text{bdy}} \ln \rho_{\text{bdy}}$$

We get a unitary result: information is conserved.

The Idea

The holographic dictionary gives us another way of computing the entropy $-\mathrm{Tr}\rho_{\mathrm{bdy}} \ln \rho_{\mathrm{bdy}}$ using the **gravitational language of the bulk**.

Basic Idea

Compute the entropy of the boundary from the gravitational theory using the holographic dictionary that relates boundary entropies to gravitational bulk quantities.

It turns out, remarkably, that a *holographic* calculation of the entropy in the semiclassical regime does, in fact, yield a unitary Page curve.

Holographic Entanglement Entropy: Semiclassical Regime

The von Neumann entropy of a density matrix $\rho_{\rm bdy}$ dual to a bulk with perturbative quantum gravity Engelhardt-Wall:

$$S_{\rm vN}[\rho_{\rm bdy}] = \frac{\text{Area}[\chi]}{4G\hbar} + S_{\rm vN}[\rho_{{\rm out}[\chi]}] \equiv S_{\rm gen}[\chi]$$

where χ is the minimal-$S_{\rm gen}$ surface which is an *extremum* of $S_{\rm gen}$. We call it a *quantum* extremal surface (QES). in the limit where the gravitational theory does not interact with the bulk quantum fields, this was first discovered by Ryu-Takayanagi, Hubeny-Rangamani-Takayanagi, and Faulkner-Lewkowycz-Maldacena

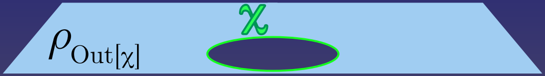
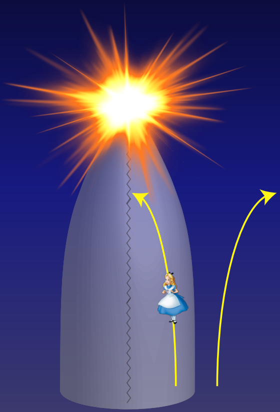


Table of Contents

Evaporating AdS Black Holes

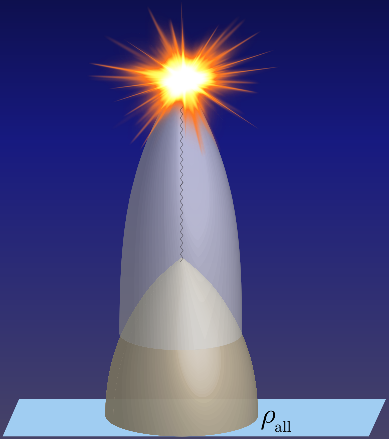
Let's consider now an evaporating AdS black hole. The next few slides are based on Almheiri, Engelhardt, Marolf, Maxfield; Penington.



Quantum Extremal Surfaces

Initially, at early times, the QES is the empty set.

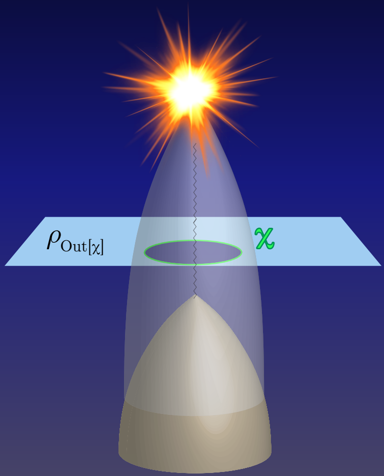
$$S_{vN}[\rho_{bdy}] = \frac{\text{Area}[\varnothing]}{4G\hbar} + S_{\text{vN}}[\rho_{\text{all}}] = S_{\text{vN}}[\rho_{\text{all}}]$$

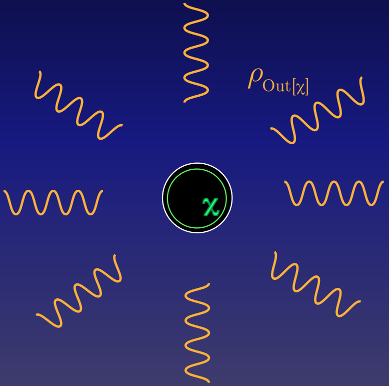


As the black hole evaporates into the external system, $S_{\text{vN}}[\rho_{\text{all}}]$ grows.

Quantum Extremal Surfaces

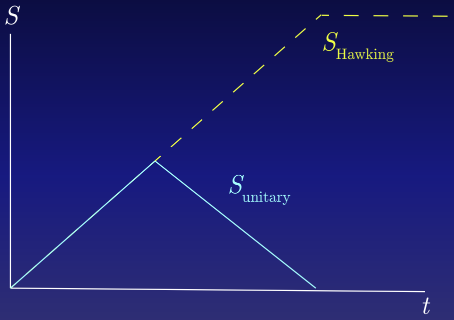
After the black hole has evaporated half its mass, a quantum extremal surface with smaller S_{gen} than the empty set appears:





As we evolve forward in time, this entropy decreases.

The Page Curve



The Page curve for unitary evolution is obtained using only semiclassical physics!

Taking a Step Back

What happened here?

$$S_{vN}[\rho_{\mathrm{bdy}}] = \frac{\mathrm{Area}[\chi]}{4G\hbar} + S_{vN}[\rho_{\mathrm{Out}[\chi]}]$$

where $\rho_{\mathrm{Out}[\chi]}$ is exactly the state that we compute from semiclassical gravity – that Hawking computed – and that gave a non-unitary answer when plugged into the non-holographic entropy formula.

$S[\rho_{\mathrm{bdy}}]$ calculates a unitarily evolving entropy.

The contribution from χ – the quantum extremal surface prescription – saves the day.

The only place where quantum gravity appeared is in the **interpretation** of $S_{\mathrm{gen}}[\chi]$ as the entropy of the radiation.

Why does this work?

Why does quantum gravity repackage unitarity in this way?

What microscopic Planckian physics is responsible for the success of the quantum extremal surface prescription?

Some Progress on this Front

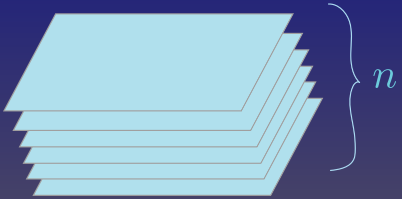
How do we generally compute entropies?

The Replica Trick

In a nongravitational theory, we can compute the von Neumann entropy using the replica trick

$$S_{\mathrm{vN}}[\rho] = -\mathrm{tr}\rho\ln\rho = -\lim_{n\to1}\left(\frac{\partial}{\partial n}\log\mathrm{Tr}[\rho^n]\right)$$

where ρ^n is the state ρ on n independent copies ("replicas") of this nongravitational theory.



Some Progress on this Front

The Replica Trick in Holography:

- In AdS/CFT, we know how to relate $\mathrm{Tr}(\rho^n)$ to the "gravitational path integral":

$$\int D g e^{S}$$

- Doing this, Penington et al; Almheiri et al find that we can justify the new QES that leads to unitarity from the gravitational path integral, but at the cost of these n replicas having nontrivial correlations.

- But these are n-independent copies of a single quantum theory! How can they be correlated?

The Gravitational Path Integral

In 2D gravity, the only exact example we understand, the gravitational path integral is computing a disorder average – an average over theories.

Analogous phenomena appear in spin glasses Lots more to say on this, please ask in Q&A.

But it is not clear what this means in more general theories of gravity.

But what I would argue *is* clear is that we have pinned down a key ingredient in the resolution of the information paradox: the gravitational path integral.

We *must* understand why it is dominated by these correlated replicas, why those give rise to unitary evolution and what their contributions mean for the gravitational theory.

Current State of the Information Paradox

Spoiler: **we haven't solved it yet.**

- What entropy is the quantum extremal surface/gravitational path integral calculating? How do we compute it directly from the formula $-\mathrm{tr}\rho \ln \rho$?

  0.3cm

- What does an observer outside the black hole actually measure?

- Is the path integral doing an averaging of some sort? Over what? Is the semiclassical limit of gravity an approximate description of an ensemble of quantum theories?

- What are the internal dynamics of an evaporating black hole?

- Where did Hawking go wrong?

Thank you!