An Apology for Firewalls

Joe Polchinski

1207.3123, Ahmed Almheiri, Don Marolf, JP, Jamie Sully
In preparation, AMPS + Douglas Stanford

Institute on Black Hole Horizons and Quantum Information
CERN, 3/22/13

APOLOGY:

1. An admission of error or discourtesy accompanied by an expression of regret.

2. Defending a position through the systematic use of information; apologetics.

APOLOGY:

1. ~~An admission of error or discourtesy accompanied by an expression of regret.~~

2. Defending a position through the systematic use of information; apologetics.

I. Review and comments on the firewall argument.

II. Problems with $A \subseteq E$

III. Problems with $A \subseteq \mathcal{H}_{\text{CFT}}$. 'Strong complementarity'

IV. Sharpening the argument from AdS/CFT

V. Problems with nonviolent nonlocality

I. Black hole complementarity (‘t Hooft, Preskill, Susskind, Thorlacius, Uglum 1993):

Information is not lost. Different observers (infalling vs. external) see the same bit in radically different places.

AdS/CFT duality (Maldacena 1997):

Information is not lost. Different observers (CFT vs. bulk) can see the same bit in radically different places.

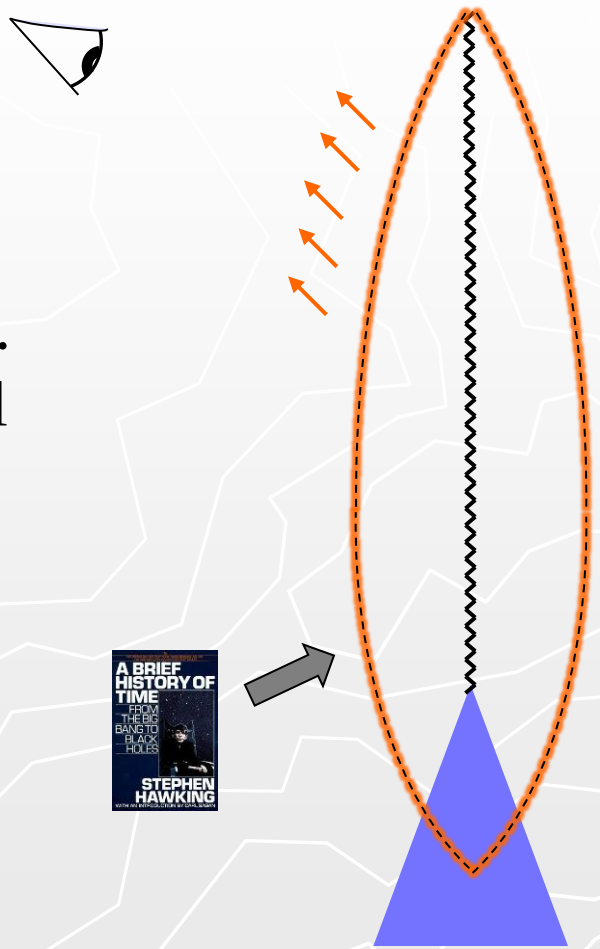
It seems like a perfect fit…

Black hole complementarity II

A further connotation: the external observer sees the 'stretched' horizon as a complicated dynamical system, which can absorb, thermalize, and reradiate information. Outside the stretched horizon, normal effective field theory applies.

The infalling observer sees nothing special at the horizon.

Challenge: to make a concrete model (cf. Mathur, Giddings). Instead, a no-go argument.

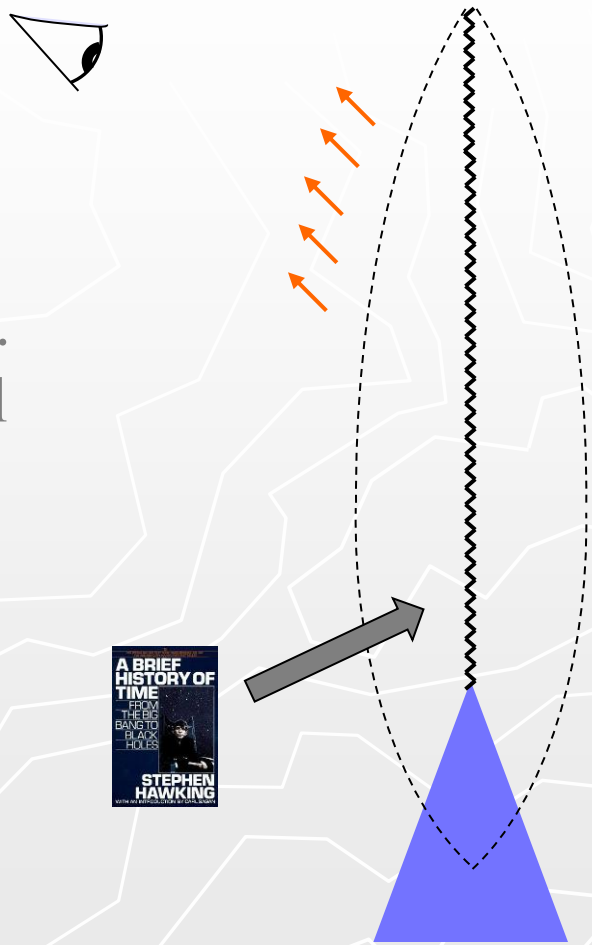


Black hole complementarity II

A further connotation: the external observer sees the 'stretched' horizon as a complicated dynamical system, which can absorb, thermalize, and reradiate information. Outside the stretched horizon, normal effective field theory applies.

The infalling observer sees nothing special at the horizon.

Challenge: to make a concrete model (cf. Mathur, Giddings). Instead, a no-go argument.

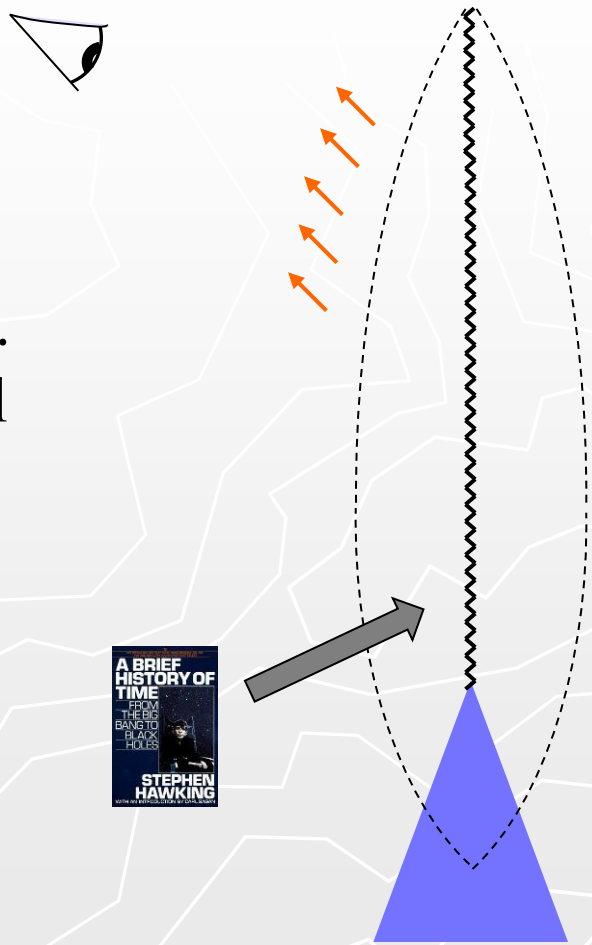


Black hole complementarity II

A further connotation: the external observer sees the 'stretched' horizon as a complicated dynamical system, which can absorb, thermalize, and reradiate information. Outside the stretched horizon, normal effective field theory applies.

The infalling observer sees nothing special at the horizon.

Challenge: to make a concrete model (cf. Mathur, Giddings). Instead, a no-go argument.

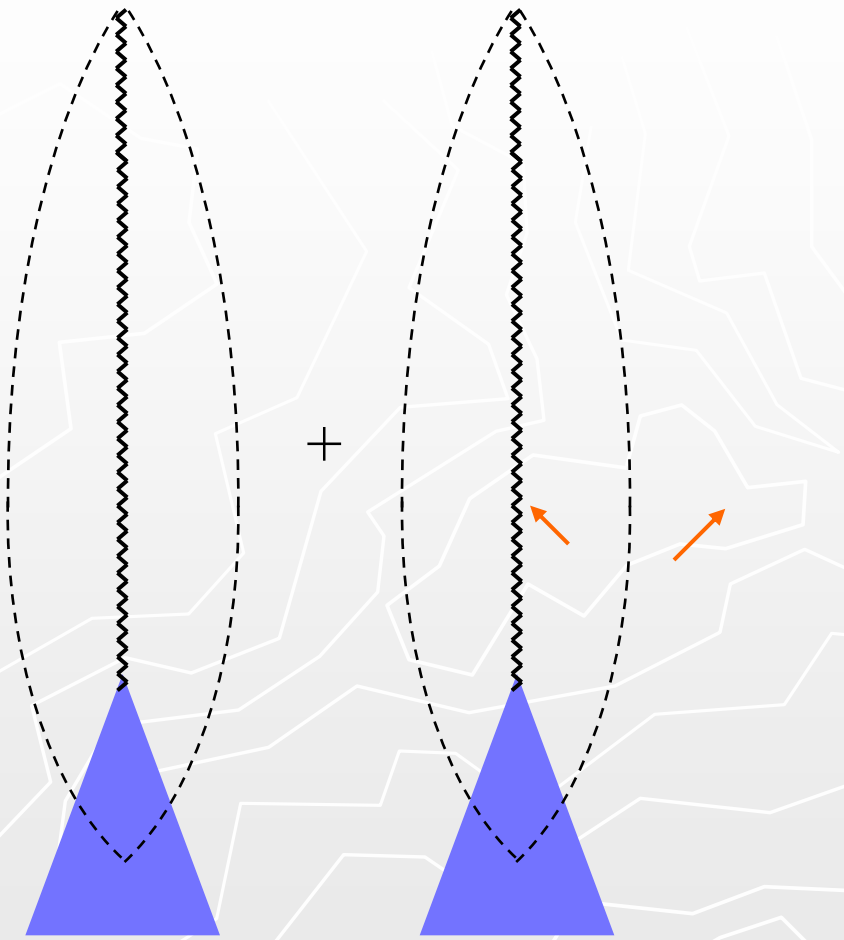


Hawking's argument:

The Hawking process is a quantum effect, and produces a superposition,

$$\approx |0', 0\rangle + e^{-\omega/kT}|1', 1\rangle$$

The two photons are entangled; the outside (unprimed) photon by itself is in a mixed state.



Hawking's argument:

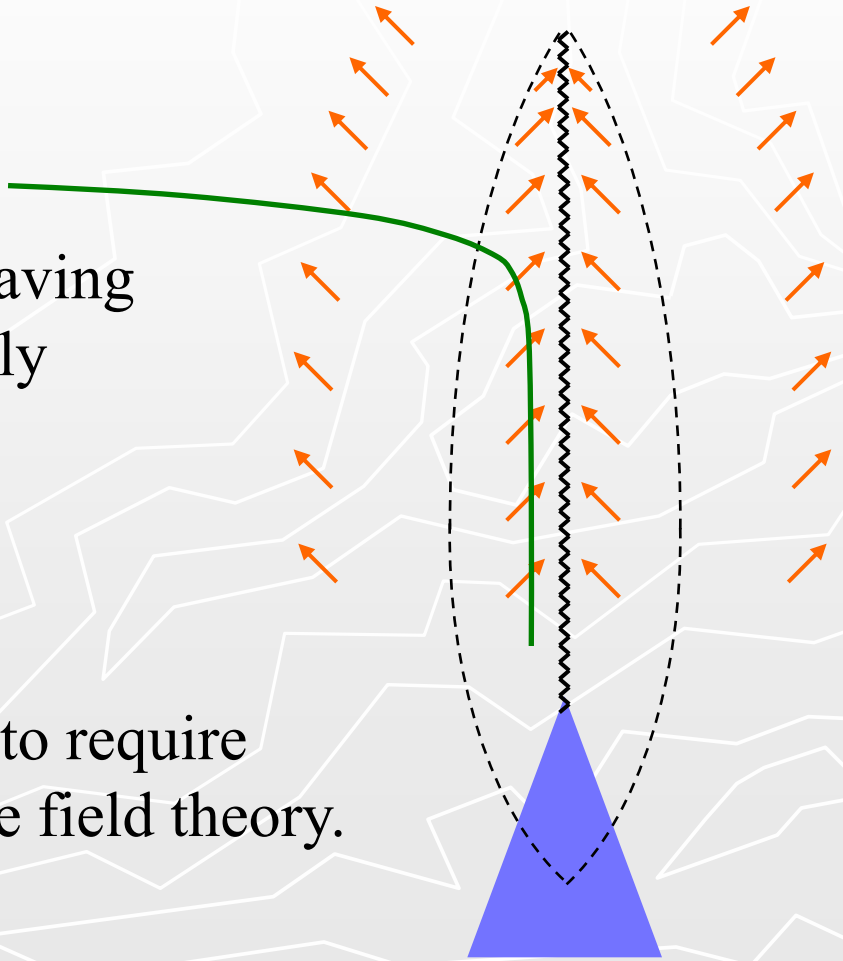
The net result is a highly entangled state, roughly

$$\left(|0',0\rangle + e^{-\omega/T}|1',1\rangle\right)\left(|0',0\rangle + e^{-\omega/T'}|1',1\rangle\right)\ldots$$

When the evaporation is completed, the inside (primed) degrees of freedom are gone, leaving the Hawking radiation in a highly mixed state.

Pure $\rightarrow$ mixed evolution.

To avoid this conclusion seems to require violation of low energy effective field theory.



$$\left(|0',0\rangle + e^{-\omega/T}|1',1\rangle\right)\left(|0',0\rangle + e^{-\omega/T'}|1',1\rangle\right)\ldots$$

Not sensitive to small corrections: we would need O(1) admixture of $|0',1\rangle, \quad |1',0\rangle$ at each step. Backreaction doesn't help:

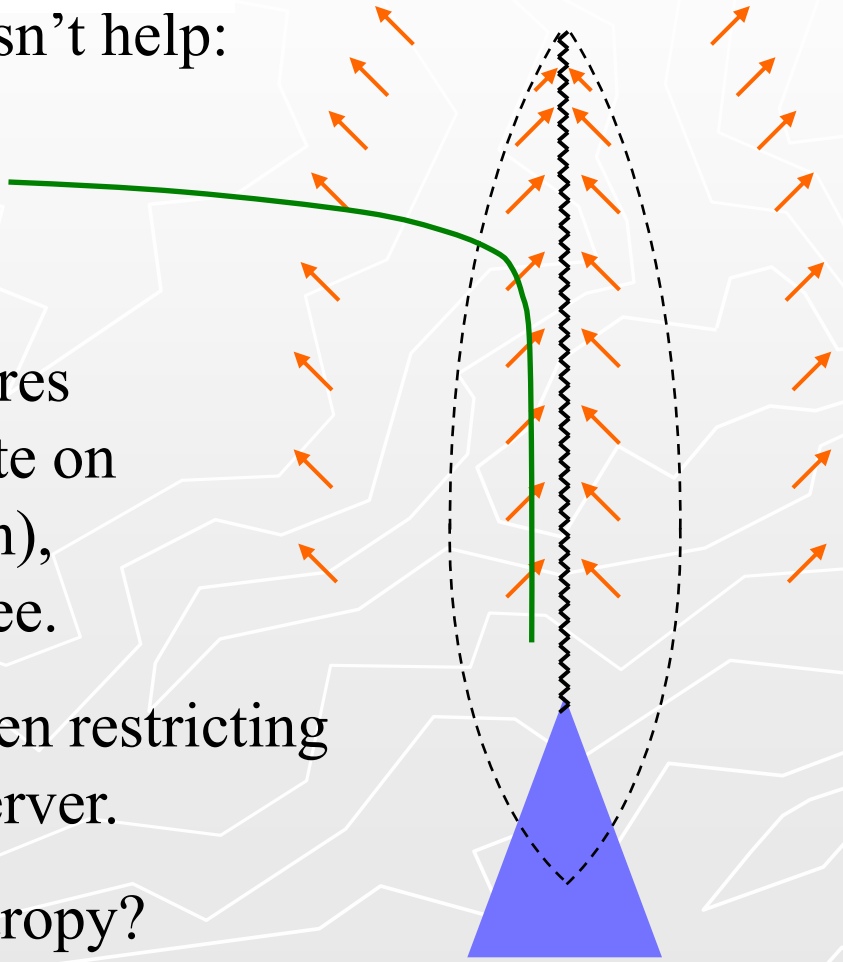
$$|0',0\rangle\psi_0 + |1',1\rangle\psi_1$$

is just as entangled.

This argument as it stands ignores complementarity: one has a state on the whole spacelike slice (green), which no single observer can see.

But one gets a contradiction even restricting to observations of a single observer.

p.s. Where is the black hole entropy?



Postulates of Black Hole Complementarity (STU, hep-th/9306069)

1) Unitarity: A distant observer sees a unitary S-matrix, which describes black hole evolution from infalling matter to outgoing Hawking-like radiation within standard quantum theory.

2) EFT: Outside the stretched horizon, physics can be described by an effective field theory of Einstein gravity plus matter.

3) The dimension of the subspace of states describing a black hole of mass M is $\exp S_{BH}(M)$.

4) No Drama: A freely falling observer experiences nothing out of the ordinary when crossing the horizon.

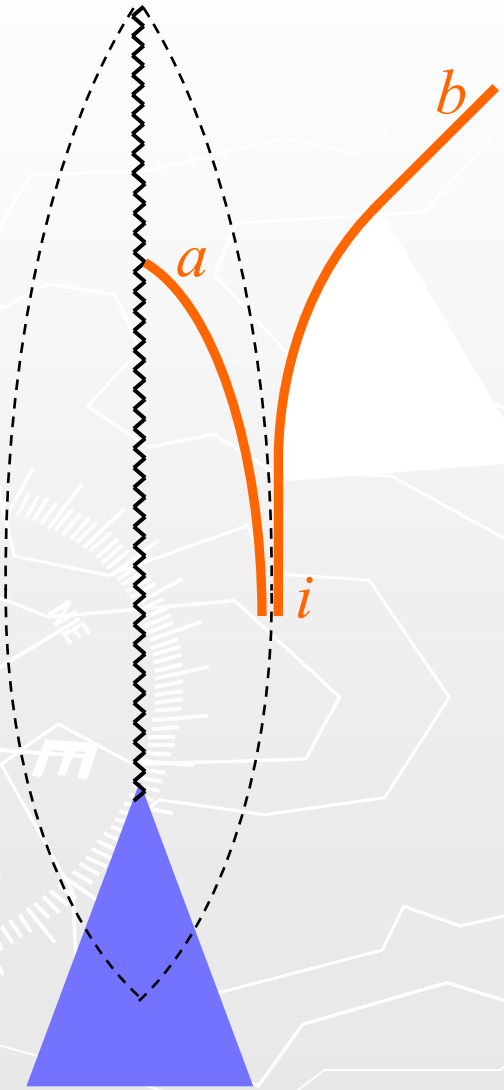
Almheiri, Marolf, Polchinski, Sully (AMPS 1207.3123):
Unitarity + EFT + No Drama are mutually inconsistent.

Note corollary: Maximal EFT implies information loss.

Consequences of No Drama + EFT

$$b = Ai + Bi^\dagger$$

$$i = Cb + Db^\dagger + C'a + D'a^\dagger$$



Creation/annihilation operators:

i: Inertial observer near horizon

b: Outgoing Hawking modes

a: Ingoing Hawking modes

Adiabatic principle/no drama:

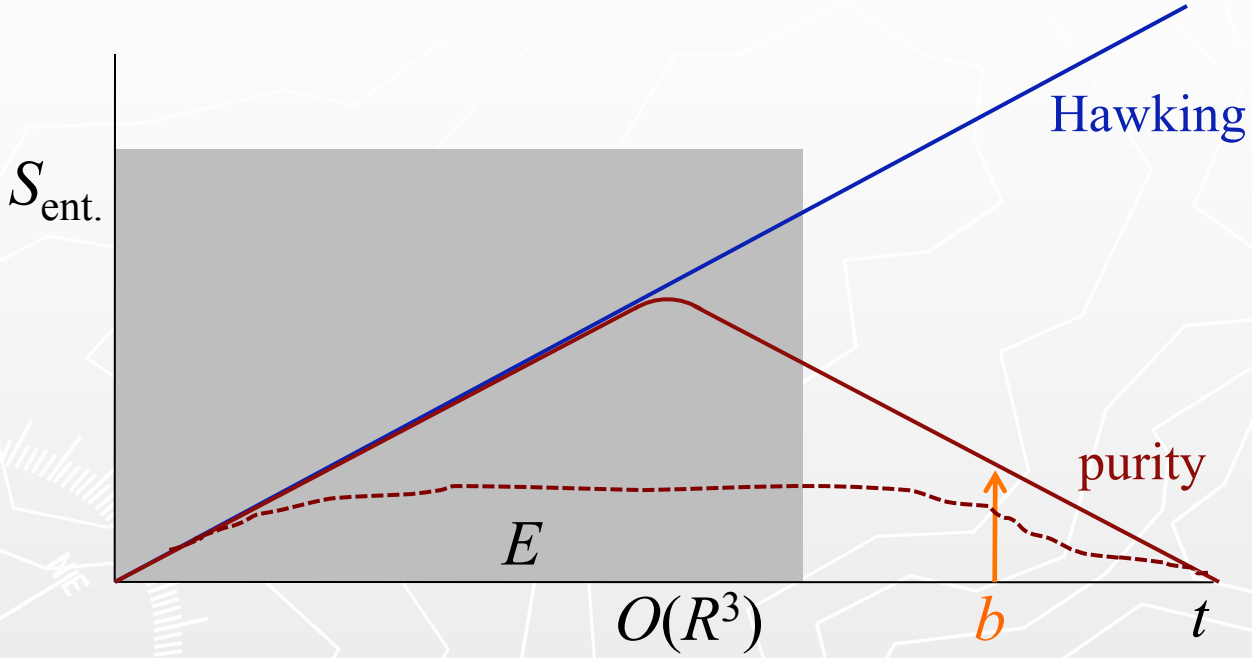
$$i|\psi\rangle = 0 \quad \text{so} \quad b|\psi\rangle \neq 0$$

This implies:

- Hawking radiation
- b and a are maximally entangled.

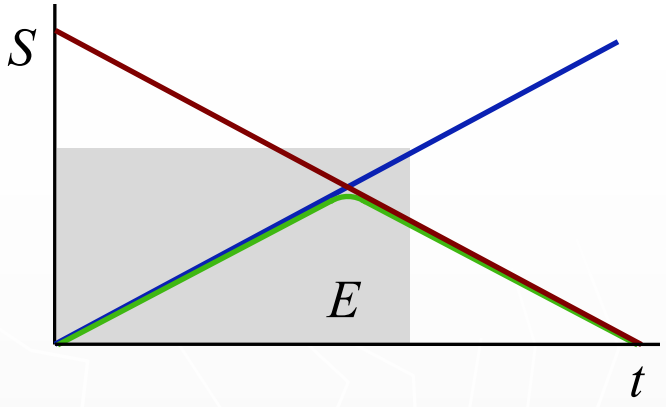
Consequences of *purity* (Page, Hayden & Preskill)

Entanglement entropy of Hawking radiation with black hole (= von Neumann entropy of HR and BH separately):



Consider the `early' Hawking radiation E, to somewhat *past* the turnover point. The state of a later Hawking mode is entangled with E (that is, b together with some sub-system b_E of E are in a pure state).

Consequences of unitarity:

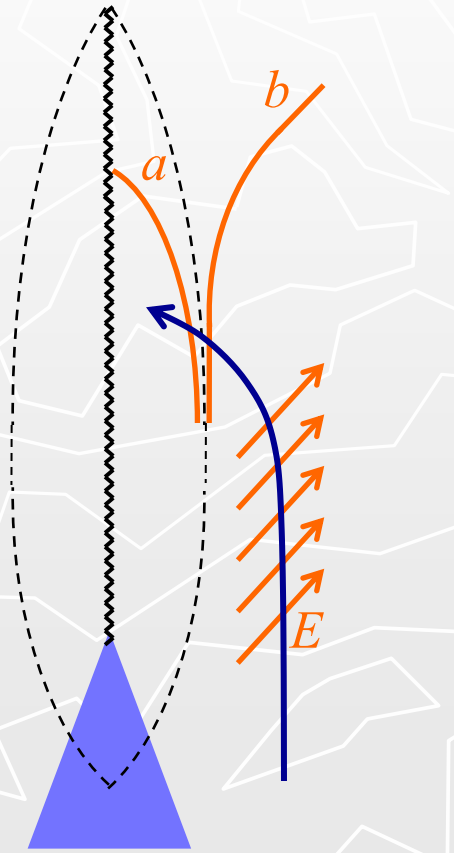


Purity: b is entangled (not necessarily maximally) with the early radiation E.

No drama: b is maximally entangled with a.

EFT: These are the same b.

Quantum mechanics doesn't allow this! Moreover, a single observer can interact with all three subsystems: E, b, and a.



Another way to state the problem:

Strong subadditivity requires (Mathur 0909.1038)

$$S_{ab} + S_{bE} \geq S_b + S_{abE}$$

No drama $\rightarrow S_{ab} = 0$ and so also $S_{abE} = S_E$. Then

$$S_{bE} \geq S_b + S_E$$

(no entanglement between b and E).
The Page curve and ignoring gray body factors gives

$$S_{bE} = S_E \square S_b$$

so we miss by a lot, but even weakening the Page assumption, and including GBF's, leaves b and E entangled.

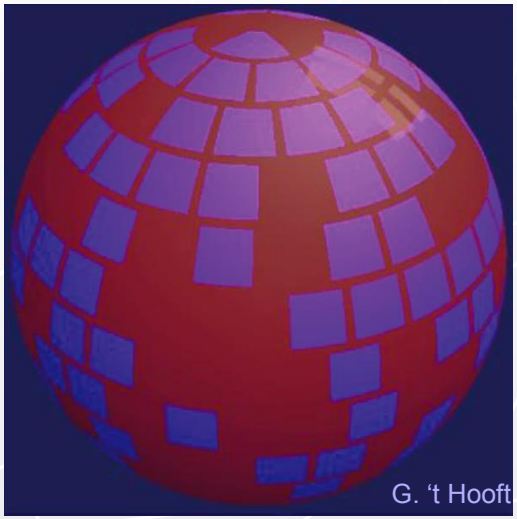
Mining the black hole:

The previous argument only applies to low partial waves, but one can do better:

Drop a box near to the horizon, let it fill with Unruh radiation, and pull it out. This defeats the centrifugal barrier, and we can make the same argument anywhere on the horizon. (Must lower the box slowly to avoid perturbing the black hole.)

So if we give up `no drama' we find excitations everywhere behind the horizon, a *firewall*. Cf. *energetic curtains*, Braunstein 0907.1190v1

Could the firewall really exist? Mechanisms: fuzzball, loss of self-entanglement of horizon (quantum memory of black hole fills up).



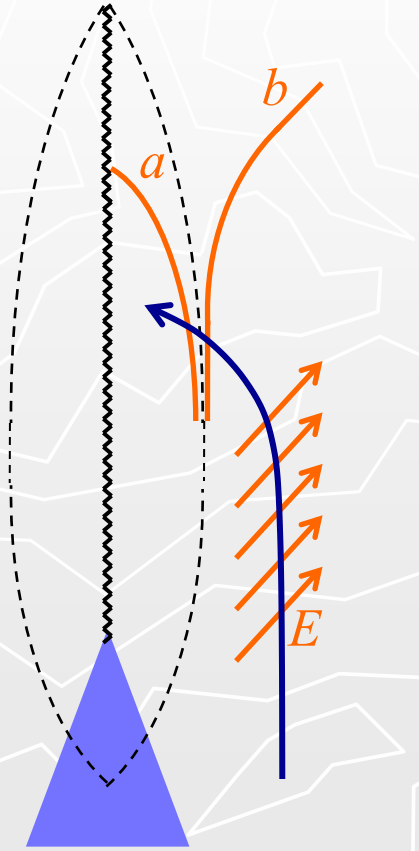
First we should try to save complementarity.

II. $a \in E$?

The paradox is that b is entangled with both a behind the horizon and the early radiation E. So maybe $a \in E$: isn't this what complementarity says? (Srednicki, Bousso, Harlow, Hayden, Nomura, Varela, Weinberg, Papadodimas, Raju, …)

Problem: a single observer can see all of a, b, E, so does not have a consistent quantum mechanics. By a quantum computation he can capture a single bit e_b entangled with b, and so have 3 bits in his lab in an impossible quantum state.

Possible out (Harlow & Hayden 1301.4504): may not be possible to carry out this computation. Will return to this later, but for now a different argument that does not require a quantum computation (AMPSS).

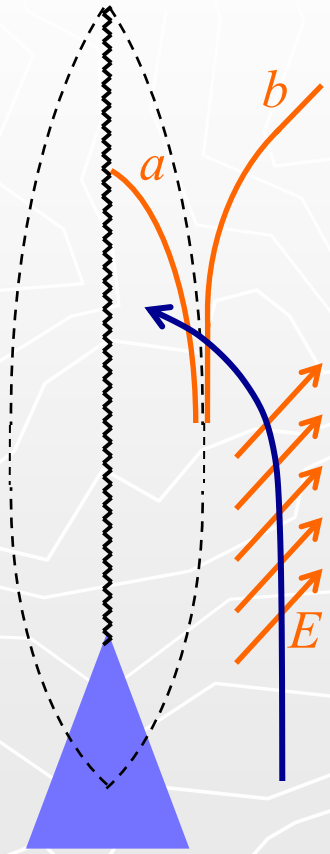


Suppose the infalling observer measures not the 'fine' bit e_b but some coarse bit e such as the state of a single Hawking photon.

Claim: $[e, a] = O(1)$ generically:

$$(-1)^{N_e} = \sigma^z \otimes I$$

$$(-1)^{N_a} = I \otimes S^0 + \sigma^x \otimes S^x + \sigma^y \otimes S^y + \sigma^z \otimes S^z$$



Averaging over embedding, all S^μ are of the order. Measuring e changes the state of every a with $O(1)$ probability (butterfly effect):

Measuring any single early bit *creates* a firewall if none is already there.

Perhaps $a \in E$ should be understood in some weaker sense, but what (note: depends on precised choice of E).

III. $a \in \mathcal{H}_{\rm CFT}$?

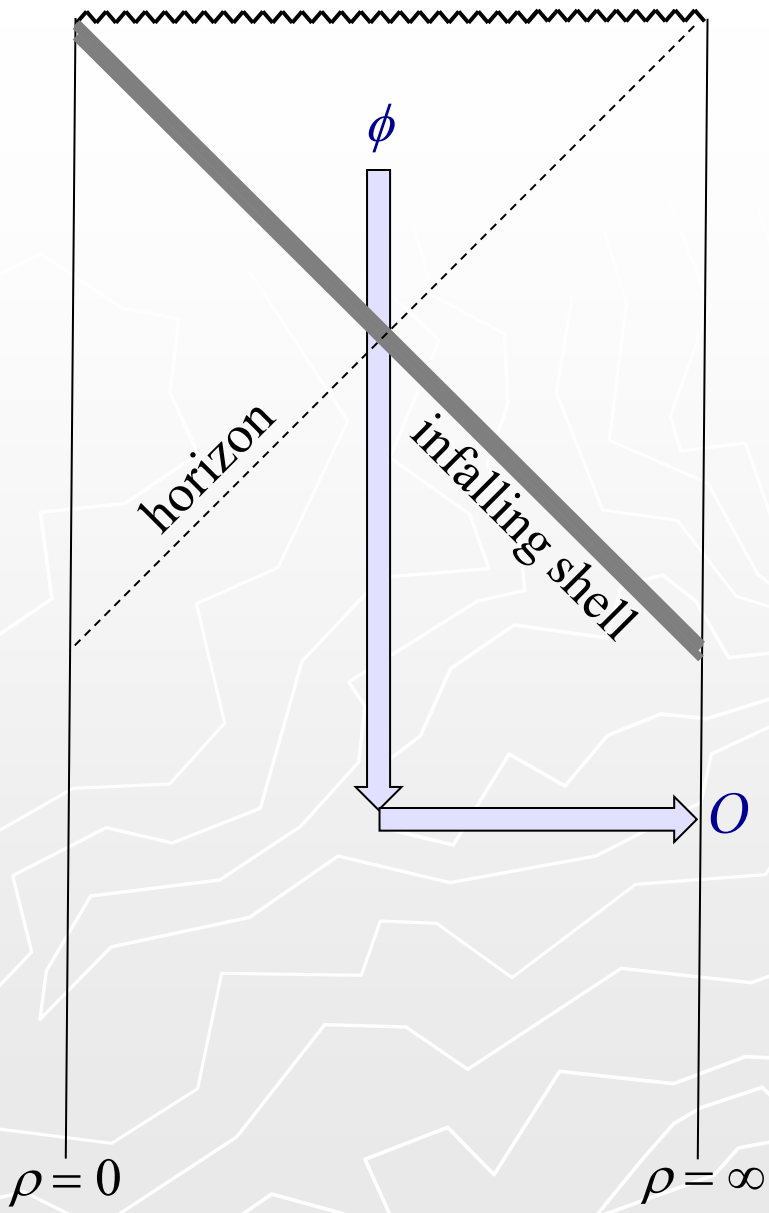
An observation independent of the firewall argument, but connected with it.

In AdS/CFT, the Gubser-Klebanov-Polyakov-Witten dictionary relates CFT fields to the boundary limit of bulk fields,

$$O(x) = \lim_{z\to 0} z^{\Delta}\phi(x,z)$$

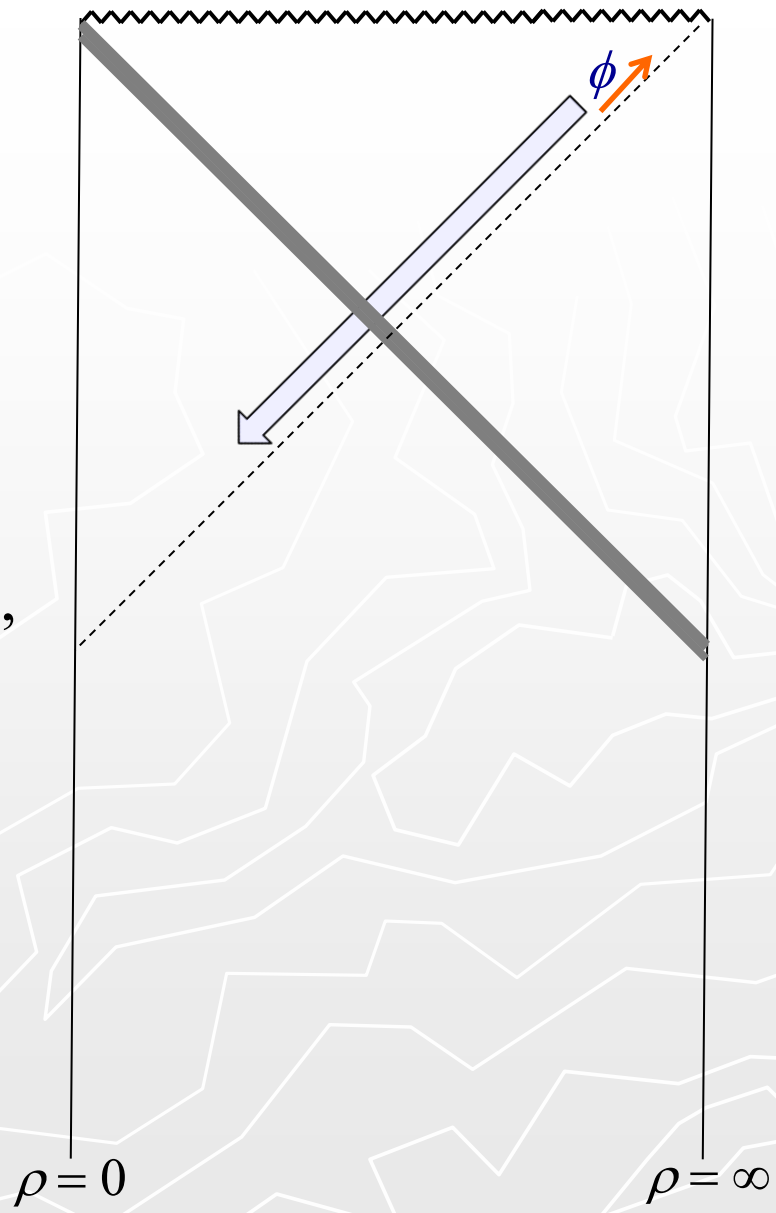
How to extend into the bulk, and behind the horizon? One approach (Banks, Douglas, Horowitz, Martinec, Balasubramanian, Kraus, Lawrence, Trivedi, Bena, Susskind, Freivogel, Hamilton, Kabat, Lifschytz, Lowe, Heemskerk, Marolf, Polchinski, Sully): extend using field equations.

For an AdS black hole formed from infalling matter, one can integrate back to before the collapse and then out to the boundary (Susskind & Freivogel; Heemskerk, Marolf, Polchinski, Sully.)



For an AdS black hole formed from infalling matter, one can integrate back to before the collapse and then out to the boundary (Susskind & Freivogel; Heemskerk, Marolf, Polchinski, Sully.)

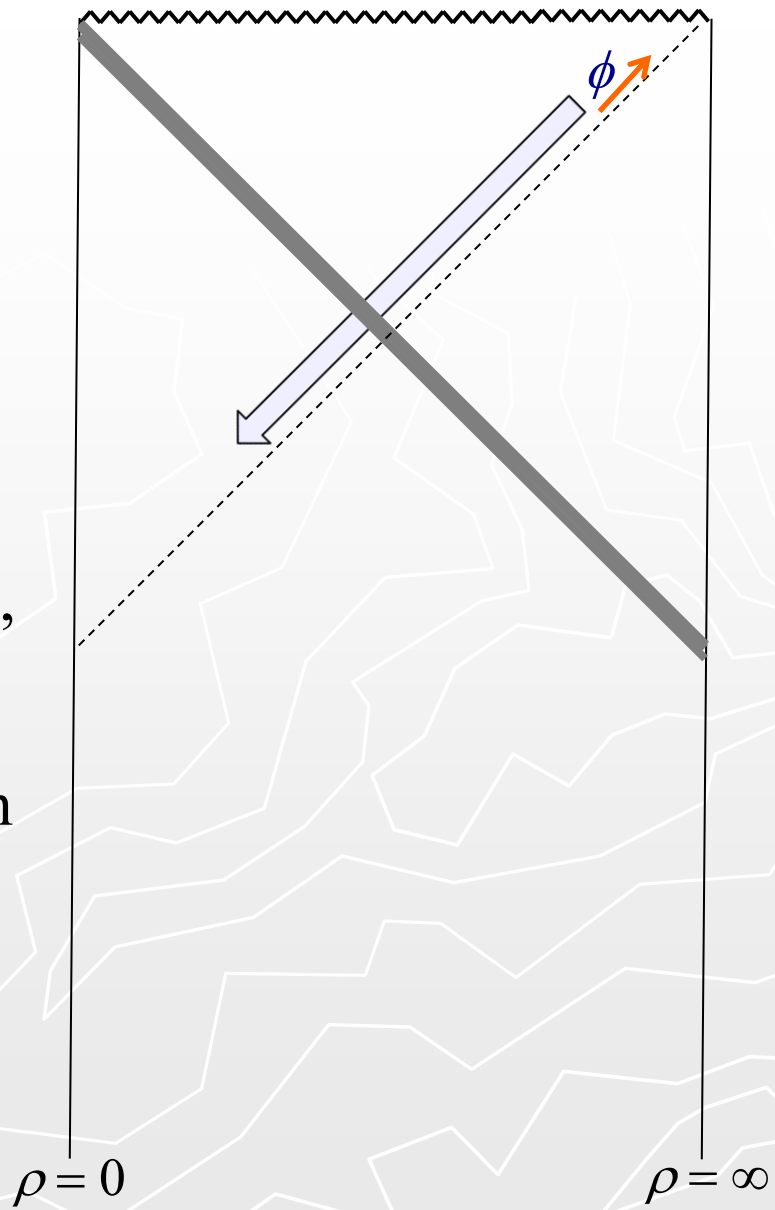
Problem: after the scrambling time, this leads to trans-Planckian physics, and can't be done explicitly.



For an AdS black hole formed from infalling matter, one can integrate back to before the collapse and then out to the boundary (Susskind & Freivogel; Heemskerk, Marolf, Polchinski, Sully.)

Problem: after the scrambling time, this leads to trans-Planckian physics, and can't be done explicitly.

In fact, ϕ cannot be constructed even in principle, in a generic black hole state: AdS/CFT describes the black hole interior (if there is one) less completely than might have been assumed.



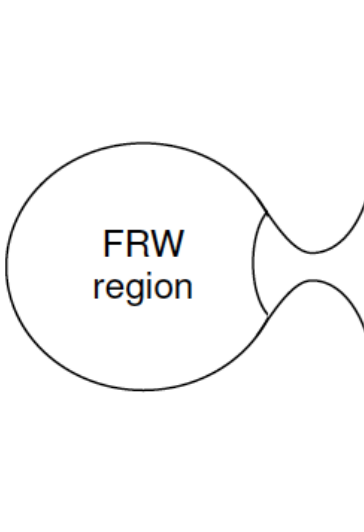
$a \notin \mathcal{H}_{\rm CFT}$: why?

Consider all AdS states $|I\rangle$ with $M < E < M + \delta M$, where $M > T_{\rm HP}$ so these look like black holes. Consider the states $a^\dagger|I\rangle$. But *interior Hawking modes have negative global energy*, so states $a^\dagger|I\rangle$ have $M-\omega < E < M-\omega + \delta M$ and so are fewer in number by $e^{-\beta\omega}$: $a^\dagger$ must annihilate the remainder. But $a^\dagger$ has a left inverse, $a/(N+1)$ so this is inconsistent: $a^\dagger$ can have no image in the CFT.

What does this mean? (Doesn't imply firewall, though it rules out some alternatives).

There is no process that can form states with just a single $a^\dagger$ excitation. The CFT describes only those states that can actually form.

(Compare Bag of Gold):



But an infalling observer wanting to describe local physics with QFT would need $a^\dagger$.

'Strong complementarity.' If there is to be an interior, it must be that the only the exterior observers sees $|I>$, and only the interior observer sees $a^\dagger$. (Or maybe superselection sectors, which commute with all CFT fields.) But this still does not resolve the basic entanglment paradox.

Interior spacetime from entanglement (Papadodimas+Raju 1211.6767, Verlinde^2 1211.6913):

Fields behind the horizon are entangled with fields b outside. Use this to identify them: for any high energy state,

$$|I\rangle = \sum e^{-\beta E_{\boldsymbol{n}}/2} |I, \boldsymbol{n}\rangle_H |\boldsymbol{n}\rangle_Z$$

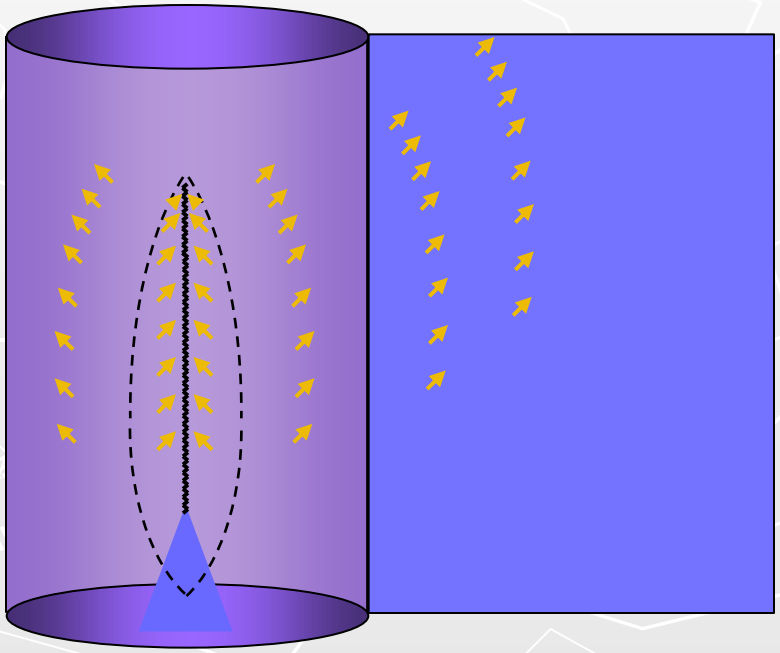
Projecting onto given $|\boldsymbol{n}\rangle_Z$, we can identify states $|I, \boldsymbol{n}\rangle_H$ with given excitation behind the horizon, and so define operators in the internal QFT. Problem: depends on choice of $|I\rangle$, e.g $a^\dagger(I)$, and so is ambiguous (consider a unitary $e^{i\theta n_b}$).

VV restrict $|I\rangle$ to 'code subspace' - how is this chosen?

IV. Sharpening the argument with AdS/CFT.

Original information problem is sharpest for large black hole in AdS (Maldacena hep-th/0106112).

Here too it helps. Large BH doesn't normally decay, but it does if we couple the boundary CFT to a second QFT with many degrees of freedom:



E.g. Rocha, 0905.4373

Advantages:

1. AdS potential confines c.m. of black hole (fluctuations ~ l_P), raised by Susskind, Yoneya, Hsu as obstacle.

2. Can take H_{CFT} to $\lambda(t)H_{CFT}$, and by making $\lambda(t)$ very small shut off CFT evolution while processing early radiation to extract e_b, so no limits from computation.

V. Nonviolent nonlocality (Giddings)

Three problems:

1. (EFT + nonlocal interactions) still has $a^\dagger$, but we have seen that this is forbidden.

2. Statistical mechanics links thermal emission to reflectivity, changing the Hawking radiation would change observable properties.

3. Falsified by mining experiments: by manipulating the outgoing Hawking bits we can turn an interaction that is supposed to reduce the black hole's entanglement entropy into one that increases it.

A simple model: Consider a basis $|\psi,h,a,b\rangle$ where a and b are the inner and outer Hawking bits, and h and ψ are 1 and $N-1$ bits from the rest of the black hole Hilbert space. Suppose that as b moves through the zone this state evolves to $|\psi,b,a,h\rangle$, swapping b and h. The ab entanglement is now internal to the black hole, and the entanglement with the early radiation is now transferred to the outmoving bit h. The black hole lies in an $N-1$ bit Hilbert space, consistent with its BH entropy.

But the mining apparatus can manipulate the outgoing bit, giving $|\psi,h,a,Ub\rangle$. The interaction takes this to $|\psi,Ub,a,h\rangle$. This requires an internal Hilbert space of more than $N-1$ bits, inconsistent with the black hole entropy.

A similar strategy works with other models.

Another apology:

I am sorry that no one has gotten rid of the firewall.
Please keep trying!