SIMPLICITY IN THE BLACK HOLE INTERIOR

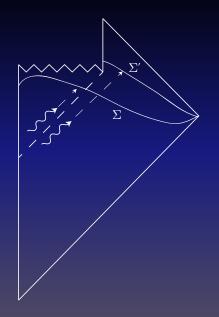
Netta Engelhardt

MIT

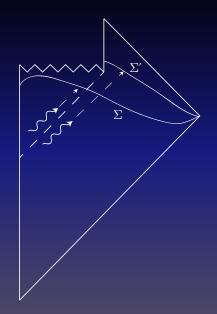


Quantum Gravity and All of That

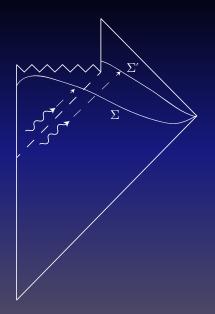
 The past two years have seen a lot of progress on the black hole information paradox



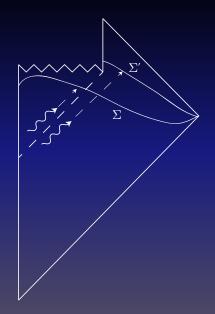
- The past two years have seen a lot of progress on the black hole information paradox
- Insights from these developments have started to teach us more about gravity in general



- The past two years have seen a lot of progress on the black hole information paradox
- Insights from these developments have started to teach us more about gravity in general
- This is of course why we are interested in the BH info paradox in the first place: to learn more about quantum gravity.



- The past two years have seen a lot of progress on the black hole information paradox
- Insights from these developments have started to teach us more about gravity in general
- This is of course why we are interested in the BH info paradox in the first place: to learn more about quantum gravity.
- But this is not (yet) a time for singing paeans: despite rapid progress, we have yet to actually resolve the paradox.



Where we stand now

 In 2019, we Almheiri, NE, Marolf, Maxfield; Penington computed the unitary Page curve from the QES formula NE, Wall '14:

$$S_{vN}[\rho_R] = rac{\operatorname{Area}[\chi_R]}{4G\hbar} + S_{\mathrm{out}}[\chi_R] = S_{\mathrm{gen}}[\chi_R]$$

where χ_R is the minimal- S_{gen} surface that extremizes S_{gen} . (In the classical case we extremize just the area RT, HRT, FLM)



Where we stand now

 In 2019, we Almheiri, NE, Marolf, Maxfield; Penington computed the unitary Page curve from the QES formula NE, Wall '14:

$$S_{vN}[\rho_R] = \frac{\text{Area}[\chi_R]}{4G\hbar} + S_{\text{out}}[\chi_R] = S_{\text{gen}}[\chi_R]$$

where χ_R is the minimal- $S_{\rm gen}$ surface that extremizes $S_{\rm gen}$. (In the classical case we extremize just the area RT, HRT, FLM)



• In simple setups, the application of this formula to the radiation, as suggested by Hayden, Penington; Penington; Almheiri Mahajan Maldacena Zhao has since been derived from the gravitational path integral (GPI) Penington et al; Almheiri et al

Hawking vs the QES

- Hawking's calculation can be done without any input from quantum gravity – it's a purely semiclassical calculation.
- While the QES calculation can be executed using only semiclassical gravity, interpreting it as a calculation of a von Neumann entropy requires input from quantum gravity.
- In particular, reconstruction of the interior rfrom the radiation equires input from quantum gravity.
- To resolve the information paradox, we need to understand what input Hawking's calculation the fully semiclassical gravity answer would need to give the same result as obtained from the QES calculation which is only operationally semiclassical.

Where is Missing from Hawking's Calculation?

We have **two** ways of computing the entropy of Hawking radiation:

- 1. The QES formula, or equivalently the gravitational path integral
- 2. Hawking's calculation (via Bogoliubov transformations, etc.).

They give different answers; so to resolve the information paradox we must ask: where do these two approaches diverge?

Where semiclassical gravity goes wrong: Pre-2019

If you had asked me before 2019 what's missing from Hawking's calculation, I probably would have said:

"Oh you know, Harlow-Hayden probably means that somewhere, somehow, Hawking's calculation lost track of any exponentially complex data in the state".

Where semiclassical gravity goes wrong: Pre-2019

If you had asked me before 2019 what's missing from Hawking's calculation, I probably would have said:

"Oh you know, Harlow-Hayden probably means that somewhere, somehow, Hawking's calculation lost track of any exponentially complex data in the state".

The logic here follows from Harlow-Hayden, Aaronson, Kim-Tang-Preskill, who showed that decoding the Hawking radiation is exponentially hard.

The idea here is that the Hawking radiation after the Page time walks like a thermal state and quacks like a thermal state, but is not in fact a thermal state. (If it were a thermal state it would of course have no information about the infalling matter)

The idea here is that the Hawking radiation after the Page time walks like a thermal state and quacks like a thermal state, but is not in fact a thermal state. (If it were a thermal state it would of course have no information about the infalling matter)

But to see that it is not a thermal state – i.e. that it in fact contains information about the infalling matter – is exponentially hard.

The idea here is that the Hawking radiation after the Page time walks like a thermal state and quacks like a thermal state, but is not in fact a thermal state. (If it were a thermal state it would of course have no information about the infalling matter)

But to see that it is not a thermal state – i.e. that it in fact contains information about the infalling matter – is exponentially hard.

It's easy to mistake the Hawking radiation for thermal.

The idea here is that the Hawking radiation after the Page time walks like a thermal state and quacks like a thermal state, but is not in fact a thermal state. (If it were a thermal state it would of course have no information about the infalling matter)

But to see that it is not a thermal state - i.e. that it in fact contains information about the infalling matter - is exponentially hard.

It's easy to mistake the Hawking radiation for thermal.

Hawking's calculation could potentially be coarse-graining over high complexity.

Where semiclassical gravity goes wrong: Post-2019

If you ask people now, most would probably say something along the lines of:

"Isn't it obvious? He used the wrong saddle in the gravitational path integral!"

The "wrong" saddle corresponds to a subdominant QES coming from the disconnected topology.



(Of course, it's not clear where in Hawking's calculation he started to indiscriminately impose ignorance of high-complexity data. And he certainly didn't use the gravitational path integral. But we're trying to bridge the gap between two calculations here.)

Hawking vs the QES, Again

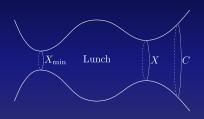
How are these two perspectives – complexity and saddle – compatible? By understanding the way in which these two mistakes are one and the same, we can make progress towards a resolution of the paradox.

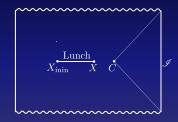
What does it mean in terms of the Lorentzian bulk geometry to (a) implement ignorance of high complexity, and (b) use the wrong saddle in the GPI?

Can these be packaged into one unified statement?

Python's Lunch Proposal

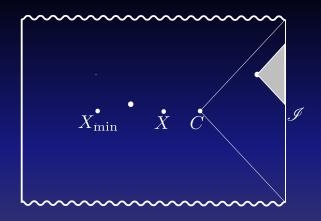
Motivated by tensor network models, Brown et al proposed that whenever there exists a nonminimal QES in the entanglement wedge, reconstruction of the region behind the nonminimal QES is exponentially complicated.



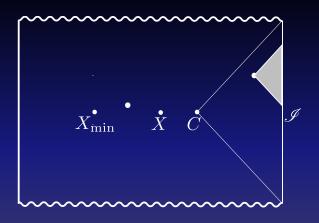


And that this complexity is given by:

$$C \propto \exp\left[\frac{1}{2}\left(S_{\mathrm{gen}}[\gamma_{\mathrm{bulge}}] - S_{\mathrm{gen}}[X]\right)\right]$$



Reconstruction up to the horizon is easy: we just causally propagate data via e.g. HKLL. Reconstruction behind X is complex by the Python's lunch.



Reconstruction up to the horizon is easy: we just causally propagate data via e.g. HKLL. Reconstruction behind X is complex by the Python's lunch. What about in-between?

The In-Between Region

In order to have a definitive connection between "coarse graining over complexity" and the nonminimal QES, we need to know more than the Python's lunch conjecture.

We need to know that it is possible to simply reconstruct *everything* up to the "appetizer".

This generically includes part of the black hole interior; in purely classical spacetimes it can include the entire interior for e.g. stellar collapse.

Recovering part of the interior

In 2017-18, Aron and I proposed that whatever reconstruction procedure you can use to obtain the data up to (and on) the event horizon can be used to reconstruct all the way up to the outermost outwards-stationary surface.



If true, this suggests a converse of the Python's lunch proposal.

Strong Python's Lunch

With Geoff and Arvin, we gave a physics-level-of-rigor proof of this in the classical regime, giving strong evidence of in favor of:

Strong Python's Lunch

Nonminimal quantum extremal surfaces are the *only* source of exponential complexity. specifically "outermost" ones that closer to the boundary than the minimal one

Current work in progress: quantum corrections (will comment on this briefly at the end, Geoff will talk about this at greater length in his talk in a month).

Table of Contents

General Idea

Toy Model: JT + Massless Scalar

Higher Dimensions

Applications

Looking Ahead

Some Terminology



A surface σ that's homologous to the asymptotic boundary is

• Marginally trapped if:

$$\theta_k \propto \frac{d\text{Area}}{d\lambda_k} = 0$$
 $\theta_\ell \propto \frac{d\text{Area}}{d\lambda_\ell} < 0$

Marginally anti-trapped if:

$$\theta_k > 0$$
 $\theta_\ell = 0$

• Extremal if

$$\theta_k = \theta_\ell = 0$$

 Outermost extremal if it lies in the "extremal wedge" of any other extremal surface.

Basic Idea NE, Wall '17, '18

Essential ingredient: reconstruction up to the event horizon is easy (just causal Lorentzian evolution of the equations of motion from data on the boundary).

Basic Idea NE, Wall '17, '18

Essential ingredient: reconstruction up to the event horizon is easy (just causal Lorentzian evolution of the equations of motion from data on the boundary).

Intuition: The outermost extremal surface is essentially what a stationary bifurcation surface looks like locally.

Basic Idea NE, Wall '17, '18

Essential ingredient: reconstruction up to the event horizon is easy (just causal Lorentzian evolution of the equations of motion from data on the boundary).

Intuition: The outermost extremal surface is essentially what a stationary bifurcation surface looks like locally. Event horizons should only fail to be stationary because of infalling matter or grav waves.

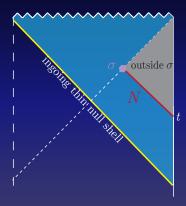
Basic Idea NE. Wall '17. '18

Essential ingredient: reconstruction up to the event horizon is easy (just causal Lorentzian evolution of the equations of motion from data on the boundary).

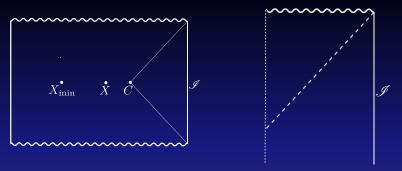
Intuition: The outermost extremal surface is essentially what a stationary bifurcation surface looks like locally. Event horizons should only fail to be stationary because of infalling matter or grav waves.

Basic Idea: We would like to show that it is possible to remove the infalling matter by turning on/off simple boundary sources (sources that create causally-propagating bulk excitations). That would then push the event horizon towards stationarity.

Illustrative Example

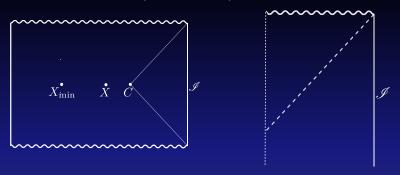


Applying to (Classical) Extremal Surfaces



Application to Extremal Surfaces: for spacetimes without a past horizon the outermost classical extremal surface is \emptyset .

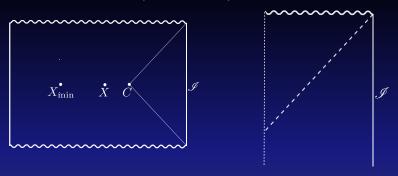
Applying to (Classical) Extremal Surfaces



Application to Extremal Surfaces: for spacetimes without a past horizon the outermost classical extremal surface is \emptyset .

For spacetimes with a nontrivial appetizer surface, want to make this work to bring the bifurcation surface to extremality. That is, we want to show that the region between C and X is simply reconstructible.

Applying to (Classical) Extremal Surfaces



Application to Extremal Surfaces: for spacetimes without a past horizon the outermost classical extremal surface is \emptyset .

For spacetimes with a nontrivial appetizer surface, want to make this work to bring the bifurcation surface to extremality. That is, we want to show that the region between C and X is simply reconstructible.

If true, this would tell us that simple (causal) reconstruction gets us all the way up to the outermost extremal surface: the appetizer is the exclusive source of high complexity.

The basic idea is as follows

1. Take to heart the proposal that the only reason for the event horizon to fail to be stationary is if something is falling across the event horizon.

The basic idea is as follows

- 1. Take to heart the proposal that the only reason for the event horizon to fail to be stationary is if something is falling across the event horizon.
- 2. "Remove" the matter by prescribing new initial data on the event horizon.

The basic idea is as follows

- 1. Take to heart the proposal that the only reason for the event horizon to fail to be stationary is if something is falling across the event horizon.
- 2. "Remove" the matter by prescribing new initial data on the event horizon.
- 3. Use HKLL to evolve this new initial data to the asymptotic boundary. This requires assuming that HKLL always works outside of event horizons. This shows that there are simple operators we can turn on/off to remove the requisite matter.

The basic idea is as follows

- 1. Take to heart the proposal that the only reason for the event horizon to fail to be stationary is if something is falling across the event horizon.
- 2. "Remove" the matter by prescribing new initial data on the event horizon.
- 3. Use HKLL to evolve this new initial data to the asymptotic boundary. This requires assuming that HKLL always works outside of event horizons. This shows that there are simple operators we can turn on/off to remove the requisite matter.
- 4. Now we have a stationary future horizon, but what we actually want is a stationary bifurcation surface.

- 1. Take to heart the proposal that the only reason for the event horizon to fail to be stationary is if something is falling across the event horizon.
- 2. "Remove" the matter by prescribing new initial data on the event horizon.
- 3. Use HKLL to evolve this new initial data to the asymptotic boundary. This requires assuming that HKLL always works outside of event horizons. This shows that there are simple operators we can turn on/off to remove the requisite matter.
- 4. Now we have a stationary future horizon, but what we actually want is a stationary bifurcation surface.
- 5. So we evolve backwards in time and repeat this procedure.

- 1. Take to heart the proposal that the only reason for the event horizon to fail to be stationary is if something is falling across the event horizon.
- 2. "Remove" the matter by prescribing new initial data on the event horizon.
- 3. Use HKLL to evolve this new initial data to the asymptotic boundary. This requires assuming that HKLL always works outside of event horizons. This shows that there are simple operators we can turn on/off to remove the requisite matter.
- 4. Now we have a stationary future horizon, but what we actually want is a stationary bifurcation surface.
- 5. So we evolve backwards in time and repeat this procedure.
- Doing this over and over again limits to the outermost extremal surface (it cannot go past it by cosmic censorship).

Table of Contents

General Idea

Toy Model: JT + Massless Scalar

Higher Dimensions

Applications

Looking Ahead

Two-Dimensional Illustration

We take JT with a minimally coupled massless scalar:

$$S = S_{JT} + \frac{1}{2} \int \sqrt{-g} (\partial \varphi)^2$$

In dilaton gravity, surfaces are points, so there's no "area"; instead, the dilaton Φ plays the role of the area in the expansion:

$$\theta_k = \partial_k \Phi$$

So, a point is extremal if $\partial_k \Phi = 0 = \partial_\ell \Phi$.

Focusing in JT+Scalar

The horizon will not be stationary, so the bifurcation surface will in general also not be stationary:

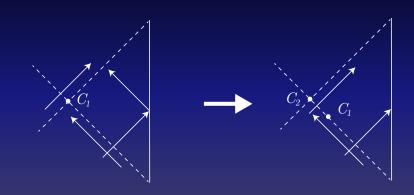
$$\partial_n \Phi|_{C_1} \neq 0$$

For any normal vector n^a .



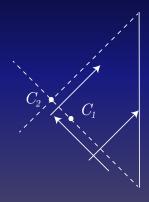
In JT+massless scalar, this can only be a result of focusing from the scalar.

We can remove the source of focusing from the future horizon by absorbing boundary conditions for the right movers.



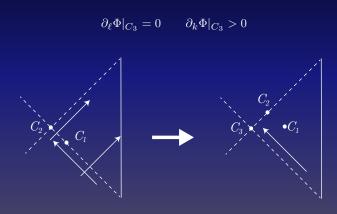
A new, deeper surface is now the bifurcation surface. This surface is marginally trapped:

$$\partial_{\ell}\Phi|_{C_2}<0 \qquad \partial_k\Phi|_{C_2}=0$$

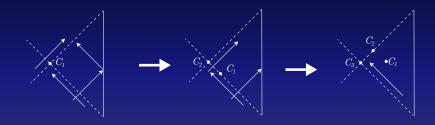


Now we evolve backwards in time, imposing boundary conditions that remove the right-movers.

This reveals a deeper bifurcation surface, which is now marginally anti-trapped:

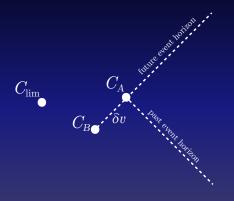


Altogether:



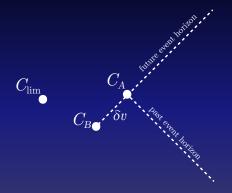
This procedure is bounded by the outermost extremal surface, but does it reach the outermost extremal surface?

Reaching the Extremal Surface

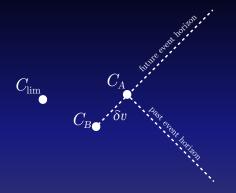


By construction, $\partial_k \Phi_{C_A} = 0$, and $\partial_\ell \Phi_{C_A} < 0$.

By construction, $\partial_k \Phi_{C_A} = 0$, and $\partial_\ell \Phi_{C_A} < 0$.



Now $\partial_{\ell}\Phi_{C_B}=0$ and $\partial_k\Phi_{C_B}>0$



In this infinitesimal neighborhood, we can approximate $ds^2 = -2dudv$, where v is an affine parameter in the k^a direction and u is an affine parameter in the ℓ^a direction.

In these coordinates we can give a lower bound to δv :

$$\delta v \geq rac{S_u T}{\partial v \partial u \Phi|_{ ext{max}}}$$
 $C_{ ext{lim}}$
 $C_{ ext{lim}}$
 $C_{ ext{lim}}$

Similarly for δu for the next iteration.

Assuming that $\partial_u \partial_v \Phi$ is bounded from below, δv and δu go to zero no slower than the expansions of the surface approach zero: so the limiting surface is indeed extremal.

Table of Contents

General Idea

Toy Model: JT + Massless Scalar

Higher Dimensions

Applications

Looking Ahead

Higher Dimensions

- Higher dimensions are harder!
- But we can still do it. The basic idea is to work perturbatively: consider some deformation δq to the event horizon.
- This perturbation needs to "open up" the lightcones so as to push the event horizon deeper in. This means, on the future horizon:

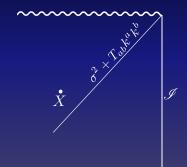
$$\delta g_{kk} \le 0$$

- The trick is to show that it is possible to find such a perturbation that satisfies the null constraint equation on the event horizon.
- We show this, but it's ugly!

Let's just do this for the future horizon.

If $\theta_k = 0$, then we are done. If $\theta_k \neq 0$, it's because of some focusing:

$$\frac{d\theta_k}{d\lambda_k} = -\frac{\theta_k^2}{D-2} - \sigma_{ab}\sigma^{ab} - T_{ab}k^a k^b$$



So there's either matter or shear sourcing the focusing. And there can be a lot of it!

So we have initial data consisting of T_{ab} and derivatives of g_{ab} on the event horizon that we can take to be part of a characteristic initial data problem:



We want to perturb the horizon initial data so as to reduce $\sigma^2 + T_{ab}k^ak^b$.

We want to perturb the horizon initial data so as to reduce $\sigma^2 + T_{ab}k^ak^b$. This means prescribing a perturbation δg that essentially dilutes the focusing, together with a matter source δT for this.

We want to perturb the horizon initial data so as to reduce $\sigma^2 + T_{ab}k^ak^b$. This means prescribing a perturbation δg that essentially dilutes the focusing, together with a matter source δT for this.

We then have to prove that this perturbed null initial data satisfies the constraints.

We want to perturb the horizon initial data so as to reduce $\sigma^2 + T_{ab}k^ak^b$. This means prescribing a perturbation δg that essentially dilutes the focusing, together with a matter source δT for this.

We then have to prove that this perturbed null initial data satisfies the constraints.

Take my word for it (or read the painful derivation in the paper): it does!

We want to perturb the horizon initial data so as to reduce $\sigma^2 + T_{ab}k^ak^b$. This means prescribing a perturbation δg that essentially dilutes the focusing, together with a matter source δT for this.

We then have to prove that this perturbed null initial data satisfies the constraints.

Take my word for it (or read the painful derivation in the paper): it does!

We then evolve this initial data back towards the asymptotic boundary using HKLL, obtaining simple sources in boundary language for preparing this perturbation.

We want to perturb the horizon initial data so as to reduce $\sigma^2 + T_{ab}k^ak^b$. This means prescribing a perturbation δg that essentially dilutes the focusing, together with a matter source δT for this.

We then have to prove that this perturbed null initial data satisfies the constraints.

Take my word for it (or read the painful derivation in the paper): it does!

We then evolve this initial data back towards the asymptotic boundary using HKLL, obtaining simple sources in boundary language for preparing this perturbation.

That allows us to then open up the lightcones on the event horizon and push it closer to the extremal surface. And we didn't have to do anything acausal for this.

So far this is for a perturbation δg . The reason that we want to work perturbatively is that this allows us to approximate the old event horizon as null in the new spacetime, so we can have well-defined characteristic initial data on it.

So far this is for a perturbation δg . The reason that we want to work perturbatively is that this allows us to approximate the old event horizon as null in the new spacetime, so we can have well-defined characteristic initial data on it.

Once we obtain the new spacetime, though, we can throw out the old one and begin from scratch, introducing a new δq that removes more focusing.

So far this is for a perturbation δg . The reason that we want to work perturbatively is that this allows us to approximate the old event horizon as null in the new spacetime, so we can have well-defined characteristic initial data on it.

Once we obtain the new spacetime, though, we can throw out the old one and begin from scratch, introducing a new δg that removes more focusing.

Using properties of the null initial data formalism, we can prove that under this approach, the bifurcation surface either becomes trivial or it asymptotes to a surface with vanishing θ_k .

So far this is for a perturbation δg . The reason that we want to work perturbatively is that this allows us to approximate the old event horizon as null in the new spacetime, so we can have well-defined characteristic initial data on it.

Once we obtain the new spacetime, though, we can throw out the old one and begin from scratch, introducing a new δg that removes more focusing.

Using properties of the null initial data formalism, we can prove that under this approach, the bifurcation surface either becomes trivial or it asymptotes to a surface with vanishing θ_k . To get the past expansion to match, we implement the same zigzag procedure as in 2D.

So far this is for a perturbation δg . The reason that we want to work perturbatively is that this allows us to approximate the old event horizon as null in the new spacetime, so we can have well-defined characteristic initial data on it.

Once we obtain the new spacetime, though, we can throw out the old one and begin from scratch, introducing a new δg that removes more focusing.

Using properties of the null initial data formalism, we can prove that under this approach, the bifurcation surface either becomes trivial or it asymptotes to a surface with vanishing θ_k . To get the past expansion to match, we implement the same zigzag procedure as in 2D.

This means that the region between the event horizon and the outermost extremal surface (which may the be empty set) is reconstructible using nothing but iterated evolution of boundary data via the equations of motion.

Table of Contents

General Idea

Toy Model: JT + Massless Scalar

Higher Dimensions

Applications

Looking Ahead

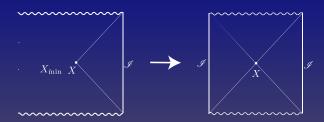
The Coarse-Grained state?

So what does the wedge of the outermost extremal surface actually correspond to?

The Coarse-Grained state?

So what does the wedge of the outermost extremal surface actually correspond to?

After executing our procedure to bring X and C together (or at least arbitrarily close to one another) , we can use a procedure from NE, Wall '17 to create a new spacetime in which the outermost extremal surface is the HRT surface:



Black Hole Uniqueness

In some cases, the procedure will only get C to limit to X, but in certain cases C and X will exactly coincide and the resulting CPT-conjugated black hole spacetime will be exactly stationary.

Theorem NE Geoff and Arvin

The causal and entanglement wedges coincide exactly if and only if the dual boundary modular Hamiltonian generates a local geometric flow with respect to a boundary Killing vector field.

Black Hole Uniqueness

In some cases, the procedure will only get C to limit to X, but in certain cases C and X will exactly coincide and the resulting CPT-conjugated black hole spacetime will be exactly stationary.

Theorem NE Geoff and Arvin

The causal and entanglement wedges coincide exactly if and only if the dual boundary modular Hamiltonian generates a local geometric flow with respect to a boundary Killing vector field.

This, coupled with our other results, gives a CFT dual to black hole uniqueness: Stationary black holes in AdS are in a one-to-one correspondence with states whose modular Hamiltonian generates an exactly local flow.

Table of Contents

General Idea

Toy Model: JT + Massless Scalar

Higher Dimensions

Applications

Looking Ahead

Quantum Corrections upcoming paper NE Geoff and Arvin

This proves everything we want in the classical regime.

What about quantum corrections? We already know that QESs don't always behave the same way that classical extremal surfaces do.

What about quantum corrections? We already know that QESs don't always behave the same way that classical extremal surfaces do. Plus, there are "obvious" examples where we don't see a QES but there is exponential complexity!

What about quantum corrections? We already know that QESs don't always behave the same way that classical extremal surfaces do. Plus, there are "obvious" examples where we don't see a QES but there is exponential complexity!

For example, consider a a non-evaporating black hole not coupled to a reservoir at late times (e.g. after it settles down). Decoding the interior outgoing modes should be hard – the blueshift as we evolve backwards results in a transplanckian problem. We expect exponential complexity, but at first glance it doesn't look like there's a QES in this spacetime.

What about quantum corrections? We already know that QESs don't always behave the same way that classical extremal surfaces do. Plus, there are "obvious" examples where we don't see a QES but there is exponential complexity!

For example, consider a a non-evaporating black hole not coupled to a reservoir at late times (e.g. after it settles down). Decoding the interior outgoing modes should be hard – the blueshift as we evolve backwards results in a transplanckian problem. We expect exponential complexity, but at first glance it doesn't look like there's a QES in this spacetime.

Our upcoming work shows (and Geoff will discuss in more detail next month in this seminar series) that a highly nonclassical QES exists in the maximally mixed state (in our code subspace).

 We want to understand the omission in Hawking's calculation in the language of Hawking's calculation.

- We want to understand the omission in Hawking's calculation in the language of Hawking's calculation.
- One way of going about this is to understand how omission of complex data (pre-2019 perspective) and omission of relevant saddles in the GPI (post-2019 perspective) agree.

- We want to understand the omission in Hawking's calculation in the language of Hawking's calculation.
- One way of going about this is to understand how omission of complex data (pre-2019 perspective) and omission of relevant saddles in the GPI (post-2019 perspective) agree.
- The Python's lunch prescription proposes that everything behind a nonminimal QES is highly complex.

- We want to understand the omission in Hawking's calculation in the language of Hawking's calculation.
- One way of going about this is to understand how omission of complex data (pre-2019 perspective) and omission of relevant saddles in the GPI (post-2019 perspective) agree.
- The Python's lunch prescription proposes that everything behind a nonminimal QES is highly complex.
- The zigzag procedure shows that everything outside of a nonminimal classical extremal surface is highly complex; we will give fairly conclusive evidence (though not a proof) that this remains true under inclusion of quantum corrections in an upcoming paper.