The black hole information paradox and the smoothness of the horizon

Kyriakos Papadodimas
CERN and University of Groningen

International Workshop on Strings, Black Holes and Quantum Information, TFC 2015

Outline:

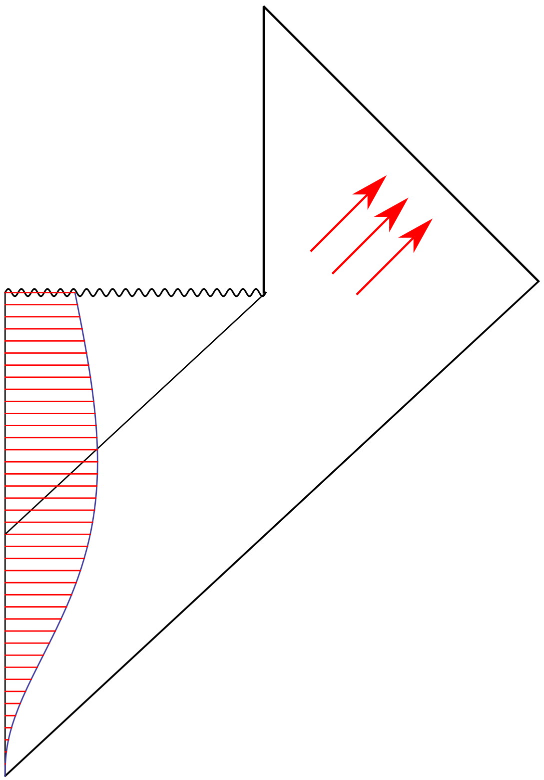
1. Review of information paradox

2. Relation to smoothness of the horizon

3. The information paradox in AdS/CFT and recent arguments against the existence of the black hole interior

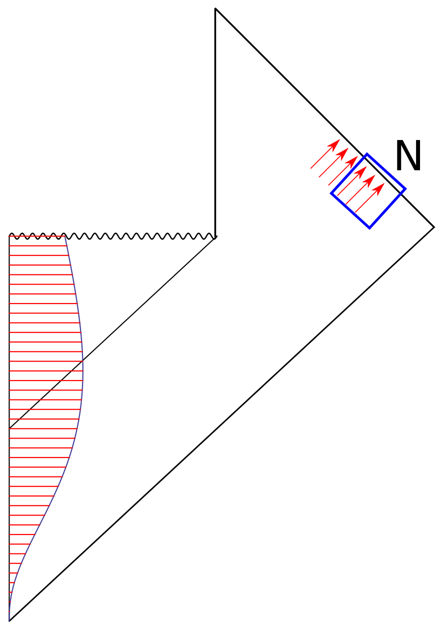
Tomorrow: A holographic reconstruction of the black hole interior
(based on work with Suvrat Raju)

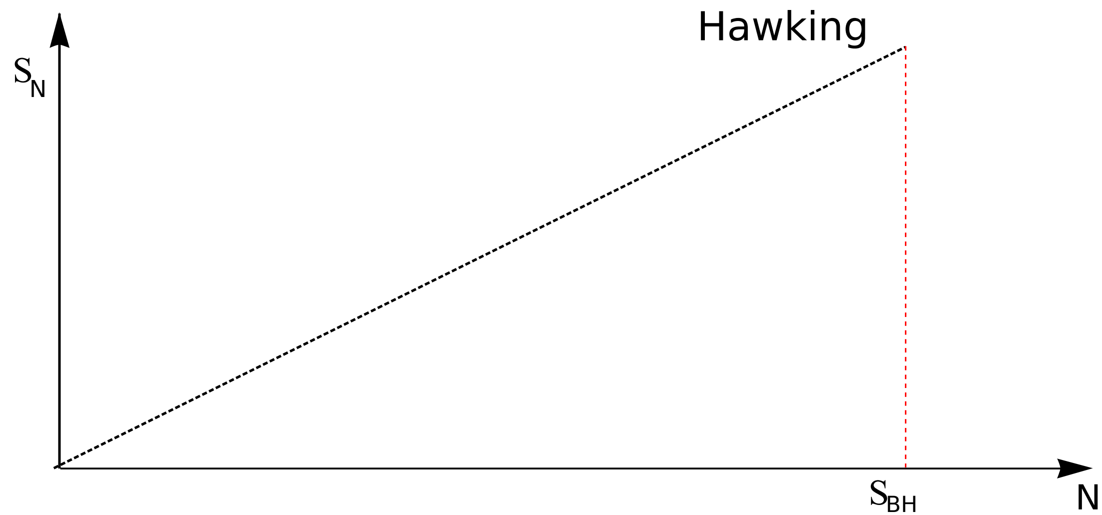
Basic formulation of the information paradox



$$|\Psi\rangle \quad \Rightarrow \quad \rho_{\text{thermal}}?$$

Inconsistent with unitary evolution in Quantum Mechanics





Consider the **entanglement entropy** of the reduced density matrix ρ_N of the first N Hawking particles

$$S_N = -\mathrm{Tr}(\rho_N \log \rho_N)$$

If BH evaporation is unitary, S_N must eventually go to zero

Possibilities:

1. Information loss

2. Remnants

3. Information encoded in outoging radiation in small correlations between particles

(similar to what happens when we burn a piece of paper)

According to Hawking's computation $\Rightarrow$ particles are uncorrelated. Is there a "mistake" in the computation of Hawking?

Hawking computation is a semiclassical approximation

Perturbative corrections

Exponentially small corrections from other "histories in path integral"

Transplanckian problem/blueshift/chaos

...

Can small corrections to Hawking's computation resolve the information paradox?

Hawking computation is a semiclassical approximation

Perturbative corrections

Exponentially small corrections from other "histories in path integral"

Transplanckian problem/blueshift/chaos

...

Can small corrections to Hawking's computation resolve the information paradox?

Two clarifications:

1) Small corrections to: simple correlation functions in EFT

2) Resolving the paradox is not the same thing as being able to compute the exact BH S-matrix

Claim: Hawking's computation is reliable for **simple** observables in effective field theory (low-point functions)

$$\langle \phi(x_1)\phi(x_2)...\phi(x_n) \rangle \qquad n \ll S_{BH}$$

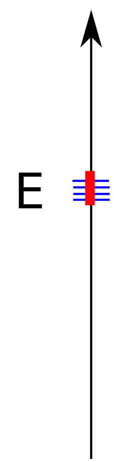
and separations of points not too small/too large

Unitarity can be restored in Hawking evaporation, at the price of introducing **exponentially small** (of order $e^{-S_{BH}}$) corrections to these observables

This claim relies on a basic property of Quantum Statistical Mechanics:

"in large systems, typical pure states look like mixed states when probed by simple observables"

Pure states vs mixed states



Consider Hilbert space $\mathcal{H}_E$ spanned by eigenstates $E_i \in (E, E + \Delta E)$, with $\mathrm{dim}\mathcal{H}_E = D = e^S \gg 1.$

Typical pure state $|\Psi\rangle = \sum_{i=1}^{D} c_i |E_i\rangle \qquad \{c_i\} \in \mathbf{S}^{2D-1}$

Microcanonical ensemble, mixed state $\rho_{\mathrm{micro}} = \frac{1}{D}\mathbf{I_E}$

Define $\langle A \rangle_{\text{micro}} = \mathrm{Tr}(\rho_{\text{micro}} A)$

We also define the average over pure states in $\mathcal{H}_E$

$$\overline{\langle \Psi | A | \Psi \rangle} \equiv \int [d\mu_\Psi] \langle \Psi | A | \Psi \rangle$$

where $[d\mu_\Psi]$ is the Haar measure. Then **for any** observable A we have

$$\overline{\langle \Psi | A | \Psi \rangle} = \langle A \rangle_{\text{micro}}$$

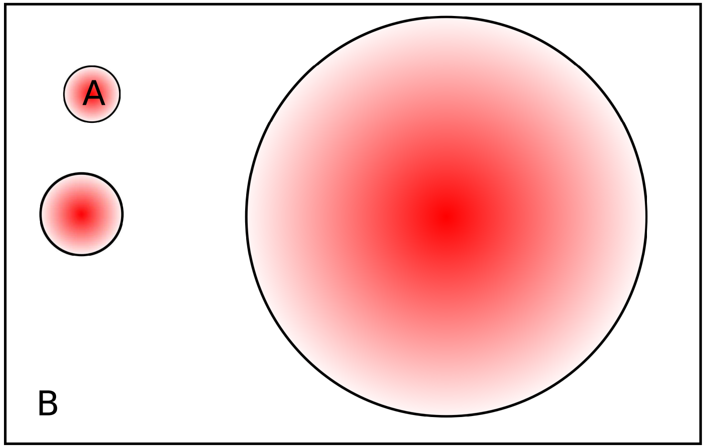
and

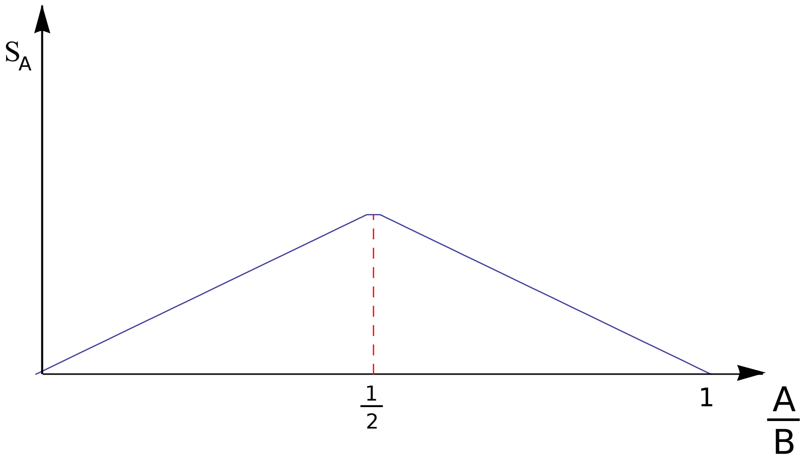
$$\text{variance} \equiv \overline{(\langle \Psi | A | \Psi \rangle^2)} - (\overline{\langle \Psi | A | \Psi \rangle})^2 = \frac{1}{e^S + 1} \left(\langle A^2 \rangle_{\text{micro}} - (\langle A \rangle_{\text{micro}})^2 \right)$$

reasonable observables have the same expectation value in most pure states, up to exponentially small corrections [Lloyd, Balasubramanian et al.] .

(relevance for Fuzzball program)

Entanglement entropy of subsystem [Page]





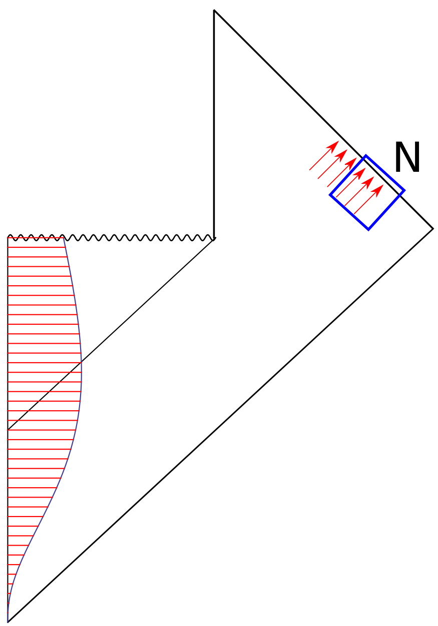
Large system B in typical pure state $|\Psi\rangle$

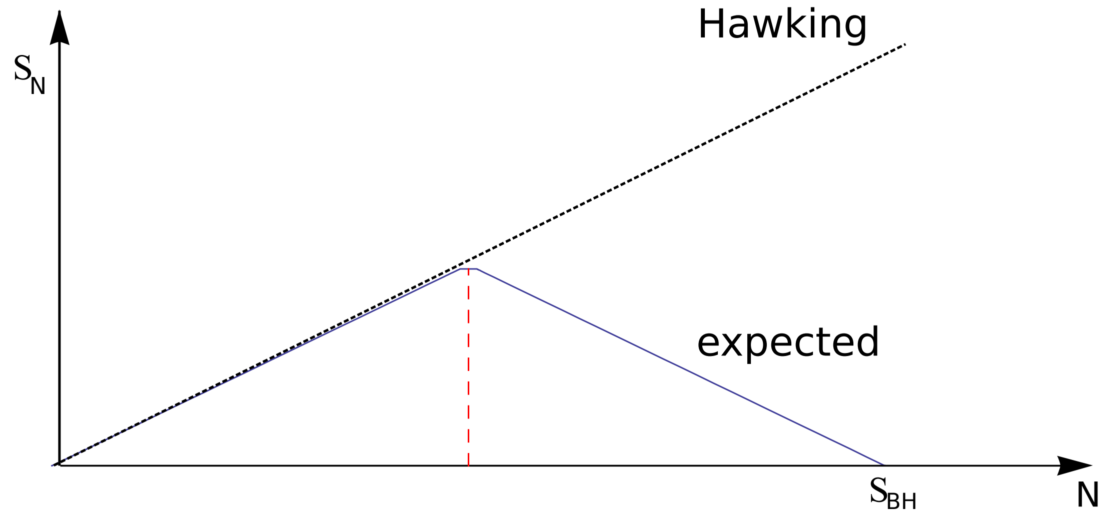
Subsystem A: if A is small, reduced density matrix is exponentially close to maximally mixed and its EE is proportional to size of A. This breaks down once $|A| > \frac{|B|}{2}$

Toy model: spin chain (off-diagonal terms, initially exponentially small, become important when $|A| > \frac{|B|}{2}$)

Resolution of basic version of information paradox

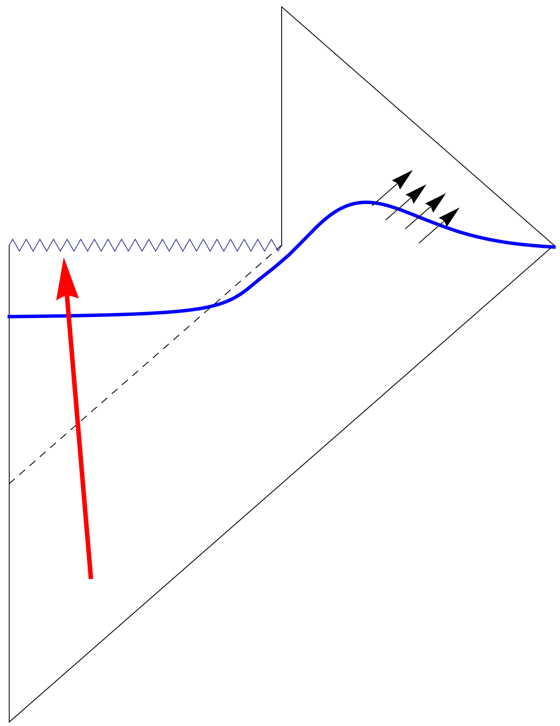
Exponentially small corrections to Hawking's computation can restore unitarity





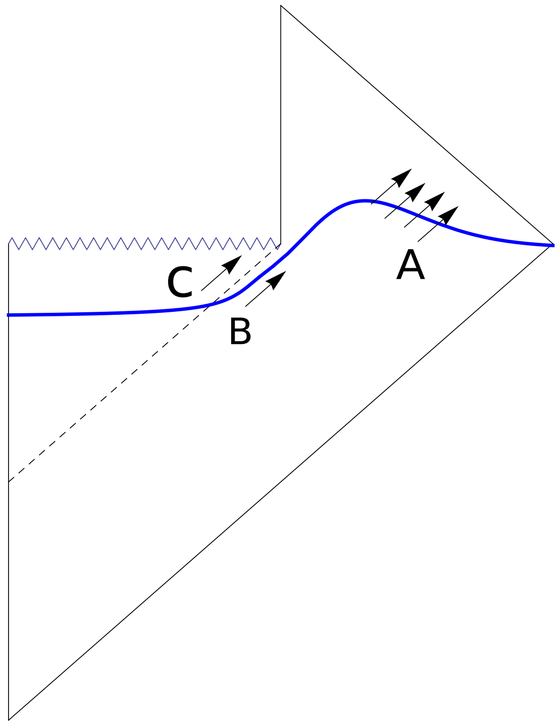
$$\langle \Psi | A | \Psi \rangle = \mathrm{Tr}[\rho A] + \mathcal{O}(e^{-S})$$

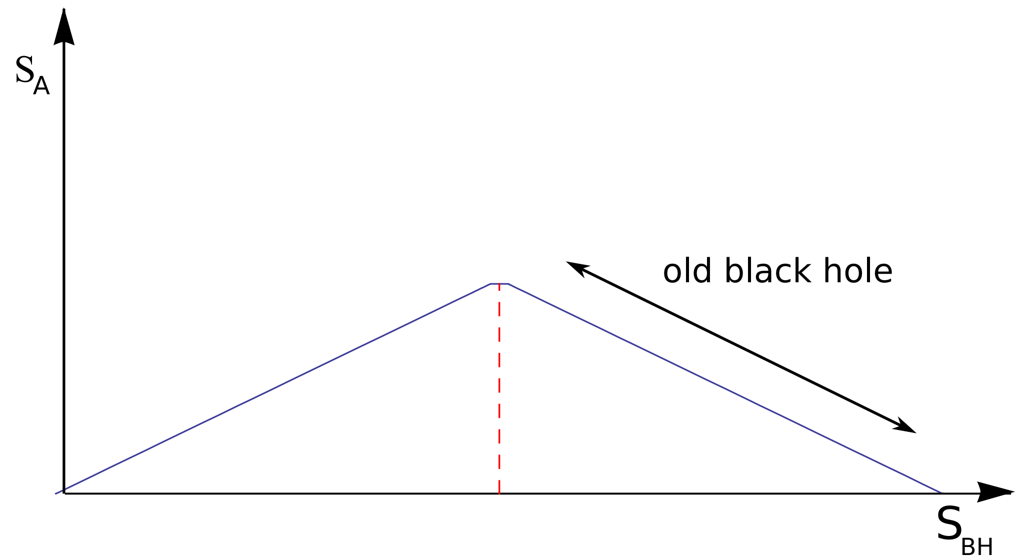
Information Paradox, "refined" formulation: interior+exterior



Quantum cloning on nice slices

Mathur/AMPS version of the information paradox





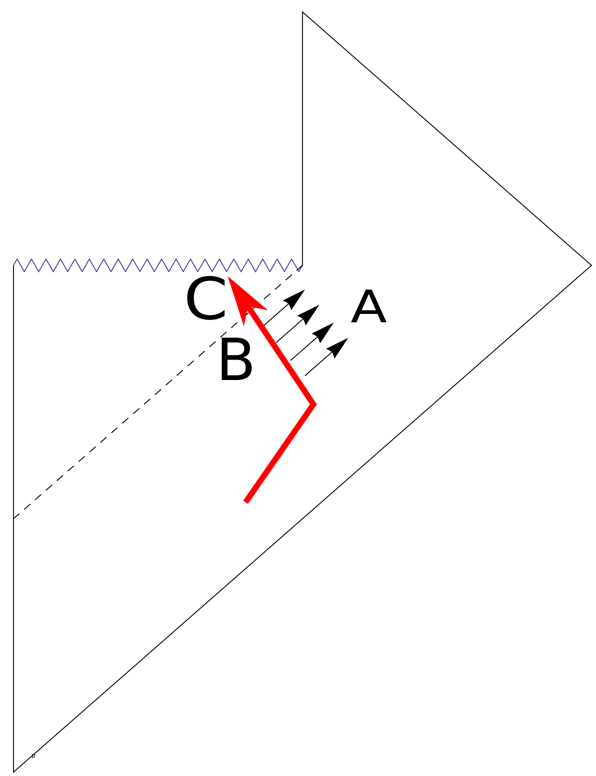
Strong subadditivity paradox [Mathur], [Almheiri, Marolf, Polchinski, Sully (AMPS)]

$$S_{AB} + S_{BC} \geq S_A + S_C$$

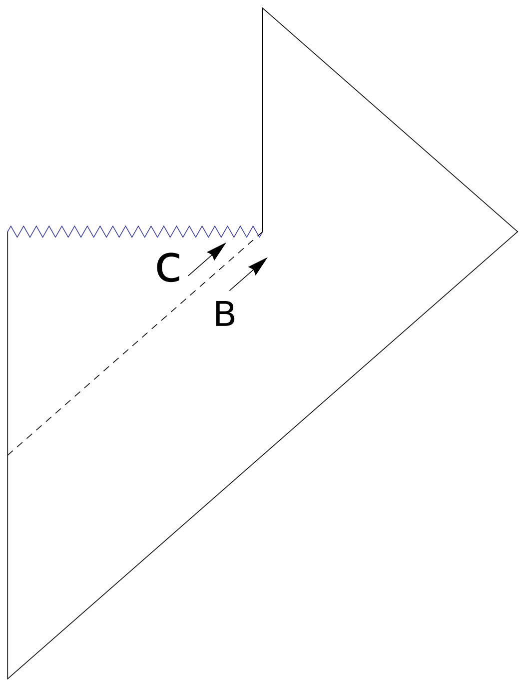
Mathur's theorem (2009): "small corrections cannot resolve the paradox" (?)

Distilling the scrambled qubit

Observer measures early Hawking radiation A to extract the "scrabmled qubit" $\widetilde{B}$ in A, with which B is entangled. Then jumps into black hole. Is C still entangled with B?
If yes, cloning. If no, excitation at horizon.



Quantum Chaos vs Entanglement

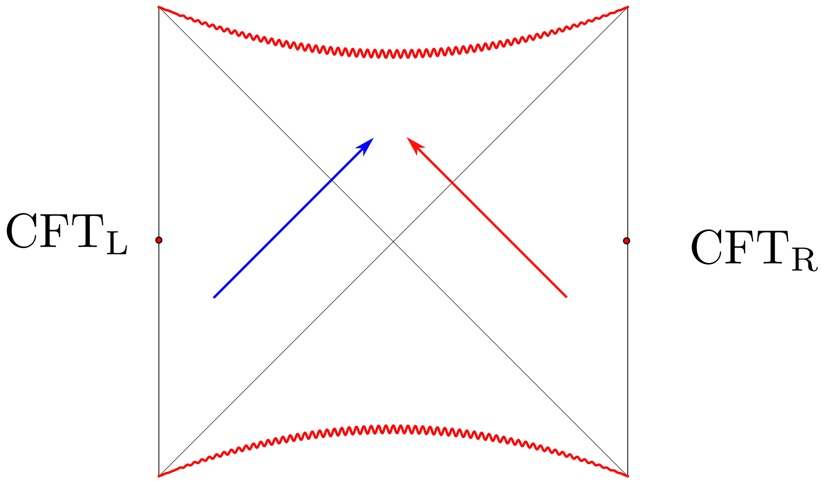


Smooth horizon requires **specific** pattern of entanglement between field operators at B and C

Fragile under perturbations due to chaotic nature of system

Hard to imagine how **typical states** will end up with the correct, **specific** entanglement needed for smoothness

Eternal black hole [Maldacena, Maldacena-Susskind]



$$H_{\text{total}} = H_L + H_R$$

$$|\mathrm{TFD}\rangle = \frac{1}{\sqrt{Z}} \sum_i e^{-\frac{\beta E_i}{2}} |E_i\rangle_L \otimes |E_i\rangle_R$$

$\exists$ general agreement that eternal BH has smooth interior.

Smoothness of the horizon depends on correct entanglement, which follows from

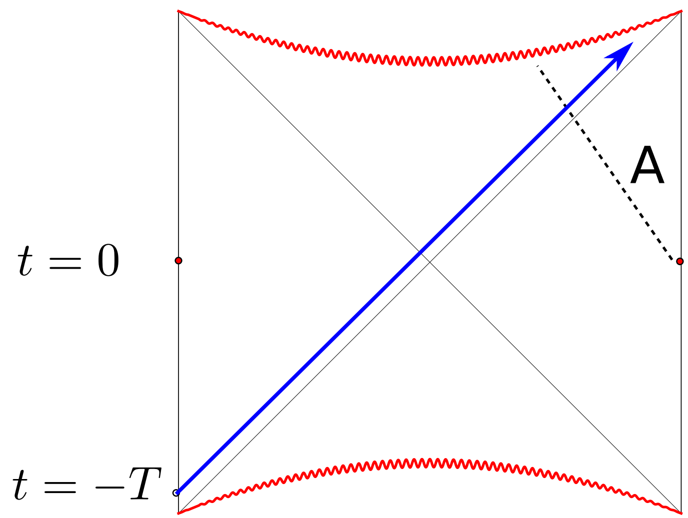
$$\langle \mathrm{TFD}| \, \mathcal{O}_L(t_L = 0) \; \mathcal{O}_R(t_R = 0) |\mathrm{TFD}\rangle \sim O(1)$$

On the other hand, for a typical state $|\Psi\rangle$ (with same amount, but different pattern of entanglement) we find using ETH

$$\langle \Psi| \, \mathcal{O}_L(t_L = 0) \; \mathcal{O}_R(t_R = 0) |\Psi\rangle \sim O(e^{-S})$$

Do they have a smooth interior?

[Shenker, Stanford]



We start with $|\mathrm{TFD}\rangle$ and perturb it by a small operator (energy of $O(1)$) at time $t_L = -T$).

For small T, effect on infalling observer A is small. But center of mass collision energy grows exponentially with T

For $T > \log S$ (scrambling time) we can no longer ignore backreaction.

The "correct entanglement" of the TFD disrupted even by small perturbations due to chaos.

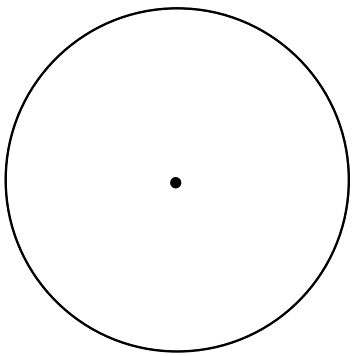
Summary:

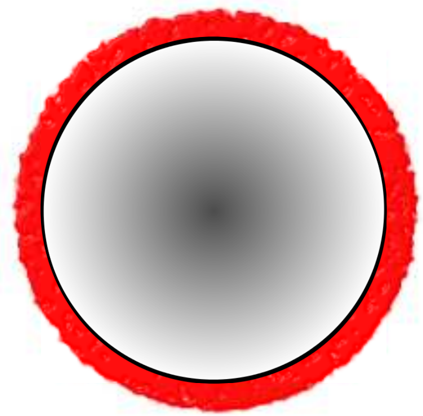
1. Information paradox from the point of view of asymptotic observer:

natural, robust resolution, consistent with generic expectations from quantum statistical mechanics
(pure $\approx$ mixed, exponentially small corrections)

2. Preserving smoothness of the horizon: more challenging. Seems to contradict generic expectations from quantum statistical mechanics

Is the horizon smooth?





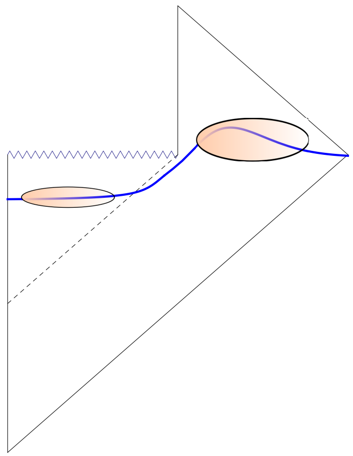
The black hole information paradox becomes more sharp if we assume that the black hole horizon is smooth

Proposals to resolve info paradox by giving up smoothness of interior: fuzzball, firewall,...

according to which an infalling observer feels deviations from GR when crossing the horizon

Black hole complementarity

The Hilbert space of Quantum Gravity does not factorize in interior $\times$ exterior (locality is approximate)



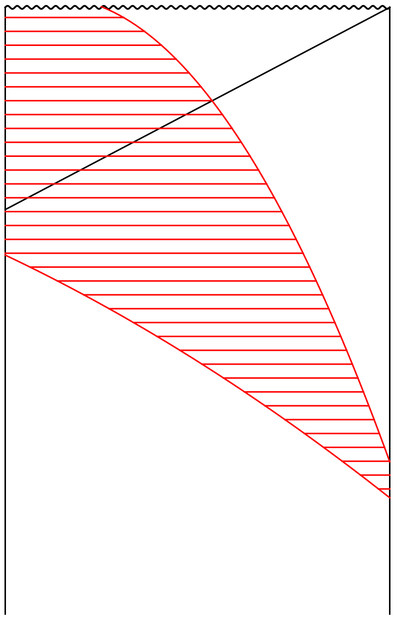
then quantum cloning problem and subadditivity Mathur/AMPS problem resolved

Is this consistent with locality in effective field theory?

What is the mathematical framework?

Some progress [K.P and S.Raju]$\Rightarrow$ New ingredient: “State-dependence”

The information paradox in AdS/CFT



Consider the $\mathcal{N} = 4$ SYM on $S^3 \times \mathrm{time}$, at large N, large λ. At some time t_0 we inject energy of order $O(N^2)$.

After some time the state seems to thermalize and equilibrate. Of course, the CFT is unitary, so we still have a pure state (some Quark-Gluon-Plasma microstate).
In the bulk we end with a big black hole in AdS.

Consider the analogue of the Hawking computation on the background of the AdS black hole.

It predicts that the bulk field is in the AdS/Hartle-Hawking state.

Bulk-bulk correlators factorize to 2-point functions, which are thermal.

Using the AdS/CFT correspondence and that $\phi \leftrightarrow \mathcal{O}$, this seems to predict that at late times, CFT correlators of the form

$$\langle\Psi|\mathcal{O}(x_1)...\mathcal{O}(x_n)|\Psi\rangle$$

factorize to products of thermal 2-point functions.

How is this possible if the state is pure? (where is the information of the state?)

From general arguments of quantum statistical mechanics we expect (for $n \ll N$) that

$$\langle\Psi|\mathcal{O}(x_1)...\mathcal{O}(x_n)|\Psi\rangle = \frac{1}{Z}\mathrm{Tr}\left[e^{-\beta H}\mathcal{O}(x_1)...\mathcal{O}(x_n)\right] + \text{small corrections}$$

which is precisely what the bulk AdS Hawking computation predicts.

When measuring simple observables (low-point functions) typical pure states and thermal states give almost the same results.

However, for complicated correlators with $n \sim S_{BH} \sim N^2$ insertions we expect to see large deviations. These complicated correlators encode the exact information of the pure state. These may not necessarily correspond to some geometric computation in the bulk.

So there is no paradox, if we keep in mind that EFT is reliable only for certain observables.

The black hole interior in AdS/CFT

The hard part of the information paradox is to reconcile unitarity with the smoothness of the horizon.

Does a big black hole in AdS have a smooth interior and can the CFT describe it?

In recent years some progress has been made:

i) Smoothness of horizon for AdS black holes relies on existence of certain operators in the CFT, with specific properties.

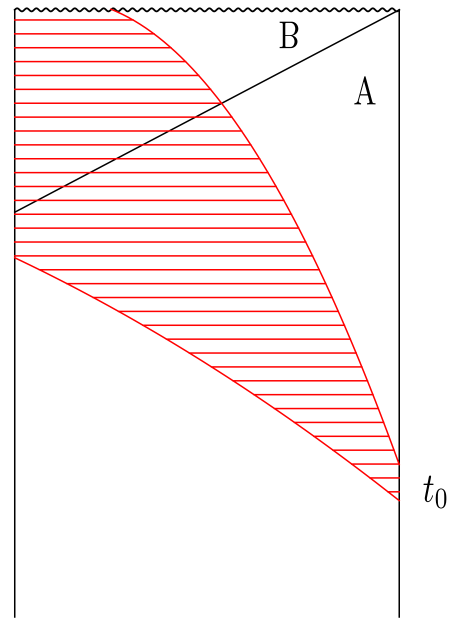
ii) Arguments against the existence of these operators (and hence in favor of "firewalls") [AMPSS: AMPS+ Stanford, Bousso, Harlow, Marolf-Polchinski]

iii) Concrete proposal on how to identifiy these operators and how these arguments can be evaded [KP, Suvrat Raju]

The black hole interior in AdS/CFT

1. The arguments of AMPSS, MP against the existence of a smooth interior for a big black hole in AdS, provide -in my opinion- the most precise formulation of the information paradox, within a non-perturbatively defined theory of quantum gravity. Any attempt to resolve the information/firewall paradox should address these arguments for the big black hole in AdS.

2. Independent of the motivation from the information/firewall paradox, developing a concrete formalism for the reconstruction of the black hole interior in AdS/CFT is an outstanding open problem, which may also have other usefel applications (BH singularity, relevance of BH interior for thermalization...)



Local observables in AdS/CFT?

Large N factorization allows us to write local* observables in empty AdS as non-local observables in CFT (smeared operators)

$$\phi_{\mathrm{CFT}}(t,\Omega,z) = \sum_m \int_{\omega>0} d\omega \ \left(\mathcal{O}_{\omega,m} \ f_{\omega,m}(t,\Omega,z) + \mathrm{h.c.}\right)$$

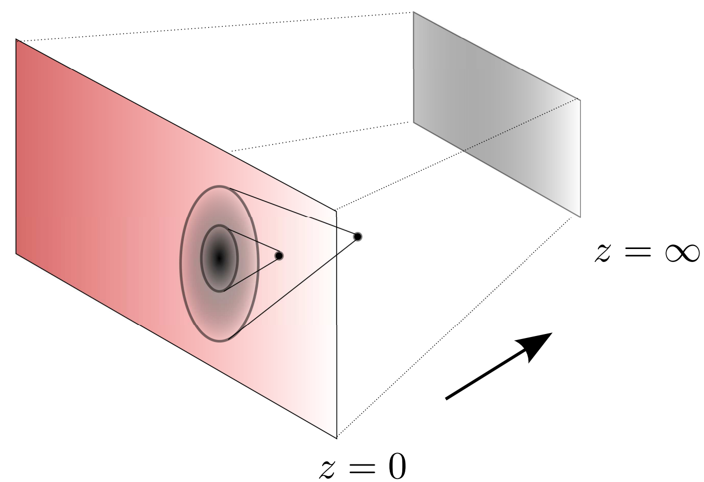
where ϕ_{CFT} obeys EOMs in AdS, and $[\phi_{\mathrm{CFT}}(P_1), \phi_{\mathrm{CFT}}(P_2)] = 0$, if points P_1, P_2 spacelike with respect to AdS metric

(based on earlier works: Banks, Douglas, Horowitz, Martinec, Bena, Balasubramanian, Giddings, Lawrence, Kraus, Trivedi, Susskind, Freivogel Hamilton, Kabat, Lifschytz, Lowe, Heemskerk, Marolf, Polchinski, Sully...)

---

* Locality is approximate:

1. (Probably) true in $1/N$ perturbation theory
2. Unlikely that $[\phi_{\mathrm{CFT}}(P_1), \phi_{\mathrm{CFT}}(P_2)] = 0$ to e^{-N^2} accuracy
3. Locality may break down for high-point functions (perhaps no bulk spacetime interpretation)



$$\phi_{\rm CFT}(t,\Omega,z) = \int dt' d\vec{x}' \; K(\,t,\Omega,z\;;\;t',\Omega')\mathcal{O}(t',\Omega')$$

where K is some known kernel (smearing function).

Subtleties: 1/N expansion, gauge invariance....

Black Hole Exterior

Consider big black hole in AdS. Expectation from bulk effective field theory (EFT) for a free scalar

$$\phi(t, r, \Omega) = \int_0^\infty d\omega \sum_{lm} b_{\omega lm}\, e^{-i\omega t} f_{\omega,l}(r) Y_{lm}(\Omega) + \mathrm{h.c.}$$

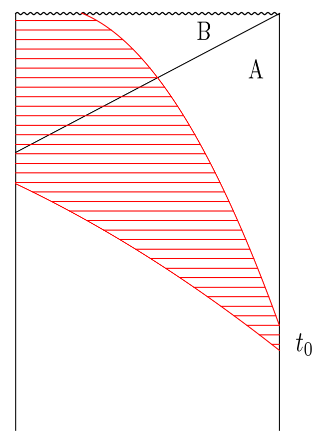
where (dropping l, m indices) we have

$$[b_\omega, b^\dagger_{\omega'}] = \delta(\omega - \omega') \qquad [H, b_\omega] = -\omega b_\omega$$

and

$$\langle b^\dagger_\omega b_\omega \rangle \sim \frac{e^{-\beta\omega}}{1 - e^{-\beta\omega}}$$

How do we reconstruct this from the CFT?



In typical QGP pure state $|\Psi\rangle$ (energy $O(N^2)$), single trace correlators factorize at large N

$$\langle\Psi|\mathcal{O}(x_1)...\mathcal{O}(x_n)|\Psi\rangle = \langle\Psi|\mathcal{O}(x_1)\mathcal{O}(x_2)|\Psi\rangle...\langle\Psi|\mathcal{O}(x_{n-1})\mathcal{O}(x_n)|\Psi\rangle + ...$$

The 2-point function in which they factorize is the thermal 2-point function, which is hard to compute, but obeys KMS condition

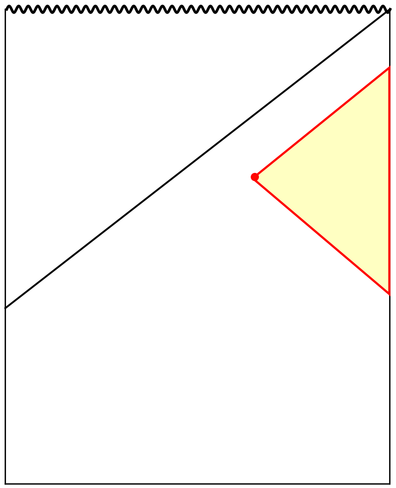
$$G_\beta(-\omega) = e^{-\beta\omega}G_\beta(\omega)$$

Consider single-trace operator $\mathcal{O}$ in CFT, dual to bulk field ϕ. Define Fourier modes

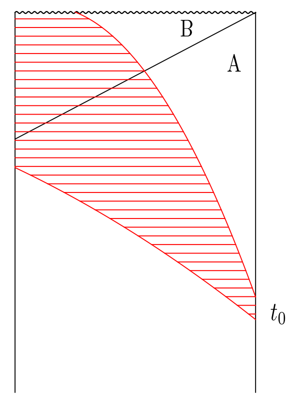
$$\mathcal{O}_{\omega lm} = \int dt d\Omega \; \mathcal{O}(t,\Omega)\, e^{i\omega t}\, Y^*_{lm}(\Omega)$$

then we identify

$$b_{\omega lm} \;\propto\; \mathcal{O}_{\omega lm}$$



(interesting subtleties about large l modes, gauge invariance, 1/N corrections etc.)



Local bulk field outside horizon of AdS black hole

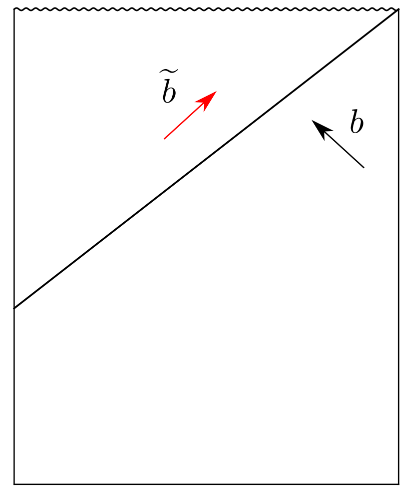
$$\phi_{\mathrm{CFT}}(t,\Omega,z) = \sum_m \int_0^\infty d\omega\, \mathcal{O}_{\omega,m}\, f^{\beta}_{\omega,m}(t,\Omega,z) + \mathrm{h.c.}$$

At large N (and late times) the correlators

$$\langle\Psi|\phi_{\mathrm{CFT}}(t_1,\Omega_1,z_1)...\phi_{\mathrm{CFT}}(t_n,\Omega_n,z_n)|\Psi\rangle$$

reproduce those of semiclassical QFT on the BH background (in AdS-Hartle-Hawking state).

Need for interior modes

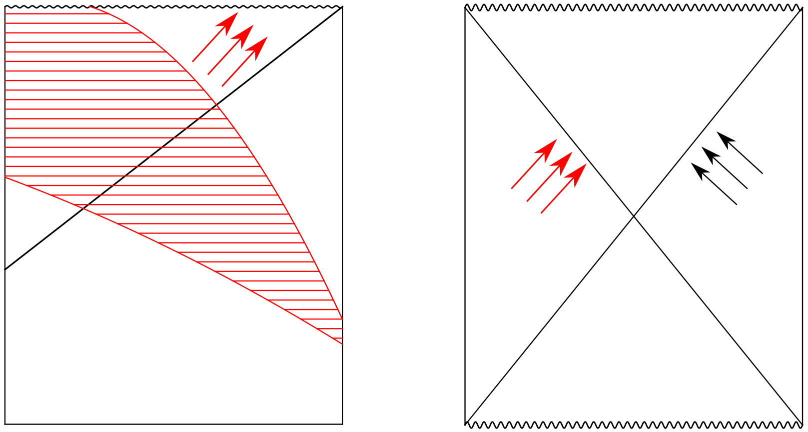


EFT: we need a new set of modes $\widetilde{b}$ which commute with b, and which are entangled with b.

We identified b with modes of $\mathcal{O}$ in CFT.

Central question:

Which CFT operators correspond to $\widetilde{b}$? — whatever these operators are, we denote them as $\widetilde{\mathcal{O}}$.



For smooth horizon we expect

$$\phi_{\mathrm{CFT}}(t,\Omega,z) = \sum_m \int_0^\infty d\omega \Big[\mathcal{O}_{\omega,m}\, e^{-i\omega t} Y_m(\Omega) g^{(1)}_{\omega,m}(z) + \mathrm{h.c.}$$

$$+ \boxed{\widetilde{\mathcal{O}}_{\omega,m}}\, e^{-i\omega t}\, Y_m(\Omega)\, g^{(2)}_{\omega,m}(z) + \mathrm{h.c.} \Big]$$

where the modes $\widetilde{\mathcal{O}}_{\omega,m}$ must satisfy certain conditions

The $\widetilde{\mathcal{O}}_{\omega,m}$'s (*mirror* or *tilde* operators) must obey the following conditions, in order to have smooth interior:

1. For every $\mathcal{O}$ there is a $\widetilde{\mathcal{O}}$
2. The algebra of $\widetilde{\mathcal{O}}$'s is isomorphic to that of the $\mathcal{O}$'s
3. The $\widetilde{\mathcal{O}}$'s commute with the $\mathcal{O}$'s
4. The $\widetilde{\mathcal{O}}$'s are "correctly entangled" with the $\mathcal{O}$'s

Equivalently:

Correlators of all these operators on $|\Psi\rangle$ must reproduce (at large N) those of the thermofield-double state

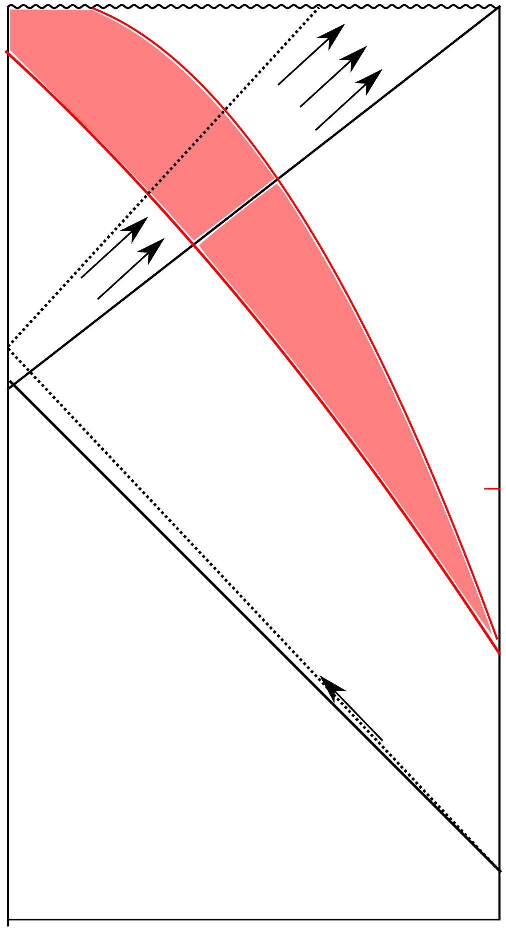
$$|TFD\rangle = \sum_i \frac{e^{-\beta E_i/2}}{\sqrt{Z}} |E_i, \widetilde{E}_i\rangle$$

$$\langle\Psi|\mathcal{O}(t_1)...\widetilde{\mathcal{O}}(t_k)..\mathcal{O}(t_n)|\Psi\rangle \approx \frac{1}{Z}\mathrm{Tr}\big[\mathcal{O}(t_1)...\mathcal{O}(t_n)\mathcal{O}(t_k + i\frac{\beta}{2})...\mathcal{O}(t_m + i\frac{\beta}{2})\big]$$

Main Question: Does the CFT contain the operators $\widetilde{\mathcal{O}}$ with the desired properties?

If so, then the CFT can describe the interior of the black hole and we have free infall through the horizon.

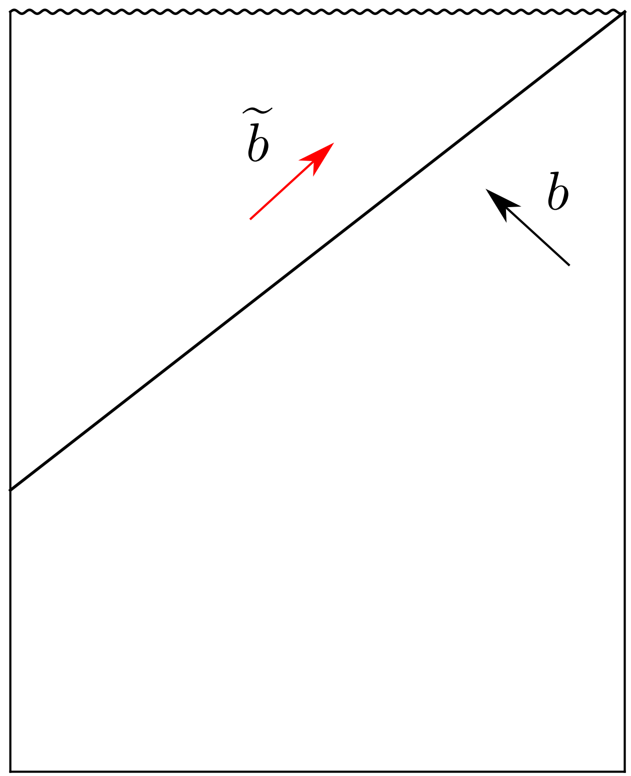
How do we find these operators?



Using bulk EFT evolution to find the $\widetilde{\mathcal{O}}$? $\Rightarrow$ Trans-planckian problem...(?)

Typical states vs states formed by collapse

Counting argument, against existence of $\widetilde{b}$ operators in CFT (AMPSS)



$$[b, b^\dagger] = 1$$
$$[H, b^\dagger] = \omega b^\dagger$$

$$[\widetilde{b}, \widetilde{b}^\dagger] = 1$$
$$[H, \widetilde{b}^\dagger] = -\omega \widetilde{b}^\dagger$$

The required algebra between $\widetilde{b}, \widetilde{b}^\dagger, H$ is inconsistent with spectrum of states in CFT

$$[\widetilde{b}, \widetilde{b}^\dagger] = 1 \qquad \Rightarrow \qquad \widetilde{b}^\dagger = \text{“creation operator”}$$

$\Rightarrow \widetilde{b}^\dagger$ should not annihilate (typical) states of the CFT $(*)$.

On the other hand

$$[H, \widetilde{b}^\dagger] = -\omega \widetilde{b}^\dagger$$

implies that $\widetilde{b}^\dagger$ lowers the energy so it maps CFT states of energy E to $E - \omega$.

But in CFT, we have $S(E) > S(E - \omega)$.

$\Rightarrow$ if $\widetilde{b}^\dagger$ is an ordinary linear operator, it must have a nontrivial kernel.

Inconsistent with statement (*).

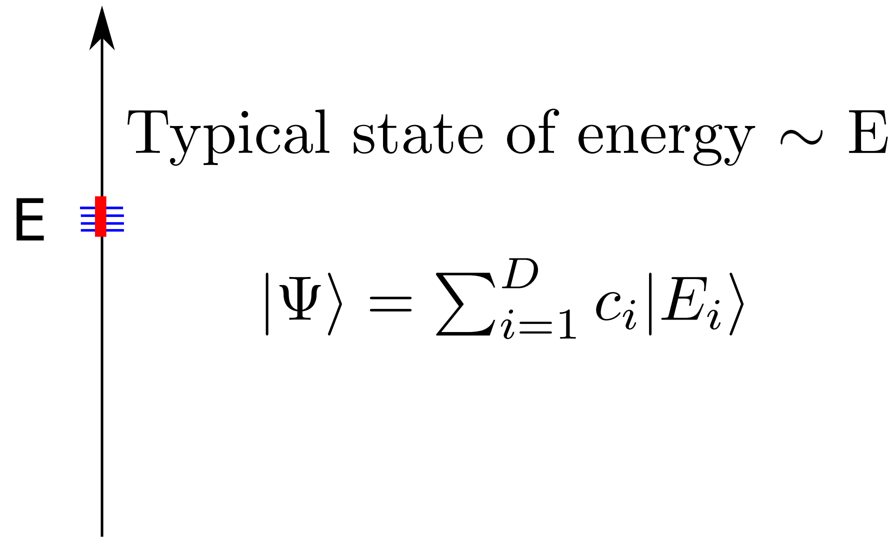
$\Rightarrow$ The CFT does not contain $\widetilde{b}$ operators and cannot describe the BH interior (?)

Previous counting argument can be made somewhat more precise (K.P and S.Raju)

Related argument $Tr[N_a] \neq 0$ (Bousso, Marolf-Polchinski)

Additional general argument: if $\widetilde{b}$ is a fixed, linear operator, it is hard to understand how $\mathbf{typical}$ CFT states can have the $\mathbf{particular, special}$ entanglement between $b, \widetilde{b}$ needed for smooth interior

Rotating phases



Marolf and Polchinski (2015): If typical states have smooth interior then $\Rightarrow$ violation of Born rule:

There are states $|\Psi_1\rangle, |\Psi_2\rangle$ with "orthogonal physical interpretation" but with state-vectors which are almost parallel

$$|\langle\Psi_1|\Psi_2\rangle| \approx 1$$

In the smooth AdS/Hartle-Hawking state bulk modes are entangled as

$$|\mathrm{HH}\rangle = \frac{1}{\sqrt{Z}} \sum_{n=0}^{\infty} e^{-\frac{\beta\omega n}{2}} |n, \widetilde{n}\rangle$$

Horizon is smooth because state is annihilated by “infalling modes”

$$a = b - e^{-\frac{\beta\omega}{2}} \widetilde{b}^{\dagger} \qquad a|\mathrm{HH}\rangle = 0$$

Consider the number operator $N_b = b^{\dagger} b$ and the new state

$$|\mathrm{HH}'\rangle \equiv e^{i\theta N_b} |\mathrm{HH}\rangle = \frac{1}{\sqrt{Z}} \sum_{n=0}^{\infty} e^{i\theta n} e^{-\frac{\beta\omega n}{2}} |n, \widetilde{n}\rangle$$

This is no longer annihilated by $a \Rightarrow$ EFT predicts that in $|\mathrm{HH}'\rangle$ the infalling observer should detect an excitation for the mode ω.
States $|\mathrm{HH}\rangle$ and $|\mathrm{HH}'\rangle$ have different physical interpretation and corresponding state-vectors have small overlap. So all seems OK in bulk EFT.

CFT analysis

Consider the Hilbert space $\mathcal{H}_E$ of states of energy approximately E

$$|\Psi\rangle = \sum_i c_i |E_i\rangle$$

For $N_\omega = \mathcal{O}_\omega^\dagger \mathcal{O}_\omega$ we have in the large N limit that $[H_{CFT}, N_\omega] = 0$.

Hence the unitary

$$U = e^{i\theta N_\omega}$$

does not change the energy of the state and maps $\mathcal{H}_E$ mostly into itself.
If we combine this statement with the other previous assumptions that:

1. Typical states in $\mathcal{H}_E$ have smooth, unexcited horizon

2. U maps unexcited to excited states

then we conclude that there must be state-vectors in $\mathcal{H}_E$ which are very close, but which have "orthogonal" physical interpretation.

This seems to violate the Born rule, and MP argue that there is an experiment where an observer can detect this violation.

Summary

1. Exponentially small corrections can change Hawking radiation from mixed to pure

2. Seems hard to reconcile this with a smooth horizon

3. Reviewed recent arguments against the existence of the black hole interior in AdS/CFT

Tomorrow: a specific proposal for reconstructing the interior operators

THANK YOU