Recent progress on the black hole information problem

Daniel Harlow

MIT

June 27, 2023

Introduction

- Hawking's black hole information problem has been one of the driving challenges of theoeretical physics for the last 50 years.

Introduction

- Hawking's black hole information problem has been one of the driving challenges of theoeretical physics for the last 50 years.
- It suggests a tension between gravity and quantum mechanics, neither of which we are willing to give up lightly.

Introduction

- Hawking's black hole information problem has been one of the driving challenges of theoeretical physics for the last 50 years.
- It suggests a tension between gravity and quantum mechanics, neither of which we are willing to give up lightly.
- In the last 10-15 years we have improved substantially our understanding of how spacetime emerges in holography, and in the last few years this has grown into a new understanding of Hawking's paradox.

Introduction

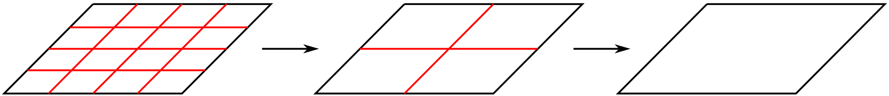
- Hawking's black hole information problem has been one of the driving challenges of theoeretical physics for the last 50 years.
- It suggests a tension between gravity and quantum mechanics, neither of which we are willing to give up lightly.
- In the last 10-15 years we have improved substantially our understanding of how spacetime emerges in holography, and in the last few years this has grown into a new understanding of Hawking's paradox.
- It is too soon to say that the problem has been fully resolved, but I think it is fair to say that many of us feel a resolution is in sight.

Introduction

- Hawking's black hole information problem has been one of the driving challenges of theoeretical physics for the last 50 years.
- It suggests a tension between gravity and quantum mechanics, neither of which we are willing to give up lightly.
- In the last 10-15 years we have improved substantially our understanding of how spacetime emerges in holography, and in the last few years this has grown into a new understanding of Hawking's paradox.
- It is too soon to say that the problem has been fully resolved, but I think it is fair to say that many of us feel a resolution is in sight.
- In this talk I will attempt to give a brief overview of what I think is the current status.

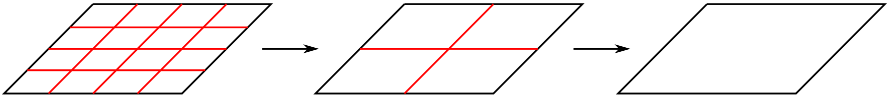
The information problem arises from a basic tension between expanding spacetimes and effective field theory.

The information problem arises from a basic tension between expanding spacetimes and effective field theory.

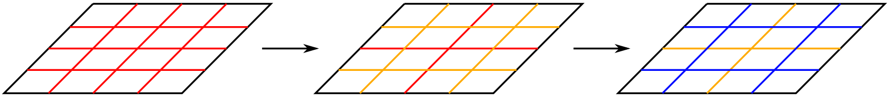


In an expanding spacetime short-wavelength modes are stretched into long-wavelength modes.

The information problem arises from a basic tension between expanding spacetimes and effective field theory.

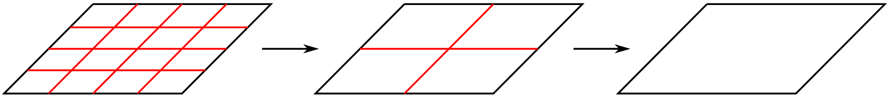


In an expanding spacetime short-wavelength modes are stretched into long-wavelength modes.

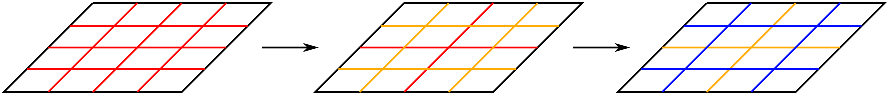


In the presence of some kind of short-distance cutoff (say at the Planck scale), this means that to preserve the cutoff scale we need to introduce new modes at short distances.

The information problem arises from a basic tension between expanding spacetimes and effective field theory.



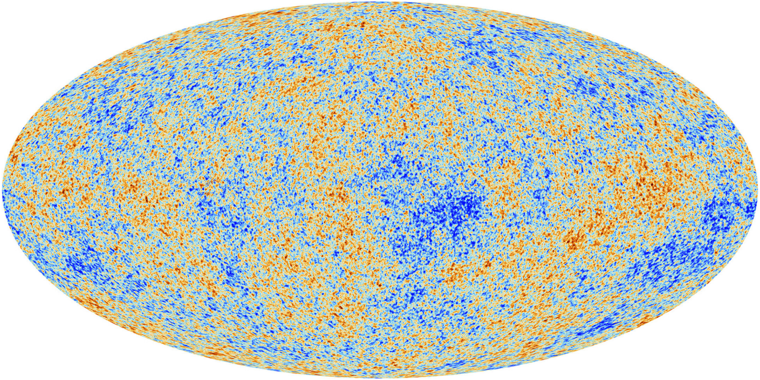
In an expanding spacetime short-wavelength modes are stretched into long-wavelength modes.



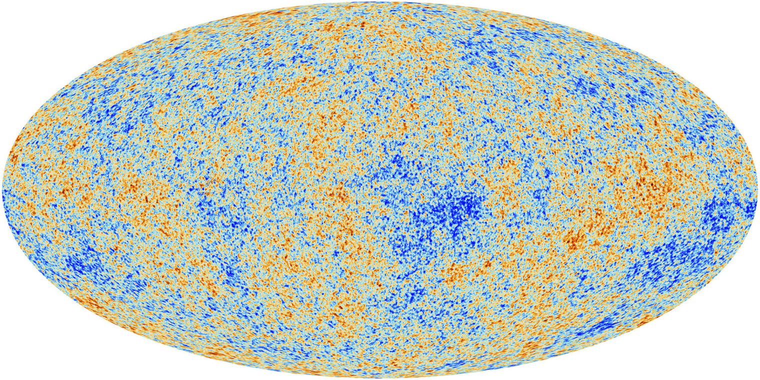
In the presence of some kind of short-distance cutoff (say at the Planck scale), this means that to preserve the cutoff scale we need to introduce new modes at short distances.

We therefore need a rule for what state these new modes are in. The only rule which seems to make any sense is to say that (roughly speaking) these new modes enter in their vacuum state.

This may just sound like some rule that I made up, but in fact it has been confirmed by observation:

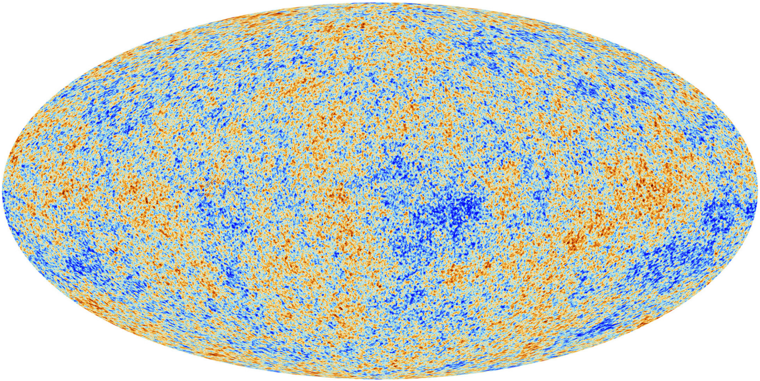


This may just sound like some rule that I made up, but in fact it has been confirmed by observation:



In the theory of inflation the structure of the universe we see today arises from exactly these vacuum fluctuations, which in most models started out substantially smaller than the Planck length.

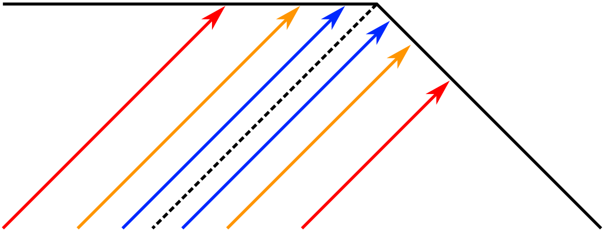
This may just sound like some rule that I made up, but in fact it has been confirmed by observation:



In the theory of inflation the structure of the universe we see today arises from exactly these vacuum fluctuations, which in most models started out substantially smaller than the Planck length.
We are all made out of sub-Planckian modes!

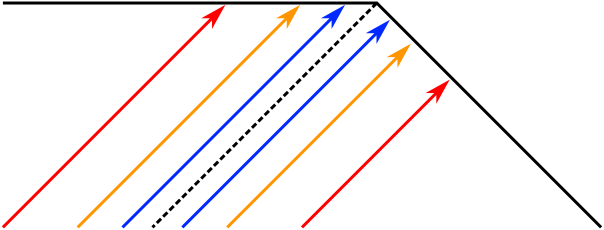
This phenomenon also arises near black hole horizons.

This phenomenon also arises near black hole horizons.



As we evolve forward in time, entangled modes straddling the horizon move away from the horizon. The interior partner falls into the singularity, while the exterior partner makes it out to infinity as Hawking radiation.

This phenomenon also arises near black hole horizons.

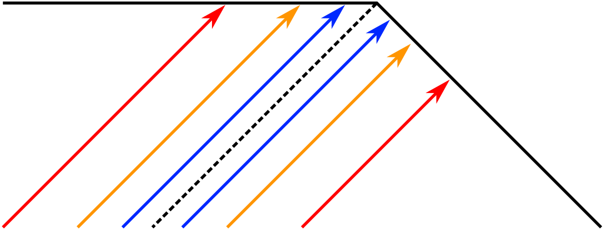


As we evolve forward in time, entangled modes straddling the horizon move away from the horizon. The interior partner falls into the singularity, while the exterior partner makes it out to infinity as Hawking radiation. The time after formation at which these modes are coming from less than the Planck distance away from the horizon is

$$t \sim \frac{\beta}{2\pi} \log S \equiv t_{scr},$$

where $\beta = 1/T \sim r_s$ and $S = \frac{A}{4G}$.

This phenomenon also arises near black hole horizons.

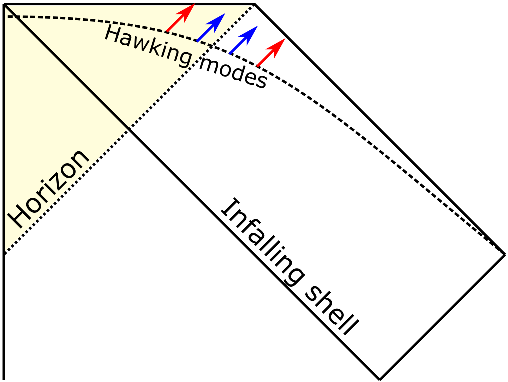


As we evolve forward in time, entangled modes straddling the horizon move away from the horizon. The interior partner falls into the singularity, while the exterior partner makes it out to infinity as Hawking radiation. The time after formation at which these modes are coming from less than the Planck distance away from the horizon is

$$t \sim \frac{\beta}{2\pi} \log S \equiv t_{scr},$$

where $\beta = 1/T \sim r_s$ and $S = \frac{A}{4G}$.
For Sagittarius A* we have $t_{scr} \sim 1000\mathrm{s}$.

Now let's recall how this leads to Hawking's paradox:



Entanglement between interior and exterior modes causes the black hole to radiate, losing energy, but this radiation cannot carry information about the infalling shell since these new modes enter in vacuum and the shell is still deep inside. By the time the black hole reaches Planckian size, it doesn't have enough energy left to return this information to the exterior.

We can describe Hawking's paradox as saying it is impossible to have the following things in one theory:

We can describe Hawking's paradox as saying it is impossible to have the following things in one theory:

(1) A finite black hole entropy

We can describe Hawking's paradox as saying it is impossible to have the following things in one theory:

(1) A finite black hole entropy

(2) A unitary black hole S-matrix

We can describe Hawking's paradox as saying it is impossible to have the following things in one theory:

(1) A finite black hole entropy
(2) A unitary black hole S-matrix
(3) Effective field theory valid away from high energy densities and/or curvatures.

We can describe Hawking's paradox as saying it is impossible to have the following things in one theory:

(1) A finite black hole entropy

(2) A unitary black hole S-matrix

(3) Effective field theory valid away from high energy densities and/or curvatures.

These are all things we really would like to be true, so any resolution of the paradox will teach us something deep!

- In some of the traditional proposals for resolving Hawking's problem, i.e. remnants and/or information loss, one gives up (1) and/or (2) to preserve (3).

- In some of the traditional proposals for resolving Hawking's problem, i.e. remnants and/or information loss, one gives up (1) and/or (2) to preserve (3).
- In particular one gives up the idea that the black hole entropy is actually given by $S = A/4G$.

- In some of the traditional proposals for resolving Hawking's problem, i.e. remnants and/or information loss, one gives up (1) and/or (2) to preserve (3).
- In particular one gives up the idea that the black hole entropy is actually given by $S = A/4G$.
- The string theory counting of black hole microstates, both directly and through the AdS/CFT correspondence, gives strong evidence that indeed we have $S = A/4G$.

- In some of the traditional proposals for resolving Hawking's problem, i.e. remnants and/or information loss, one gives up (1) and/or (2) to preserve (3).
- In particular one gives up the idea that the black hole entropy is actually given by $S = A/4G$.
- The string theory counting of black hole microstates, both directly and through the AdS/CFT correspondence, gives strong evidence that indeed we have $S = A/4G$.
- Moreover AdS/CFT give strong evidence that the S-matrix is unitary.

- In some of the traditional proposals for resolving Hawking's problem, i.e. remnants and/or information loss, one gives up (1) and/or (2) to preserve (3).
- In particular one gives up the idea that the black hole entropy is actually given by $S = A/4G$.
- The string theory counting of black hole microstates, both directly and through the AdS/CFT correspondence, gives strong evidence that indeed we have $S = A/4G$.
- Moreover AdS/CFT give strong evidence that the S-matrix is unitary.
- Aesthetically, it would really be a pity if black hole thermodynamics were fake: why should black holes behave like they have entropy $A/4G$ if they don't?

- In some of the traditional proposals for resolving Hawking's problem, i.e. remnants and/or information loss, one gives up (1) and/or (2) to preserve (3).
- In particular one gives up the idea that the black hole entropy is actually given by $S = A/4G$.
- The string theory counting of black hole microstates, both directly and through the AdS/CFT correspondence, gives strong evidence that indeed we have $S = A/4G$.
- Moreover AdS/CFT give strong evidence that the S-matrix is unitary.
- Aesthetically, it would really be a pity if black hole thermodynamics were fake: why should black holes behave like they have entropy $A/4G$ if they don't?

Our challenge is thus to understand what replaces (3).

It is worth emphasizing how drastic the necessary violation of (3) must be.

It is worth emphasizing how drastic the necessary violation of (3) must be.

- The most extreme modification would be to say that the interior is destroyed already at t_{scr} - there is a "firewall" at the horizon* which destroys anyone who falls in.

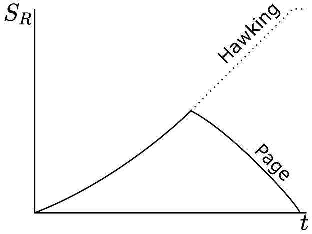
It is worth emphasizing how drastic the necessary violation of (3) must be.

- The most extreme modification would be to say that the interior is destroyed already at t_{scr} - there is a "firewall" at the horizon* which destroys anyone who falls in.
- It is not clear however why this should happen for black holes but not for the CMB, and anyways we shouldn't accept this unless we really have no other choice.

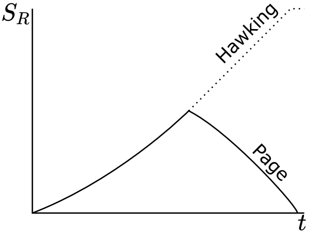
It is worth emphasizing how drastic the necessary violation of (3) must be.

- The most extreme modification would be to say that the interior is destroyed already at t_{scr} - there is a "firewall" at the horizon* which destroys anyone who falls in.
- It is not clear however why this should happen for black holes but not for the CMB, and anyways we shouldn't accept this unless we really have no other choice.
- On the other hand if we think Hawking's picture is valid until times of order the evaporation time, then at these late times the infalling shell is spacelike-separated at great distance from the Hawking radiation. For information to get out, *severe* non-locality is necessary: at distances of order $10^{97}\mathrm{m}$ for Sagittarius A*!

Another illustration of the level of modification which is necessary is the "Page curve", which plots the von Neumann entropy of the radiation as a function of time:

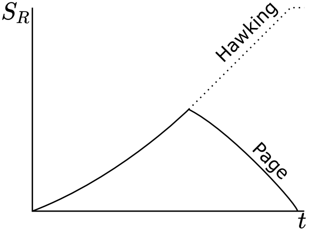


Another illustration of the level of modification which is necessary is the "Page curve", which plots the von Neumann entropy of the radiation as a function of time:



Restoring unitarity changes the slope of the curve at $O(1)$, *not* at $O(e^{-S})$!

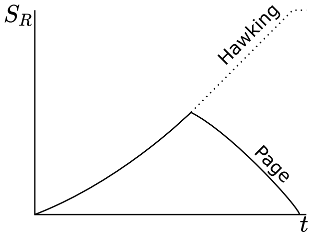
Another illustration of the level of modification which is necessary is the "Page curve", which plots the von Neumann entropy of the radiation as a function of time:



Restoring unitarity changes the slope of the curve at $O(1)$, *not* at $O(e^{-S})$! The time at which this curve turns over is called the Page time, and it is of order

$$t_{page} \sim \beta S.$$

Another illustration of the level of modification which is necessary is the "Page curve", which plots the von Neumann entropy of the radiation as a function of time:



Restoring unitarity changes the slope of the curve at $O(1)$, *not* at $O(e^{-S})$! The time at which this curve turns over is called the Page time, and it is of order

$$t_{page} \sim \beta S.$$

At this time the entropy in the interior modes (which purify the Hawking radiation) exceeds the entropy of the black hole, which seems to present a serious obstruction to the idea that $S = \frac{A}{4G}$.

We can thus organize discussion of the breakdown of effective field theory in the black hole in terms of what time we think it happens:

We can thus organize discussion of the breakdown of effective field theory in the black hole in terms of what time we think it happens:

- $t_{scr} \sim \beta \log S$: time at which Hawking radiation comes from sub-Planckian modes.

We can thus organize discussion of the breakdown of effective field theory in the black hole in terms of what time we think it happens:

- $t_{scr} \sim \beta \log S$: time at which Hawking radiation comes from sub-Planckian modes.
- $t_{page} \sim \beta S$: time at which there are more interior modes than black hole microstate degrees of freedom

We can thus organize discussion of the breakdown of effective field theory in the black hole in terms of what time we think it happens:

- $t_{scr} \sim \beta \log S$: time at which Hawking radiation comes from sub-Planckian modes.
- $t_{page} \sim \beta S$: time at which there are more interior modes than black hole microstate degrees of freedom
- $t_{evap} \sim \beta S$: time at which the black hole itself has Planckian size

We can thus organize discussion of the breakdown of effective field theory in the black hole in terms of what time we think it happens:

- $t_{scr} \sim \beta \log S$: time at which Hawking radiation comes from sub-Planckian modes.
- $t_{page} \sim \beta S$: time at which there are more interior modes than black hole microstate degrees of freedom
- $t_{evap} \sim \beta S$: time at which the black hole itself has Planckian size
- $t_{exp} \sim \beta e^S$: (non-evaporating black holes only) the time at which the number of mutually orthogonal interior states exceeds the *total* number of microstates (including superpositions).

We can thus organize discussion of the breakdown of effective field theory in the black hole in terms of what time we think it happens:

- $t_{scr} \sim \beta \log S$: time at which Hawking radiation comes from sub-Planckian modes.
- $t_{page} \sim \beta S$: time at which there are more interior modes than black hole microstate degrees of freedom
- $t_{evap} \sim \beta S$: time at which the black hole itself has Planckian size
- $t_{exp} \sim \beta e^S$: (non-evaporating black holes only) the time at which the number of mutually orthogonal interior states exceeds the *total* number of microstates (including superpositions).

What we seem to be learning is that the effective field theory description of the black hole interior is good until t_{evap}/t_{exp}, and in the remainder of this talk I will try to give a sense of how this works.

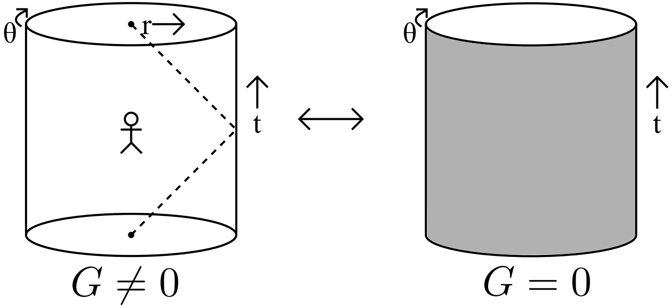
Akers/Engelhardt/Harlow/Penington/Vardhan 2022

Emergent spacetime and AdS/CFT

In understanding what might replace (3) (the validity of EFT away from singularities), it is useful to note that in our best theory of quantum gravity so far, AdS/CFT, it not obvious that *any* version of (3) holds:

Emergent spacetime and AdS/CFT

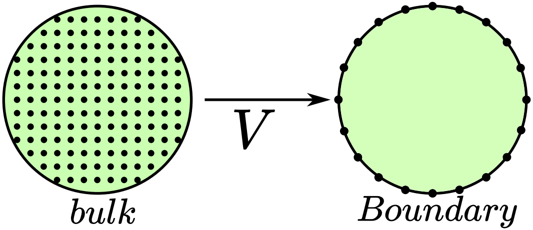
In understanding what might replace (3) (the validity of EFT away from singularities), it is useful to note that in our best theory of quantum gravity so far, AdS/CFT, it not obvious that *any* version of (3) holds:



Quantum gravity in asymptotically-AdS space is equal to quantum field theory living on the asymptotic boundary, so the bulk spacetime is at best *emergent*: it makes sense only in certain situations and only in some approximation. Maldacena 1998

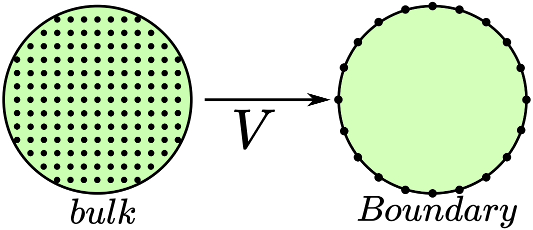
It has gradually been understood that the correct mathematical framework for describing the emergence of the bulk spacetime in AdS/CFT is *quantum error correction*. Almheiri/Dong/Harlow 2014

It has gradually been understood that the correct mathematical framework for describing the emergence of the bulk spacetime in AdS/CFT is *quantum error correction*. Almheiri/Dong/Harlow 2014



In particular for sufficiently low-energy states (no black holes) there is a holographic encoding map $V : \mathcal{H}_{bulk} \to \mathcal{H}_{boundary}$, where $\mathcal{H}_{bulk}$ is the set of low-energy bulk states and V is approximately (up to $O(e^{-N^2})$) an isometry.

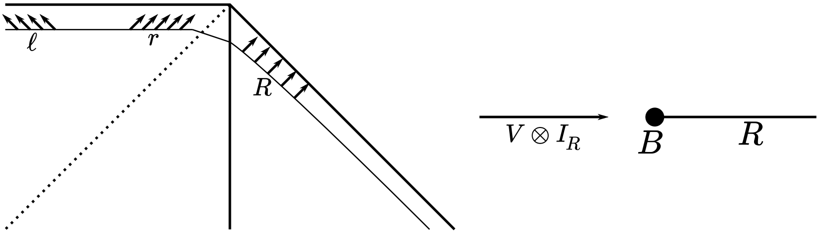
It has gradually been understood that the correct mathematical framework for describing the emergence of the bulk spacetime in AdS/CFT is *quantum error correction*. Almheiri/Dong/Harlow 2014



In particular for sufficiently low-energy states (no black holes) there is a holographic encoding map $V : \mathcal{H}_{bulk} \to \mathcal{H}_{boundary}$, where $\mathcal{H}_{bulk}$ is the set of low-energy bulk states and V is approximately (up to $O(e^{-N^2})$) an isometry.
(Recall that an isometry is a linear map $V : \mathcal{H}_A \to \mathcal{H}_B$ that preserves the inner product. Isometries can exist only if $|A| \leq |B|$.)

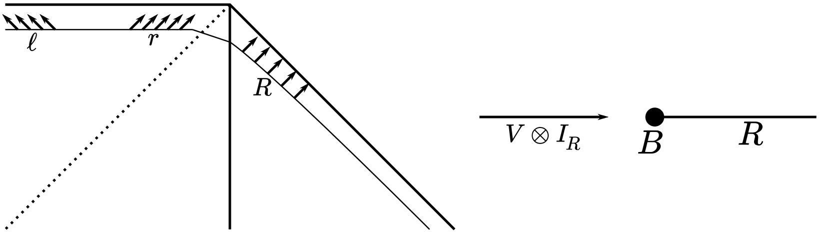
An encoding map for the BH interior

We'd like to construct a similar holographic encoding map $V : \mathcal{H}_\ell \otimes \mathcal{H}_r \to \mathcal{H}_B$ for the black hole interior:



An encoding map for the BH interior

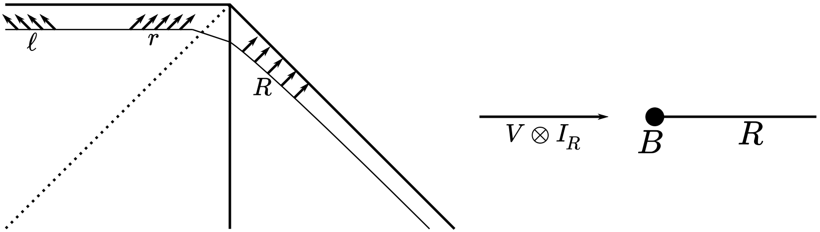
We'd like to construct a similar holographic encoding map $V : \mathcal{H}_\ell \otimes \mathcal{H}_r \to \mathcal{H}_B$ for the black hole interior:



The basic problem, which is really the essence of the information problem, is that at late times we have $|\ell||r| \gg |B|$, so the map V cannot be an isometry.

An encoding map for the BH interior

We'd like to construct a similar holographic encoding map $V : \mathcal{H}_\ell \otimes \mathcal{H}_r \to \mathcal{H}_B$ for the black hole interior:



The basic problem, which is really the essence of the information problem, is that at late times we have $|\ell||r| \gg |B|$, so the map V cannot be an isometry.

The main thing we have learned in the last few years is that we should embrace this non-isometric nature of V: it may sound scary, but understood properly it explains the difference between Page and Hawking without giving up too much on EFT!

In more detail the essence of our proposal is the following:

There is a large set of "null states" in the Hilbert space of effective field theory inside a black hole, each of which is annihilated by the holographic map to the fundamental degrees of freedom. This however cannot be detected by any observer who does not perform an operation of exponential complexity.

In more detail the essence of our proposal is the following:

There is a large set of "null states" in the Hilbert space of effective field theory inside a black hole, each of which is annihilated by the holographic map to the fundamental degrees of freedom. This however cannot be detected by any observer who does not perform an operation of exponential complexity.

In the previous language we advocate the following replacement:

~~(3) EFT valid wherever there is not a large energy density/curvature.~~

(3*) EFT valid *for sub-exponential states/observables* wherever there is not large energy density/curvature.

In more detail the essence of our proposal is the following:

There is a large set of "null states" in the Hilbert space of effective field theory inside a black hole, each of which is annihilated by the holographic map to the fundamental degrees of freedom. This however cannot be detected by any observer who does not perform an operation of exponential complexity.

In the previous language we advocate the following replacement:

- ~~(3) EFT valid wherever there is not a large energy density/curvature.~~
- (3*) EFT valid *for sub-exponential states/observables* wherever there is not large energy density/curvature.

Indeed we can construct models where appropriate analogous of (1), (2), and (3*) are all proveably true. They are thus compatible, and so if we are willing to accept (3*) then the information problem is resolved in these models.

An illustration

Here is a simple model that illustrates the basic idea.

An illustration

Here is a simple model that illustrates the basic idea.

- Say that we have an orthonormal basis of e^S states $|1\rangle, |2\rangle, \ldots, |e^S\rangle$.

An illustration

Here is a simple model that illustrates the basic idea.

- Say that we have an orthonormal basis of e^S states $|1\rangle, |2\rangle, \ldots, |e^S\rangle$.
- We learn in kindergarten that there are no states that are orthogonal to all of these, but when S is large it is easy to make a state which is *nearly* orthogonal to all of them:

$$|\psi\rangle = e^{-S/2} \sum_{n=1}^{e^S} |n\rangle.$$

An illustration

Here is a simple model that illustrates the basic idea.

- Say that we have an orthonormal basis of e^S states $|1\rangle, |2\rangle, \ldots, |e^S\rangle$.
- We learn in kindergarten that there are no states that are orthogonal to all of these, but when S is large it is easy to make a state which is *nearly* orthogonal to all of them:

$$|\psi\rangle = e^{-S/2} \sum_{n=1}^{e^S} |n\rangle.$$

- Here is another one, which is also nearly orthogonal to $|\psi\rangle$:

$$|\phi\rangle = e^{-S/2} \sum_{n=1}^{e^S} (-1)^n |n\rangle.$$

More generally we can make a toy encoding map of the black hole interior ℓr into the microstates B:

$$V|i\rangle_{\ell r} = \frac{1}{\sqrt{|B|}} \sum_b e^{i\theta(i,b)}|b\rangle.$$

More generally we can make a toy encoding map of the black hole interior ℓr into the microstates B:

$$V|i\rangle_{\ell r} = \frac{1}{\sqrt{|B|}} \sum_b e^{i\theta(i,b)}|b\rangle.$$

Let's compute the inner product of two such states:

$$\begin{aligned}\langle j|V^\dagger V|i\rangle &= \frac{1}{|B|} \sum_b e^{i\theta(i,b)-i\theta(j,b)} \\ &= \begin{cases} 1 & i=j \\ O(1/\sqrt{|B|}) & i \neq j \end{cases}.\end{aligned}$$

Noting that $|B| = e^S$, we see that these states are orthogonal up to exponentially small corrections.

More generally we can make a toy encoding map of the black hole interior ℓr into the microstates B:

$$V|i\rangle_{\ell r} = \frac{1}{\sqrt{|B|}} \sum_b e^{i\theta(i,b)}|b\rangle.$$

Let's compute the inner product of two such states:

$$\begin{aligned}\langle j|V^\dagger V|i\rangle &= \frac{1}{|B|} \sum_b e^{i\theta(i,b)-i\theta(j,b)} \\ &= \begin{cases} 1 & i=j \\ O(1/\sqrt{|B|}) & i\neq j \end{cases}.\end{aligned}$$

Noting that $|B| = e^S$, we see that these states are orthogonal up to exponentially small corrections.

You can fit a lot more nearly orthogonal states into Hilbert space than you might have thought!

Conclusion

This idea can be further developed into models which have many of the desired features of a quantum theory of black holes:

Conclusion

This idea can be further developed into models which have many of the desired features of a quantum theory of black holes:

- Black hole entropy is finite

Conclusion

This idea can be further developed into models which have many of the desired features of a quantum theory of black holes:

- Black hole entropy is finite
- There is a unitary S-matrix describing the formation and evaporation of black hole physics.

Conclusion

This idea can be further developed into models which have many of the desired features of a quantum theory of black holes:

- Black hole entropy is finite
- There is a unitary S-matrix describing the formation and evaporation of black hole physics.
- Effective field theory is valid in the interior for all observables of sub-exponential complexity.

Conclusion

This idea can be further developed into models which have many of the desired features of a quantum theory of black holes:

- Black hole entropy is finite
- There is a unitary S-matrix describing the formation and evaporation of black hole physics.
- Effective field theory is valid in the interior for all observables of sub-exponential complexity.
- The Page curve of the radiation can be computed explicitly, and agrees with the results of the "quantum extremal surface" calculation of Almheiri/Engelhardt/Marolf/Maxfield, Penington 2019.

Conclusion

This idea can be further developed into models which have many of the desired features of a quantum theory of black holes:

- Black hole entropy is finite
- There is a unitary S-matrix describing the formation and evaporation of black hole physics.
- Effective field theory is valid in the interior for all observables of sub-exponential complexity.
- The Page curve of the radiation can be computed explicitly, and agrees with the results of the "quantum extremal surface" calculation of Almheiri/Engelhardt/Marolf/Maxfield, Penington 2019.

There are still many details to fill in, but it is an exciting time and there is a palpable feeling in the community that the crux of the problem has been dealt with.

Conclusion

This idea can be further developed into models which have many of the desired features of a quantum theory of black holes:

- Black hole entropy is finite
- There is a unitary S-matrix describing the formation and evaporation of black hole physics.
- Effective field theory is valid in the interior for all observables of sub-exponential complexity.
- The Page curve of the radiation can be computed explicitly, and agrees with the results of the "quantum extremal surface" calculation of Almheiri/Engelhardt/Marolf/Maxfield, Penington 2019.

There are still many details to fill in, but it is an exciting time and there is a palpable feeling in the community that the crux of the problem has been dealt with.

Thanks for listening!