AdS/CFT and Black Hole Complementarity

or

What I learned at the Bits and Branes program

Joseph Polchinski

with

Idse Heemskerk, Don Marolf, James Sully, Ahmed Almheiri

1201.3664 and in progress

Black Holes and Information Conference

KITP, May 21, 2012

Things I believed at the beginning of the Bits and Branes program:

1. I understand, more-or-less, what black hole complementarity (BHC) means.
2. BHC resolves all paradoxes in black hole decay.
3. AdS/CFT duality implements BHC.

Goal during program: make these more precise.

What I learned:

What I learned:

1. I am confused.

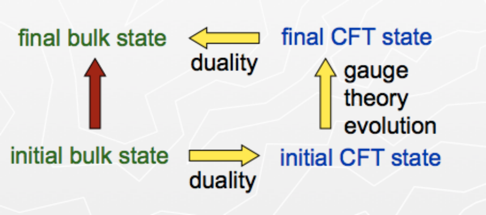
Outline

I. Cats behind horizons

II. Confusions about complementarity

Where was Hawking's 'mistake'?

For quantum gravity in AdS spacetime, AdS/CFT duality implies that the process of black hole formation and decay is unitary, and that the information is emitted with the Hawking radiation. But where does the argument for information loss break down?



This diagram doesn't seem to commute!

I. Cats behind horizons

If Schrodinger's cat were behind the horizon of an AdS black hole, could we determine its state by a measurement in the dual CFT?

Yes: Banks, Douglas, Horowitz, Martinec, Balasubramanian, Ross, Susskind, Toumbas, Maldacena, Hubeny, Kraus, Ooguri, Shenker, Kleban, Freivogel, Hamilton, Kabat, Lifschytz, Lowe, Marolf, Lawrence, Silverstein, Heemskerk, Polchinski

One strategy: construct field operators $\phi(y)$ for points behind the horizon in terms of operators in the CFT.

Bulk operators

First, review construction for scalar fields in the bulk of AdS (Banks, Douglas, Horowitz, Martinec; Balasubramanian, Kraus, Lawrence, Trivedi; Bena; *Hamilton, Kabat, Lifschytz, Lowe*)

Consider any bulk Green's function

$$(\Box' - m^2)G(y|y') = \frac{1}{\sqrt{-g}}\delta^{d+1}(y-y')$$

Then at leading order in $1/N$ (free bulk fields),

$$\begin{aligned}\phi(y) &= \int d^{d+1}y' \sqrt{-g'}\phi(y')(\Box' - m^2)G(y|y') \\ &= \text{surface term} \\ &= \int d^dx' \, K(y|x')\mathcal{O}(x') \,.\end{aligned}$$

$$\phi(y) = \int d^d x' \, K(y|x') \mathcal{O}(x') + \dots$$

Extends order-by-order in $1/N$.

Form of smearing function K depends on the choice of Greens function (it's convenient to choose one that is nonvanishing only in spacelike directions) but result is equivalent.

Extends to other fields (in gauge-fixed form).

This is an operator relation, so it extends to nontrivial backgrounds, as long as they are close to AdS.

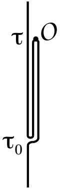
$$\phi(y) = \int d^d x' \, K(y|x') \mathcal{O}(x') + \ldots$$

This represents the bulk operator in terms of a superposition of local CFT operators at different times. By using the CFT evolution,

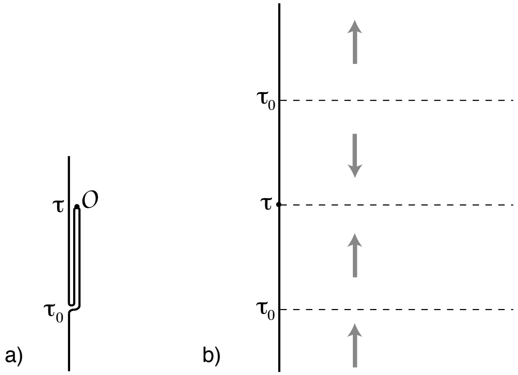
$$\mathcal{O}(\tau, \vec{x}) = e^{-iH(\tau - \tau_0)} \mathcal{O}(\tau_0, \vec{x}) e^{iH(\tau - \tau_0)}.$$

we can express it in terms of operators at a single time. These ‘precursor’ operators are necessarily nonlocal, and gauge invariant, so it is conventional to call them Wilson loops. However, this may not be the best description: the paths are extremely irregular.

We can also give a path integral construction, inserting a ‘fold’ into the space:

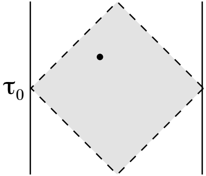


This is dual to a folded bulk.



τ
O
τ0
a)
τ0
τ
τ0
b)

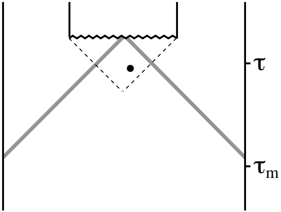
A nice feature is that the mapping between CFT operators at time τ_0 and operators in the bulk region spacelike with respect to τ_0 is independent of the Hamiltonian at other times.



Two arguments: (1) Backward/forwards evolution (2) The bulk and boundary operator lie on the same Cauchy surface, so the relation between them is determined purely by the bulk constraints (cf. path integral).

Behind the horizon

This immediately tells us that we can make measurements behind horizons. E.g. the thought experiment of Hubeny:



We can measure the bulk field at CFT time τ_m exactly as before and then throw in a shell such that field is behind a horizon.

More generally, for a black hole formed in collapse we can evolve the field equations backwards to a time before the collapse, and then use the AdS construction. For an eternal black hole, bulk operators behind the horizon would involve operators in both CFT's.

Conclusion: yes, we can measure the state of the cat behind the horizon.

Lesson: the Hilbert space of an observer behind the horizon can be embedded in the full CFT Hilbert space. But after the black hole evaporates, this is just the CFT of the outgoing Hawking radiation: this is black hole complementarity.

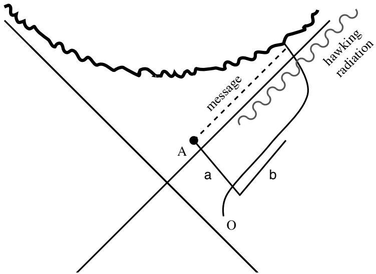
A weakness in this construction: it requires that we are able to solve *both* the bulk and boundary dynamics. Can we do better?

Hamilton, Kabat, Lifshitz, Lowe suggest that spacelike commutativity of $\phi(y)$ and $\phi(y')$ (appropriately modified to take account of the gravitational constraints), together with the AdS/CFT boundary dictionary, is enough to determine the fields.

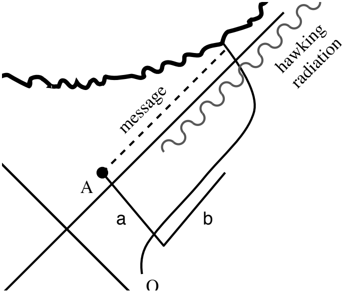
But this depends implicitly on knowing the dynamics — if we use the wrong dynamics, we can construct commuting fields that have nothing to do with the physics (Susskind).

II. Confusions about complementarity

Review original idea (figure from Susskind and Thorlacius).



Throw a bit a into a black hole. If info is conserved, this same bit is encoded in the Hawking radiation. No observer can see both copies (no-cloning). This is consistent if the bit takes at least a time $r_S \ln r_S$ to thermalize.



The two copies of the bit lie on a single spacelike hypersurface. The full low energy Hilbert space on this surface would have two copies of the bit — this is not consistent with QM. One should talk only about the smaller Hilbert spaces available to individual observers.

This seems to be morally consistent with AdS/CFT: the starting point is nonlocal, and does not approximate quantum field theory in curve spacetime. The Hilbert space of the observer behind the horizon is embedded in the CFT Hilbert space, which we identify with the Hilbert space of the external observer.

Bit models

One can also frame the black hole information paradox in terms of the naturally produced Hawking pairs. The bit models of Mathur; Czech, Larjo, Rozali; Giddings; Avery abstract the black hole down to these pairs.

$$|\hat{0}\hat{0}\hat{1}\hat{0}\hat{0}\hat{1}\hat{0}\hat{0}\hat{0}\hat{0}\hat{1}\hat{1}\hat{1}\hat{1}\hat{1}\hat{0}\hat{1}\hat{1}\hat{0}\hat{1}\hat{0}\hat{1}\hat{0}100010001000010110100011\rangle$$

Theorem of Mathur: in this context, if the information is emitted in the Hawking radiation then one must have an $O(1)$ modification of the quantum mechanics of the modes near the horizon: not just the Hawking state $|\hat{0},0\rangle + |\hat{1},1\rangle$ but also $|\hat{0},1\rangle$, $|\hat{1},0\rangle$, and $|\hat{0},0\rangle - |\hat{1},1\rangle$ are produced. (Mathur: the horizon cannot be *information-free*)

Such a large modification of the local physics is surprising. The bit models are based on a 'large' Hilbert space, which BHC excludes. Can we develop an alternative picture based on Hilbert spaces of individual observers?

The problem (Mathur, ...)

Once the black hole has lived $O(1/2)$ of its mass, all successive Hawking photons must be maximally entangled with the the first half of the Hawking radiation (Page).

But in the Hawking process, each photon is strongly entangled with its partner.

This is inconsistent with strong subadditivity of entropy, and tantamount to cloning.

Strong subadditivity

$$S_{AB} + S_{BC} \geq S_{ABC} + S_B$$

Apply this with A = first half of Hawking radiation, B = later Hawking photon, C = partner of B behind horizon. BC is in a pure state (in the Hawking process), so $S_{BC} = 0$ and $S_{ABC} = S_A$. $S_B > 0$ because B is entangled with C. So $S_{AB} > S_A$, but we need $S_{AB} < S_A$ to get the information out.

In order for BHC to save us, we would need that no observer could see all of A, B, C. But this requires only that they see the first half of the Hawking radiation and the pair at the horizon. This seems easy for an infalling observer (cf. Susskind and Thorlacius), but may be an out.

Thinking about this seems to lead to a bigger problem...

A bigger problem

Once the black hole has radiated $O(1/2)$ of its mass, its internal state is fully entangled with the Hawking radiation: we can make a measurement on the early radiation that will tell us whether a given Hawking state will be populated (Page; Hayden & Preskill).

But a state with a definite number of outgoing Hawking quanta must have high energy quanta near the horizon. The outgoing mode operators are related to those near the horizon by

$$b_\omega = \int_0^\infty d\omega' \left(\alpha^*_{\omega\omega'} a_{\omega'} - \beta^*_{\omega\omega'} a^\dagger_{\omega'} \right) ,$$

so if e.g. $b_\omega|\psi\rangle = 0$ then there must be $a^\dagger$ excitations that can be seen by an infalling observer.

Burning up at the horizon

Further conjecture: It isn't necessary to actually make the measurement on the early Hawking radiation, the fact that we could do so is enough: the radiation has already decohered the states of the later Hawking photons. The conclusion seems to be that if the Hawking radiation is pure, infalling observers burn up. (Other arguments due to Marolf.)

The usual argument against this is that there is nothing special about the horizon locally, only globally. But we expect locality to break down, maybe this is the manifestation.

I don't really believe this, but at this point I see several alternatives all unsatisfactory...

Conclusions:

I. We can make quantum measurements behind horizons using operators in the CFT, but, with the current state of knowledge, this requires a complete understanding of the bulk and boundary dynamics separately.

IIA. It is not obvious how BHC avoids $O(1)$ modifications of the quantum dynamics near the horizon as found in the bit models. Perhaps an infalling observer cannot actually make the necessary measurements?

IIB. How do we keep the infalling observer from burning up? Or is this the actual situation?