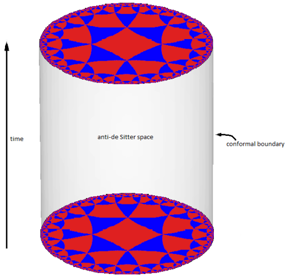
Quantum Mechanics and the Black Hole Horizon

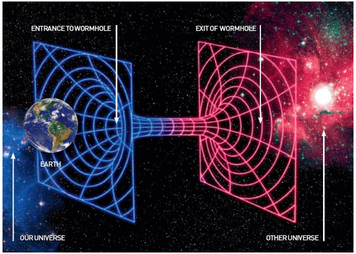
Kyriakos Papadodimas

CERN and University of Groningen

9th Aegean summer school: Einstein's theory of gravity and its modifications

Space-time, gravity and locality





String theory, AdS/CFT: space-time and gravity emergent

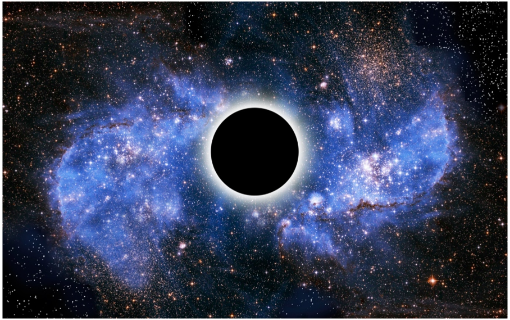
What are the fundamental principles?

Role of entanglement and quantum information

Space-time behind the horizon

Quantum black holes: statistical mechanics, information and limitations of locality

Motivations



Black hole information paradox

What happens when crossing the horizon?

How can we describe the black hole interior in AdS/CFT?

Motivations

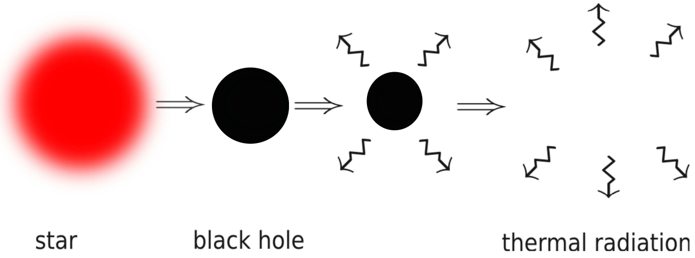
I will describe a proposal (developed with S. Raju) for describing the black hole interior, which may have implications towards the resolution of the information paradox

[JHEP 1310 (2013) 212], [PRL 112 (2014) 5], [Phys.Rev. D89 (2014)], [PRL 115 (2015)],[Int.J.Mod.Phys. D22 (2013)], [JHEP 1605 (2016), KP, S.Raju, J.W. Bryan, S. Banerjee]

Key physical principles:
i) Locality in quantum gravity is not exact
ii) State dependence of physical observables

More recent developments: a new class of non-equilibrium BH states, connection to traversable wormholes
[1708.06328, KP]
[1708.09370, Rik van Breukelen and KP],
work in progress with J. de Boer, S. Lokhande, R. van Breukelen, E. Verlinde

The information paradox

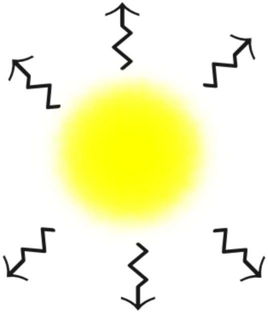


$$|\Psi\rangle \quad \stackrel{?}{\to} \quad \rho = \frac{e^{-\beta H}}{Z}$$

Inconsistent with **unitary** evolution in quantum mechanics

$$|\Psi(t)\rangle = e^{-iHt}|\Psi(0)\rangle$$

Normal “burning“ process



Why no information loss problem?

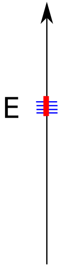
Radiation appears to be thermal

There are correlations (entanglement) between photons.

Typical size e^{-S} for small number of photons [Page]

The entanglement among all outgoing photons carries the full information of initial state

Pure vs Mixed states



$$|\Psi\rangle = \sum_i^N c_i |E_i\rangle \qquad \text{vs} \qquad \rho_{micro} = \frac{1}{N}\mathrm{I}$$

c_i = random coefficients

Theorem: In a large quantum system, for most pure states, and simple observables A, we have

$$\langle\Psi|A|\Psi\rangle = \mathrm{Tr}(\rho_{micro}A) + O(e^{-S})$$

but notice that for complicated observables where $n \approx S$

$$\langle\Psi|A_1 ... A_n|\Psi\rangle = \mathrm{Tr}(\rho_{micro}A_1 ... A_n) + O(e^{-(S-n)})$$

[S.Lloyd]

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

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

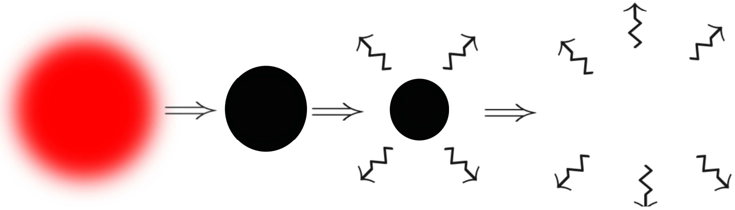
$$\overline{\langle\Psi|A|\Psi\rangle} = \langle A\rangle_{\rm micro}$$

and

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

"reasonable" observables have the same expectation value in most pure states, up to exponentially small corrections.

Unitarity from small corrections

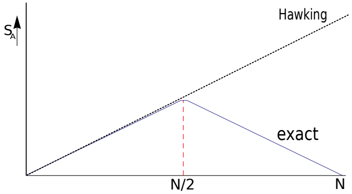


Hawking's computation is semiclassical, we do expect corrections

$$\rho = \rho_{\text{thermal}} + \rho_{\text{cor}}$$

Statistical Mechanics: Even if corrections ρ_{cor} were sufficiently large to restore unitarity, they would generally only lead to exponentially small $(e^{-S_{BH}})$ deviations from Hawking's predictions for simple observables.
Reminder: for solar mass BH $S_{BH} \approx 10^{77}$

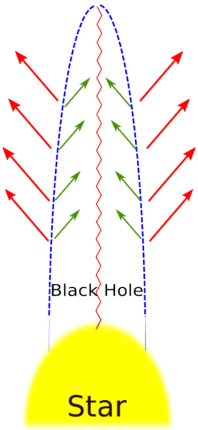
Comments



In the scenario of unitarization of BH evaporation via small corrections to Hawking's computation:

- Hawking predictions for simple observables may be accurate up to $e^{-S_{BH}}$ deviations
- There may be important deviations for complicated observables (for example correlators between $O(S_{BH})$ Hawking particles — significant entanglement)
- Hawking computation does not lead to a sharp paradox for observables in Effective Field Theory.
- So far we have not said anything about the BH interior...

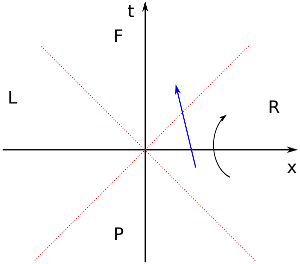
Entanglement near the horizon

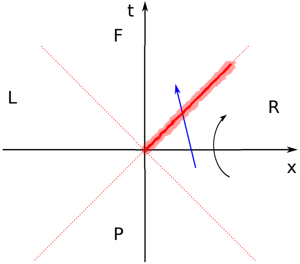


Hawking particles are produced in **entangled pairs**

This entanglement is **necessary** for the smoothness of spacetime near the horizon

Example: flat space, Unruh effect

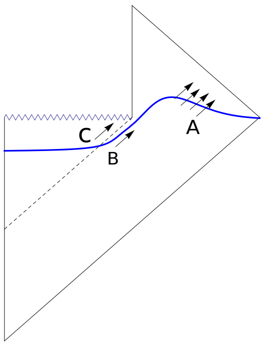




$$|0\rangle_M = \sum_{n=0}^{\infty} e^{-\pi\omega n} |n\rangle_L \otimes |n\rangle_R \qquad |\Psi\rangle = |0\rangle_L \otimes |0\rangle_R \rightarrow \langle T_{\mu\nu}\rangle \neq 0$$

Modern info paradox, infalling observer

[Mathur, 2009], [Almheiri, Marolf, Polchinski, Sully, 2012]



General Relativity: smooth horizon, B entangled with C

Quantum Mechanics: no information loss, B entangled with A

B violates monogamy of entanglement

Violation of strong subadditivity of entanglement entropy: for 3 independent systems A, B, C we have

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

Mathur's theorem: "small corrections cannot fix the problem " (?)

Unitarity or smooth horizon?

Giving up B-C entanglement?

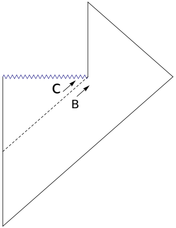
Firewall, fuzzball* proposals $\Rightarrow \langle T_{\mu\nu} \rangle$ at horizon is very large, BH interior geometry is completely modified (maybe no interior at all)

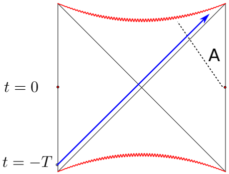
Infalling observer "burns" upon impact on the horizon.

Dramatic modification of General Relativity/Effective Field Theory over macroscopic scales, due to quantum effects

Chaos vs "specific entanglement"

Black Holes are Chaotic Quantum Systems





How can **typical states** have **specific** entanglement between B, C which is needed for smoothness?

Correct entanglement fragile under perturbations due to chaotic nature of system
[Shenker, Stanford]

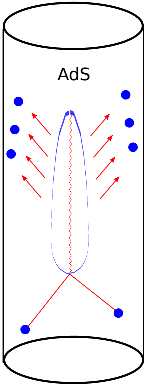
Summary

- The modern version of the info paradox, is intimately related to the smoothness of the horizon and to what happens to the infalling observer.

- We have a conflict between QM and General Relativity because it seems impossible to have the correct entanglement of quantum fields, needed for smoothness, near the horizon.

- We will study the problem in AdS/CFT.

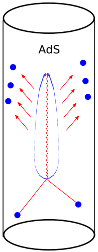
Black Holes in AdS/CFT



Non-perturbative Black Hole S-matrix encoded in CFT correlators

Manifestly Unitary

Black Hole interior in AdS/CFT?



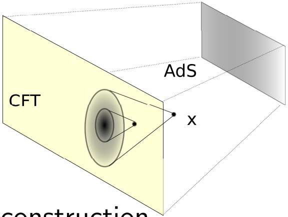
The modern information paradox is related to the smoothness of the BH horizon.

Can we study the black hole horizon/interior in AdS/CFT?

Until recently it was not known how to do this.

In work with S.Raju we proposed a new class of CFT operators which are able to describe the BH interior.

Local observables in AdS

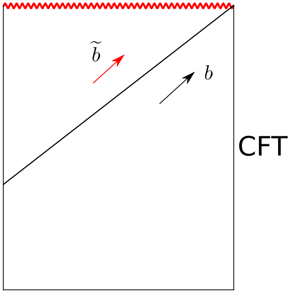


[Hamilton, Kabat, Lifshytz, Lowe] construction

$$\phi(x) = \int dY\, K(x,Y)\, \mathcal{O}(Y)$$

$\mathcal{O}=$ local single trace operator
K =known kernel
Locality in bulk is approximate:

1. True in $1/N$ perturbation theory
2. $[\phi(P_1), \phi(P_2)] = 0$ only up to e^{-N^2} accuracy
3. Locality may break down for high-point functions



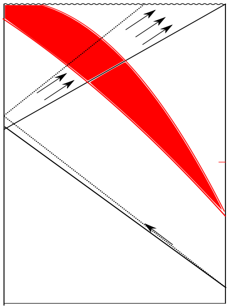
For smooth horizon effective field theory requires:

I) $\widetilde{b}$ commute with b AND II) $\widetilde{b}$ entangled with b

$$b \Leftrightarrow \mathcal{O}$$
$$\widetilde{b} \Leftrightarrow ?$$

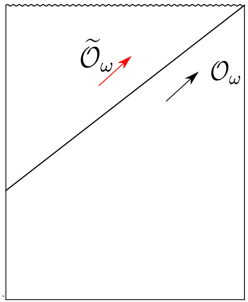
Which CFT operators $\widetilde{\mathcal{O}}$ correspond to $\widetilde{b}$? Why is operator algebra "doubled"?

Direct reconstruction?



- Transplanckian problem
- States formed by collapse form a small subset of typical BH states.

Firewall paradox for large AdS black holes



- [AMPSS, Marolf-Polchinski] paradox: effective field theory implies $[H, \widetilde{\mathcal{O}}^\dagger_\omega] = -\omega \widetilde{\mathcal{O}}^\dagger_\omega$. This leads to

$$Tr[e^{-\beta H} \widetilde{\mathcal{O}}^\dagger_\omega \widetilde{\mathcal{O}}_\omega] < 0$$

  which is inconsistent
- Notice that this is a firewall paradox for big, stable AdS black holes.

Is there a way out?

A construction of the BH interior

[KP and S.Raju]

- If we take a CFT state $|\Psi\rangle$ of $O(N^2)$ energy, we expect that at late times it will thermalize.

$$\langle\Psi|\mathcal{O}_1(x_1)...\mathcal{O}_n(x_n)|\Psi\rangle \approx Z^{-1}\mathrm{Tr}(e^{-\beta H}\mathcal{O}_1(x_1)...\mathcal{O}_n(x_n))$$

- This is true only for simple observables $n \ll N$

- Thermalization of pure state $\Rightarrow$ must have the notion of a small algebra of observables

- In a large N gauge theory, natural small "algebra" $\mathcal{A}$ = products of few, single trace operators

Intuitive picture

- Even though we are in a single CFT in a pure state, the small algebra of single trace operators probes the pure state $|\Psi\rangle$ as if it were an entangled state

$$\langle\Psi|\mathcal{O}\mathcal{O}...|\Psi\rangle \approx Tr[e^{-\beta H}\mathcal{O}\mathcal{O}...] \quad \leftrightarrow \quad |TFD\rangle = \sum_E \frac{e^{-\beta E/2}}{\sqrt{Z}}|E\rangle \otimes |E\rangle$$

- Operator algebra seems to be doubled! 2nd copy $\rightarrow$ operators behind horizon
- Usually thought of as a mathematical trick. In my work with S.Raju, we proposed a physical interpretation:

  The $O(N^2)$ d.o.f. of the CFT play the role of the "heat bath" for the small algebra of single trace operators. The second copy of the thermofield formalism represents this heat bath.
- Whatever operators the single trace operators are entangled with, will play the role operators behind the horizon.
- How do we identify these operators mathematically?

Small algebra of observables

Small algebra generated by single trace operators

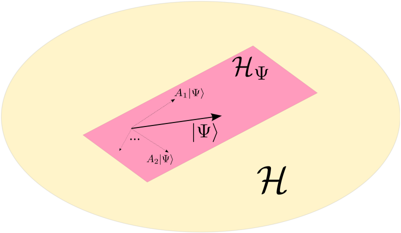
$$\mathcal{A} \equiv \mathrm{span}[\mathcal{O}(x_1), \mathcal{O}(x_1)\mathcal{O}(x_2), ...]$$

If $|\Psi\rangle$ is a BH microstate, we have nontrivial property

$$A|\Psi\rangle \neq 0 \qquad \forall A \in \mathcal{A}\ ,\ A \neq 0$$

Physically this means that the state seems to be entangled when probed by the algebra $\mathcal{A}$.

The small Hilbert space



$$\mathcal{H}_\Psi = \mathrm{span}\{\mathcal{A}|\Psi\rangle\}$$

Which was called “code subspace” in later works by other authors.

Effective Field Theory in bulk takes place within this subspace

$$\phi(x_1)...\phi(x_n)|\Psi\rangle$$

Tomita-Takesaki modular theory

Algebra $\mathcal{A}$ acts on $\mathcal{H}_\Psi$. It has two properties:
i) By acting on $|\Psi\rangle$ the algebra $\mathcal{A}$ generates $\mathcal{H}_\Psi$
ii) The algebra $\mathcal{A}$ cannot annihilate state $|\Psi\rangle$.

Theorem: The representation of the algebra on $\mathcal{H}_\Psi$ is reducible, and the algebra has an isomorphic commutant (2nd copy) acting on the same space.

Define antilinear map

$$SA|\Psi\rangle = A^\dagger|\Psi\rangle$$

and

$$\Delta = S^\dagger S \qquad J = S\Delta^{-1/2}$$

Then for any $\mathcal{O} \in \mathcal{A}$, the operators

$$\boxed{\widetilde{\mathcal{O}} = J\mathcal{O}J}$$

i) commute with elements of $\mathcal{A}$
ii) are entangled with $\mathcal{O}$ (non-zero 2-point functions)
These are the operators that we need for the Black Hole interior.

The modular Hamiltonian

The operator $\Delta = S^\dagger S$ is a positive, hermitian operator and can be written as

$$\Delta = e^{-K}$$

where

$$K = \text{modular Hamiltonian}$$

for the small algebra

Using the large N expansion and the KMS condition for thermal correlators in equilibrium states

$$K = \beta(H_{CFT} - E_0)$$

In practice

$$\widetilde{\mathcal{O}}_\omega|\Psi\rangle = e^{-\frac{\beta\omega}{2}}\mathcal{O}_\omega^\dagger|\Psi\rangle$$

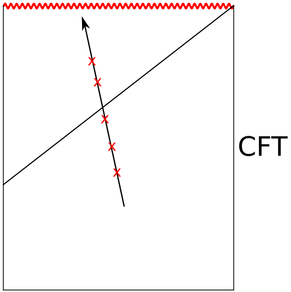
$$\widetilde{\mathcal{O}}_\omega\mathcal{O}....\mathcal{O}|\Psi\rangle = \mathcal{O}...\mathcal{O}\widetilde{\mathcal{O}}_\omega|\Psi\rangle$$

$$[H,\widetilde{\mathcal{O}}_\omega]\mathcal{O}....\mathcal{O}|\Psi\rangle = \omega\widetilde{\mathcal{O}}_\omega\mathcal{O}....\mathcal{O}|\Psi\rangle$$

These equations define the operators $\widetilde{\mathcal{O}}$ on a subspace $\mathcal{H}_\Psi \subset \mathcal{H}_{\rm CFT}$, which is relevant for EFT around BH microstate $|\Psi\rangle$

$$\mathcal{H}_\Psi = \text{span}\mathcal{A}|\Psi\rangle$$

Equations admit solution because the algebra $\mathcal{A}$ cannot annihilate the state $|\Psi\rangle$



Bulk field inside BH

$$\phi(t,r,\Omega) = \int_0^\infty d\omega \Big[\mathcal{O}_\omega\, f_\omega(t,\Omega,r) + \widetilde{\mathcal{O}}_\omega g_\omega(t,\Omega,r) + \mathrm{h.c.}\Big]$$

Correlation functions of these operators reproduce those of effective field theory in the exterior/interior of the black hole

Smooth spacetime at the horizon, no firewall/fuzzball. At the same time, Unitarity is OK

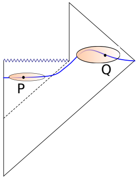
What about previous paradoxes?

Non-locality in Quantum Gravity

$\widetilde{\mathcal{O}}$ were constructed based on the fact that we restricted our attention to a "small algebra" of $\mathcal{O}$'s. The construction breaks down if the "small algebra" is enlarged to include all operators

$[\mathcal{O}, \widetilde{\mathcal{O}}] = 0$ only on $\mathcal{H}_\Psi$, not as operator equation

Operators $\widetilde{\mathcal{O}}$ = complicated combinations of $\mathcal{O}$. Realization of BH complementarity



$$[\phi(P), \phi(Q)] \sim 0$$

$$[\phi(P), \Phi^{\text{complex}}(Q)] = O(1)$$

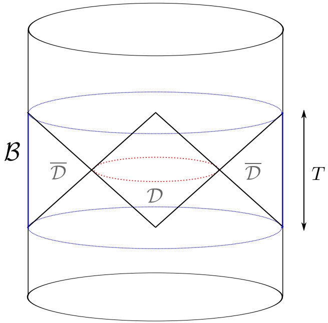
The Hilbert space of Quantum Gravity does not factorize as $\mathcal{H}_{\text{inside}} \otimes \mathcal{H}_{\text{outside}}$

1) Solves problem of Monogamy of Entanglement (and avoids Mathur's theorem)

2) Is consistent with locality in EFT, concrete mathematical realization of complementarity

A toy model of complementarity

[JHEP 1605 (2016), KP, S.Raju, J.W. Bryan, S. Banerjee]



Global AdS: operators in $\mathcal{D}$ can be represented as complicated operators in the time band $\mathcal{B}$

State-dependence

- Interior operators defined by

$$\widetilde{\mathcal{O}}_\omega|\Psi\rangle = e^{-\frac{\beta\omega}{2}}\mathcal{O}_\omega^\dagger|\Psi\rangle$$

$$\widetilde{\mathcal{O}}_\omega\mathcal{O}....\mathcal{O}|\Psi\rangle = \mathcal{O}...\mathcal{O}\widetilde{\mathcal{O}}_\omega|\Psi\rangle$$

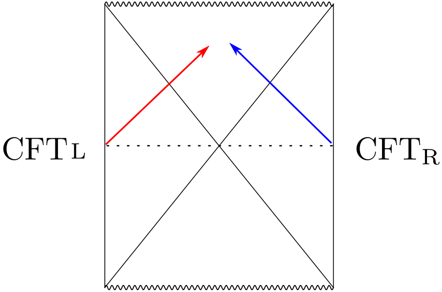
$$[H,\widetilde{\mathcal{O}}_\omega]|\Psi\rangle = \omega|\Psi\rangle$$

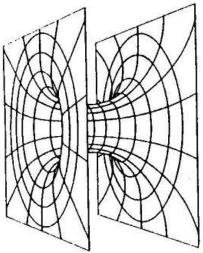
- Solution defined only on $\mathcal{H}_\Psi$, depends on reference state $|\Psi\rangle$
- Operators cannot be upgraded to "globally defined" operators
- State-dependence solves Chaos vs Entanglement problem naturally: operators are selected by the entanglement!
- Novel QM feature of black hole interior?

Connection to ER = EPR

[K.P and S.Raju (1503.08825)]

Entanglement & Wormholes (Maldacena, Susskind, Raamsdonk)

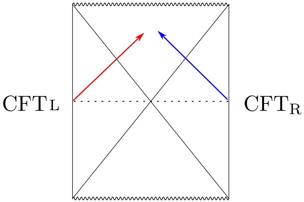




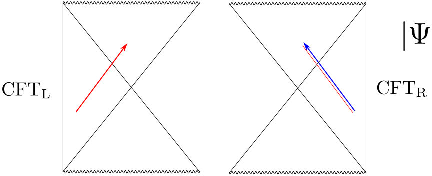
$$H = H_L + H_R$$

$$|\text{TFD}\rangle = \sum_E \frac{e^{-\beta E/2}}{\sqrt{Z}} |E\rangle_L \otimes |E\rangle_R$$

ER=EPR



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



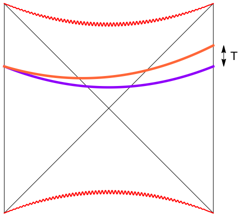
$$|\Psi\rangle = \sum_{ij} c_{ij} |E_i\rangle_L \otimes |E_j\rangle_R$$

$$c_{ij} = \text{generic}$$

Time-shifted wormholes

[K.P and S.Raju, PRL 115 (2015)]

$$|\Psi_T\rangle \equiv e^{iH_R T}|\mathrm{TFD}\rangle$$



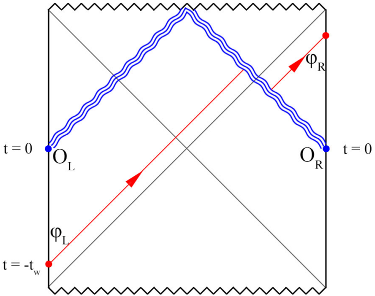
The states $|\Psi_T\rangle$ are related to $|\mathrm{TFD}\rangle$ by a large diffeomorphism. They should* be as smooth as $|\mathrm{TFD}\rangle$.

We showed that it is impossible to find fixed operators, for all states $|\Psi_T\rangle$, describing the BH interior

Strong evidence in favor of state-dependence

Proof using traversable wormholes

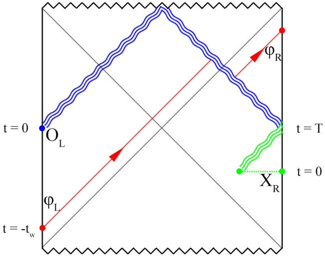
[Gao-Jafferis-Wall],[Maldacena, Stanford, Yang]
[1708.09370, Rik van Breukelen, KP]



Evidence for smoothness of $|\mathrm{TFD}\rangle$ state.

Traversable wormholes and state-dependence

[1708.xxxxx, Rik van Breukelen, KP]



$$|\Psi_T\rangle = e^{iH_R T}|TFD\rangle$$

couple two CFTs at $t = 0$ with

$$U = e^{igO_L(t=0)X_R(t=0)}$$

where $X_R \equiv e^{iH_R T} O_R e^{-iH_R T}$

This shows that indeed a very large class of states

$$|\Psi_T\rangle = e^{iH_R T}|TFD\rangle$$

are smooth! As mentioned in this previous slide this can only happen if the interior operators are state dependent.

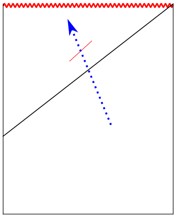
Hence this new result **confirms** state-dependence within this class of states.

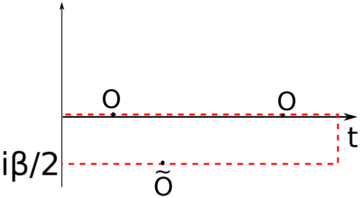
Summary on state dependence

- Solves the firewall paradox, provides reconstruction of BH interior in AdS/CFT.
- New feature in QM, needs to be understood better.
- Quantum measurement theory for the infaller (observer is part of the system)
- Time evolution for observer crossing the horizon
  (is infaller Hamiltonian state-dependent, if so, what principle selects it?)

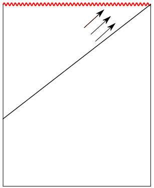
Thermalization in gauge theories

[KP 1708.06328]



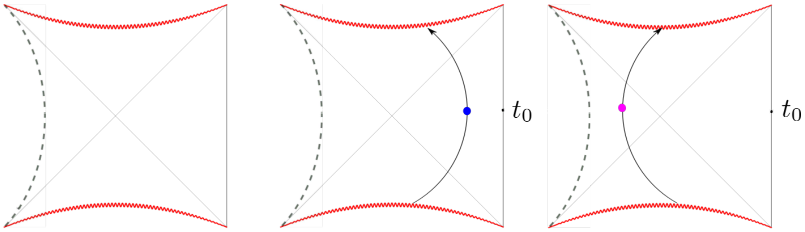


A new class of non-equilibrium states



$$|\Psi\rangle = U(\widetilde{\mathcal{O}})\,|\Psi_0\rangle = e^{-\frac{\beta H}{2}}\,U(\mathcal{O})e^{\frac{\beta H}{2}}\,|\Psi_0\rangle$$

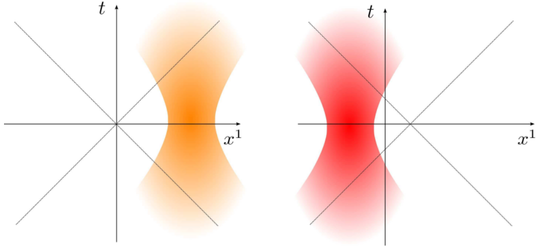
A new class of non-equilibrium states



$|\Psi_0\rangle$ $\qquad$ $U(\mathcal{O})|\Psi_0\rangle$ $\qquad$ $e^{-\frac{\beta H}{2}}U(\mathcal{O})e^{\frac{\beta H}{2}}|\Psi_0\rangle$

- $|\Psi_0\rangle$ = equilibrium state
- $U(\mathcal{O})|\Psi_0\rangle$ = standard non-equilibrium state
- $e^{-\frac{\beta H}{2}}U(\mathcal{O})e^{\frac{\beta H}{2}}|\Psi_0\rangle$ = new type of non-equilibrium state

Localized states in Rindler space

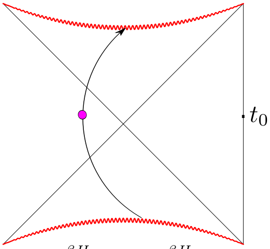


For Rindler space, modular Hamiltonian is Lorentz boost generator M in t, x^1 plane.
Unruh inverse temperature

$$\beta = 2\pi$$

$$e^{-\pi M} U_R e^{\pi M} |0\rangle = U_L' |0\rangle$$

Properties of the new states



$e^{-\frac{\beta H}{2}} U(\mathcal{O}) e^{\frac{\beta H}{2}} |\Psi_0\rangle$

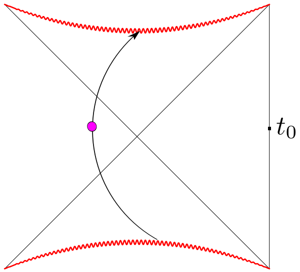
- They seem to be in equilibrium in terms of single-trace correlators

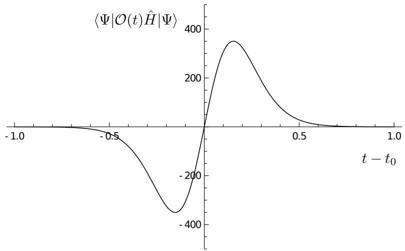
$$\frac{d}{dt}\langle\Psi|\mathcal{O}(t)|\Psi\rangle = 0$$

- It can be seen that they are out of equilibrium by incuding H in the correlator

$$\frac{d}{dt}\langle\Psi|\mathcal{O}(t)H|\Psi\rangle \neq 0$$

Example

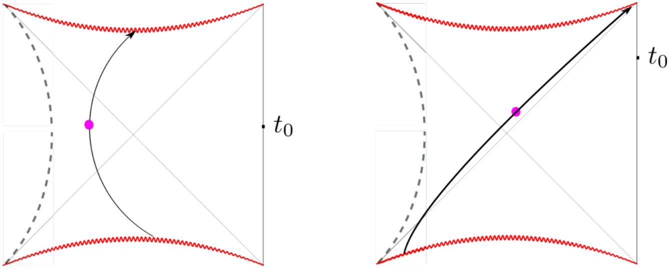




Consider a 2d CFT on $\mathbb{S}^1 \times R$ on a state $|\Psi\rangle = e^{-\frac{\beta H}{2}} U(\mathcal{O}) e^{\frac{\beta H}{2}} |\Psi_0\rangle$, with $U = e^{i\theta\mathcal{O}(t_0)}$. Then at large c we find

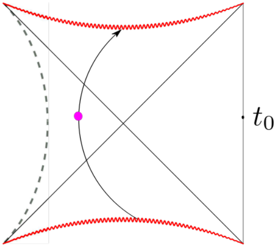
$$\langle\Psi|\mathcal{O}(t)\hat{H}|\Psi\rangle = \theta\, 2\Delta \left(\frac{2\pi}{\beta}\right)^{2\Delta+1} \sum_{m=-\infty}^{+\infty} \frac{\sinh\left(\frac{2\pi(t-t_0)}{\beta}\right)}{\left[2\cosh\left(\frac{4\pi^2 m}{\beta}\right) + 2\cosh\left(\frac{2\pi(t-t_0)}{\beta}\right)\right]^{\Delta+1}}$$

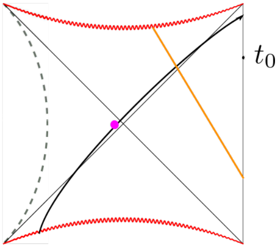
Extracting the particle behind the horizon



[in progress with J. de Boer, S. Lokhande, R. van Breukelen]

Testing the conjecture





We can create negative energy shockwaves by acting with

$$e^{ig\mathcal{O}\widetilde{\mathcal{O}}}$$

on the state

$$e^{-\frac{\beta H}{2}}U(\mathcal{O})e^{\frac{\beta H}{2}}|\Psi_0\rangle$$

The excitation should be detected in the CFT with usual single trace operators.
See also recent work of [Kourkoulou, Maldacena] for similar states in SYK model

Non-equilibrium states and the black hole interior

- We have identified a class of states in the Hilbert space of the boundary CFT, which correspond black holes with excitations behind the horizon.
- They can be simply written as

$$e^{-\frac{\beta H}{2}} U(\mathcal{O}) e^{\frac{\beta H}{2}} |\Psi_0\rangle$$

  without having to use $\widetilde{\mathcal{O}}$.
- Their existence gives additional evidence that BH interior can be described in the CFT
- They contain information about part of the "left region" for a 1-sided black hole!
- These states may be interesting more generally from the point of view of statistical mechanics

Summary and outlook

- The modern version of the info paradox has to do with entanglement at the horizon
- I described a proposal suggesting how it might be resolved.
- This proposal provides a reconstruction of the BH interior in AdS/CFT
- Key principles: non-locality and state-dependence
- Interesting connections with non-equilibrium states and thermalization
- The "traversable wormhole" protocols open up new exciting ways of testing these ideas and probing the black hole interior via scattering experiments.
- New evidence in favor of state-dependence.

THANK YOU