Introductory Lecture on the BHIP → Paradox

(B → black, H → Hole, I → information)

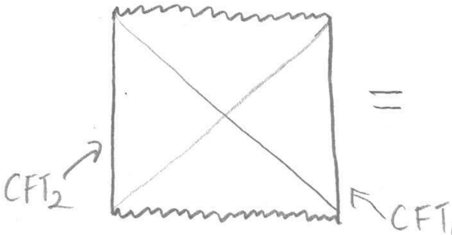
Section 1: Why AdS/CFT doesn't save us   01/06/12

You may have heard by now that AdS/CFT is a theory that defines quantum gravity in AdS via a "dual" formulation as a CFT. The dictionary is:

$$\boxed{Z^{AdS}_{gravity}[\partial M = \Sigma] = Z_{CFT}[\Sigma]}$$

where M is the bulk manifold with boundary Σ

Statement: The eternal Schwarzschild black hole is dual to the thermofield double state of a pair of CFTs, which are uncoupled but have the same spectrum

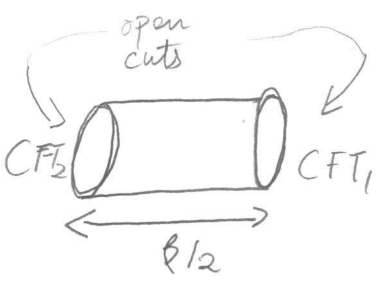


$$= \sum_n e^{-\beta E_n/2} |n_1\rangle |n_2\rangle \equiv |TFD\rangle$$

β is the temperature of the black hole, $|n\rangle$ is an energy eigenstate of the CFT Hamiltonian.

This state can be prepared by a Euclidean path integral. Let d be the dimension of the CFT and let the CFT states live on a spatial S^{d-1} (For 2d CFT, this is just a circle).

The Euclidean path interval that prepares this state is one where we evolve over a Euclidean time interval of length $\beta/2$.

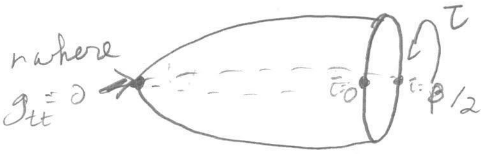


This is clear from the definition above.

How do we know this is dual to the eternal black hole?

Using the AdS/CFT dictionary we want a geometry with $Z^{AdS}_{gravity}[\partial M = \boxed{}\!\!\!\supset]$

This is exactly half the Euclidean Schwarzschild BH



$$ds^2 = \left(1 + \frac{r^2}{\ell^2} - \frac{2M}{r^{d-1}}\right) d\tau^2 + \frac{dr^2}{1 - \frac{2M}{r^{d-1}} + \frac{r^2}{\ell^2}} + r^2 d\Omega^2$$

$$\tau \sim \tau + \beta$$

Important $Z^{AdS}_{grav}[\partial M = \boxed{}\!\!\!\supset]$ will have other contributions but when G_N is small, this will be the leading saddle that dominates the path integral.

→ we use this to define our initial quantum state. For Lorentzian evolution.

Now we want to compute a correlation function

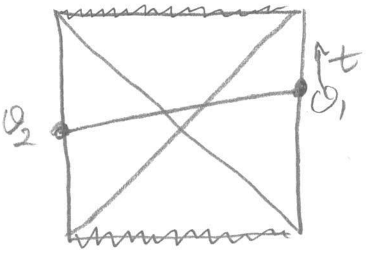
$$\langle TFD | \mathcal{O}_1(t)\, \mathcal{O}_2(0) | TFD \rangle = \sum_{n,m} e^{-\frac{\beta}{2}(E_n + E_m)} \langle n_1 | \langle n_2 | e^{iH_1 t} \mathcal{O}_1(0) e^{-iH_1 t} \mathcal{O}_2(0) | m_1 \rangle \times | m_2 \rangle$$

$$= \sum_{n,m} e^{-\frac{\beta}{2}(E_n + E_m) + i(E_n - E_m)t} \langle n_1 | \mathcal{O}_1 | m_1 \rangle \langle n_2 | \mathcal{O}_2 | m_2 \rangle$$

⤳ The large phases in this sum can make the correlation function get very small at late times, but it can't be exactly zero if the spectrum is finite

Let us compare this to a gravity calculation

You have seen by now that, correlation functions in CFT can be approximated in gravity by geodesics (or Entanglement entropies as extremal surfaces)



$$\langle \mathcal{O}_1 \mathcal{O}_2 \rangle_{TFD} = e^{-\text{Geodesic length}} = \# e^{-\# t/\beta}$$

⇝ goes to zero at late times since geodesic grows without bound

→ This geodesic always stays well away from the singularity, so it shouldn't probe strong gravitational effects.

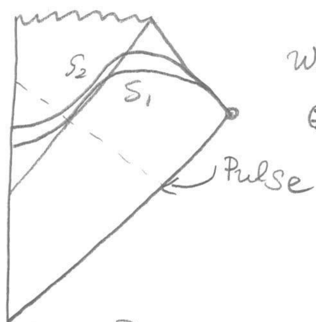
→ Other geometries contribute? If so, what? How does AdS gravity avoid this apparent violation of unitarity?

→ Do we need to modify the geometry inside the BH at late times (firewall?)

Since I've now convinced you that AdS/CFT still has an info problem, let us try to understand the real statement of the problem in its original context: flat space.

Flat space information paradox & a crisis of entanglement: 0909.1038

Consider a black hole formed from collapse:



we need to define a set of "niceness conditions" ~~condition~~ such that we can rely on semiclassical physics to some high accuracy

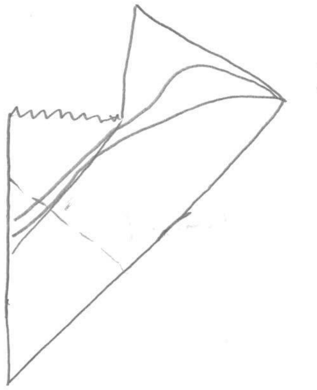
$D = \#$ of bulk spacetime dimension

1) Quantum states are defined on spacelike slices with Intrinsic curvature $R^{(D-1)} \ll \frac{1}{l_p^2}$

2) The extrinsic curvature of the slice should also be small $K \ll \frac{1}{l_p^2}$

3) The full curvature of the spacetime in the vicinity of the slice should also satisfy ${}^{(D)}R \ll 1/l_p^2$

4) Under time evolution, a nice slice should evolve into another nice slice



To make such a slice:

Intermediate geometry $ds^2 = -\left(1-\frac{2M}{r}\right)dt^2 + \frac{dr^2}{(1-2M/r)} + r^2 d\Omega_2$

1) $r > 4M$ $\quad t = \text{const}$

2) $r < 2M$ $\quad r = r_1 = \text{const}$ w/ $M/2 < r_1 < 3M/2$
s.t. it's not close to the horizon or the singularity

3) connect these by a <u>smooth connector</u>

4) $r = r_1$ is spacelike since $r < 2M$.
Follow this until before the matter formed a black hole then connect it smoothly to $r = 0$ where there is no singularity @ early times.

Can check that under time evolution, these slices satisfy the niceness conditions.

<u>Stretching on the slice</u>

All the stretching between successive slices happens in a given place: the connector region

→ Then the QFT vacuum on one slice isn't the natural vacuum on a later slices

⇝ This creates ~ 1 hawking pair at wavelength $\lambda \sim L$ which characterizes the stretching region

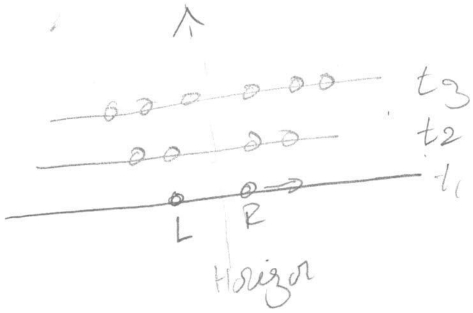
Then $\boxed{|\Psi\rangle_{pair} = C e^{\# a_L^\dagger a_R^\dagger} |0_L\rangle|0\rangle_R}$

where L / R denote inside / outside the BH

We can approximate $|\Psi\rangle_{pair} \simeq \frac{1}{\sqrt{2}}\left(|0_L\rangle|0_R\rangle + |1_L\rangle|1_R\rangle\right)$

<u>Locality of matter</u> we will assume the infalling matter is far from the stretching region s.t. $|\Psi\rangle_{tot} = |\Psi_m\rangle \otimes |\Psi\rangle_{pair}$

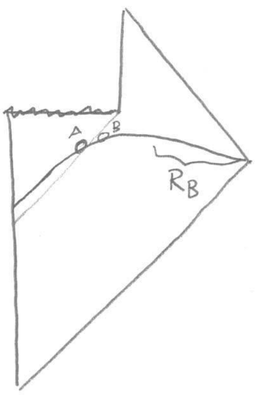
In the exercises, you will show the following: evolving for N time steps the quantum state of the R system has entanglemen entropy



$$\boxed{S_{ent} = N \log 2}$$

If the black hole has fully evaporated, there is no L subsystem to trace over & the R system is the full state. ⇝ If the initial state $|\Psi_{in}\rangle$ was pure, the final subsystem must be mixed since $S_{ent} \neq 0$. This can not happen under unitary time evolution.

AMPS argument (1207.3123)



Let us try to change the argument around: Let us consider that the evolution is unitary such that a very late Hawking quantum (labeled B in the figure) is purified by the radiation that has come out (R_B in figure)

In equations: $S(\rho_{B+R_B}) \approx 0$

Now recall some definitions from the previous lessons:

mutal information: $I_{A,B} = S_A + S_B - S_{AB} \geqslant 0$

Implies $|S_A - S_B| \leqslant S_{AB}$

Strong subadditivity: $S_{ABC} + S_B \leq S_{AB} + S_{BC}$

or: $I_{A,B} \leqslant I_{A,BC}$

Now: $I_{A,BR_B} = S_A + \cancel{S_{BR_B}}^{\,0} - S_{ABR_B} \geqslant 0$ & $|S_A - \cancel{S_{B,R_B}}| \leqslant S_{ABR_B}$

$\Rightarrow \boxed{S_A = S_{ABR_B}}$

So the mutual information $I_{A, BR_B} = 0$

Homework: Show $I_{A, BR_B} = 0$ iff $\rho_{ABR_B} = \rho_A \otimes \rho_{BR_B}$

If you stare at this long enough you will notice a contradiction with what we were saying earlier: namely that Hawking quanta are produced in entangled pairs. This is incompatible (!) with purification at late times. We must choose. If a quantum B is highly entangled with the early radiation, it cannot be entangled with the interior mode A.

Q But the entanglement between A & B was derived by assuming the validity of effective field theory on a weakly curved slice. Surely that can't break down, can it?

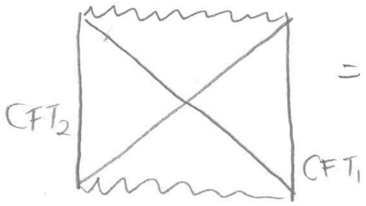
A Unfortunately we can't have our cake and eat it. The reasonable assumption of the validity of low energy effective field theory led to the unreasonable assumption of non-unitary time evolution

The reasonable assumption of unitarity restoration leads us to consider a pattern of entanglement that starkly contrasts with the predictions of effective field theory.

→ This led AMPS to conclude that there must be a firewall that forms behind the horizon to save unitarity.

ER = EPR & the future

One idea to take seriously is ER = EPR



= $|TFD\rangle$, which is a state with a lot of entanglement. In fact, one could not easily imagine reconstructing the interior of the BH from the exterior of side 1 without access to CFT_2

Let us think of A radiation quantum in this setup (in system 1). We know by the TFD construction that it is entangled with radiation in system 2, in other words B is not entangled with R_B, but with A

In an evaporating black hole perhaps one way around the problem is to think of A & R_B as the same subsystem such that there's no contradiction. This violates our notion of locality in a strong sense, but I don't make up the rules.