Black Holes: Complementarity or Firewalls?

Joseph Polchinski
Kavli Institute for Theoretical Physics
University of California at Santa Barbara

1207.3123, AMPS (Ahmed Almheiri, Don Marolf, JP, James Sully)

UC Davis, 2/25/13

The information paradox (Hawking, 1976): sharp conflict of QM vs. general relativity.

Paradox → AdS/CFT duality → partial resolution of paradox

Unanswered questions…

Now, a new paradox.

Outline

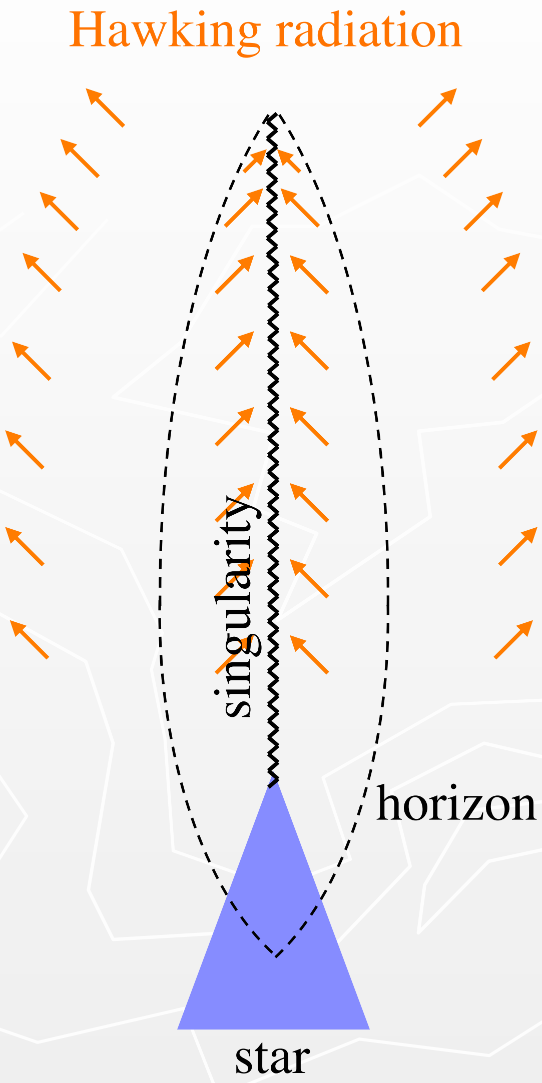
Review

- The information paradox
- AdS/CFT and black hole complementarity

The AMPS contradiction

What to give up? Possible resolutions

The life of a black hole:

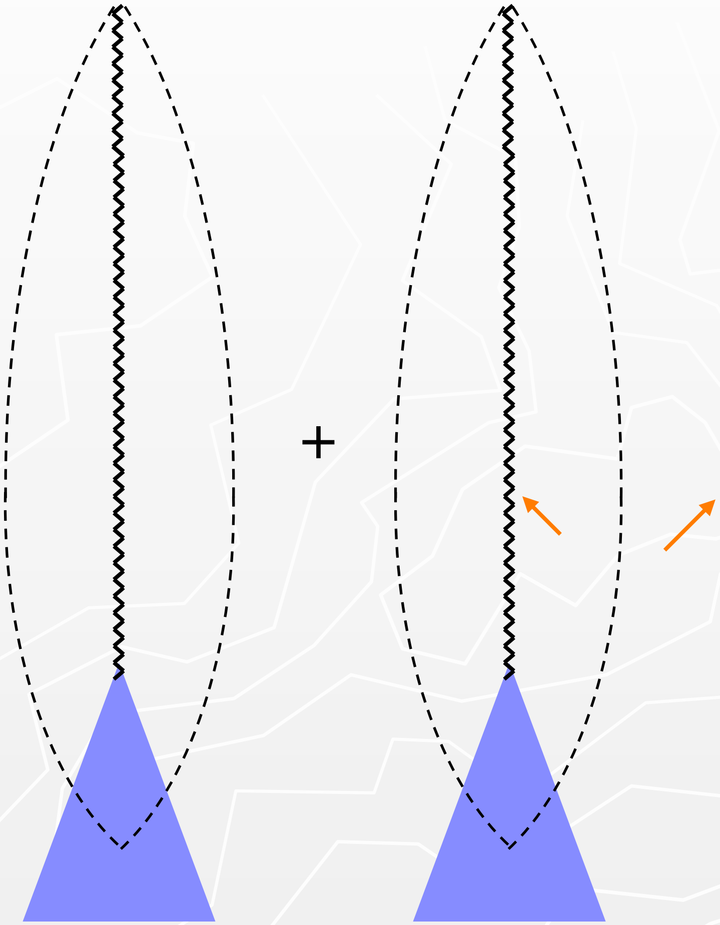


Hawking’s argument:

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

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

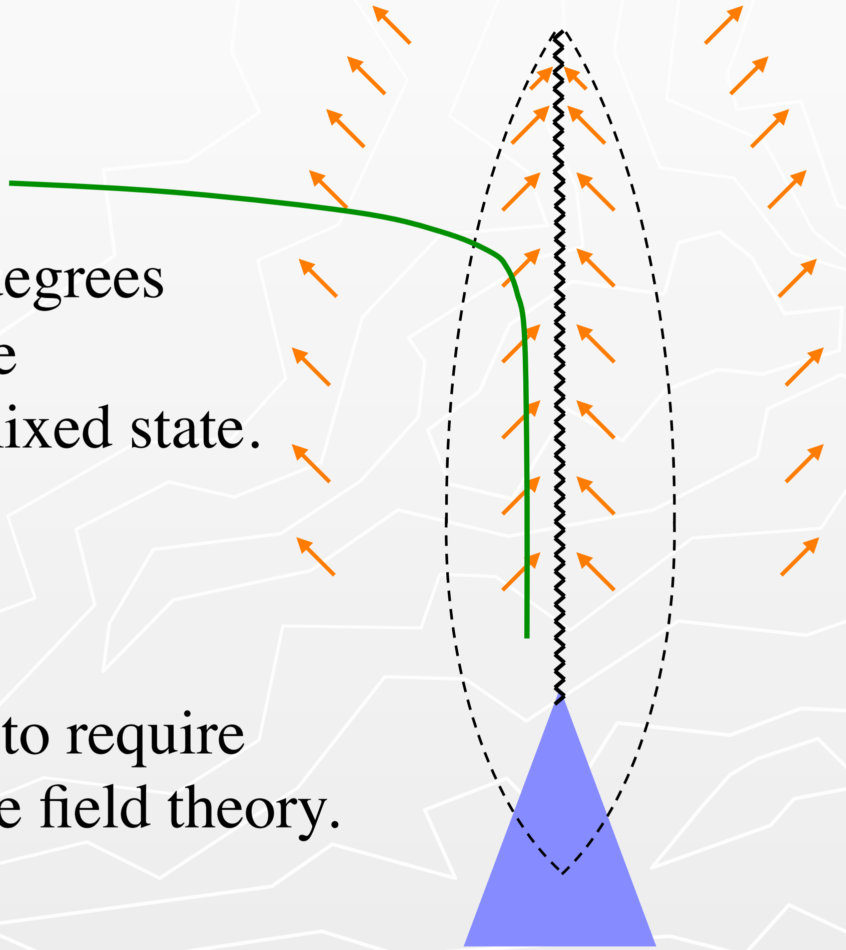
The two photons are entangled; the outside photon by itself is in a mixed state.



Hawking's argument:

The net result is a highly entangled state, roughly

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



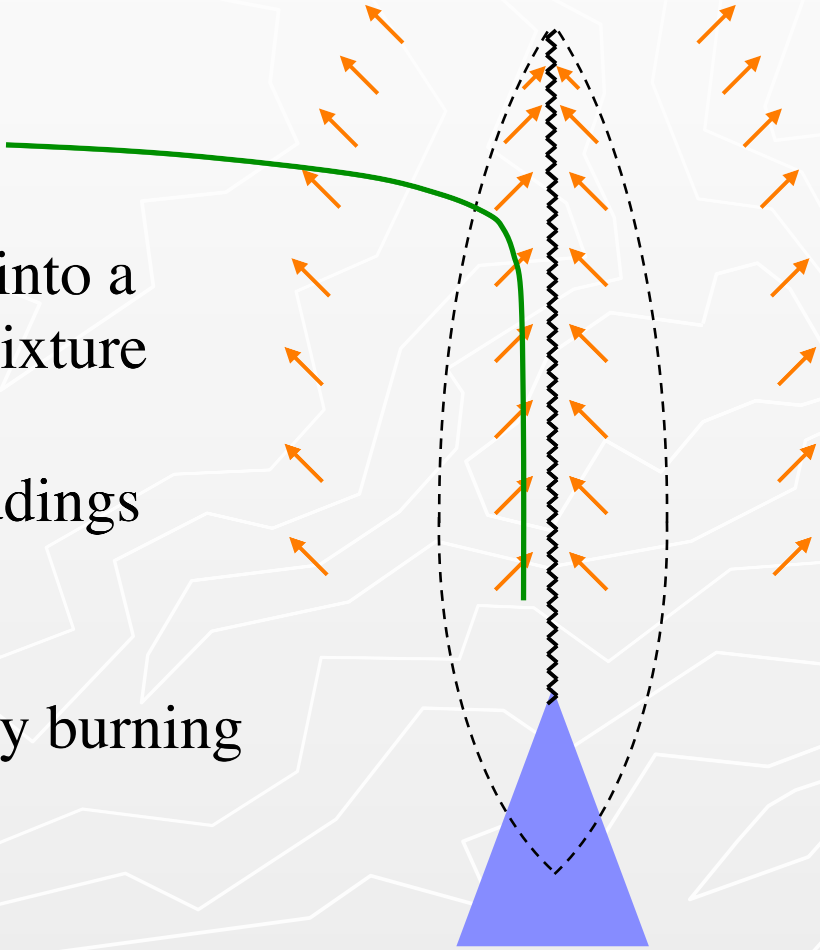
When the evaporation is completed,the inside (primed) degrees of freedom are gone, leaving the Hawkingradiation in a highly mixed state.

Pure → mixed evolution.

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

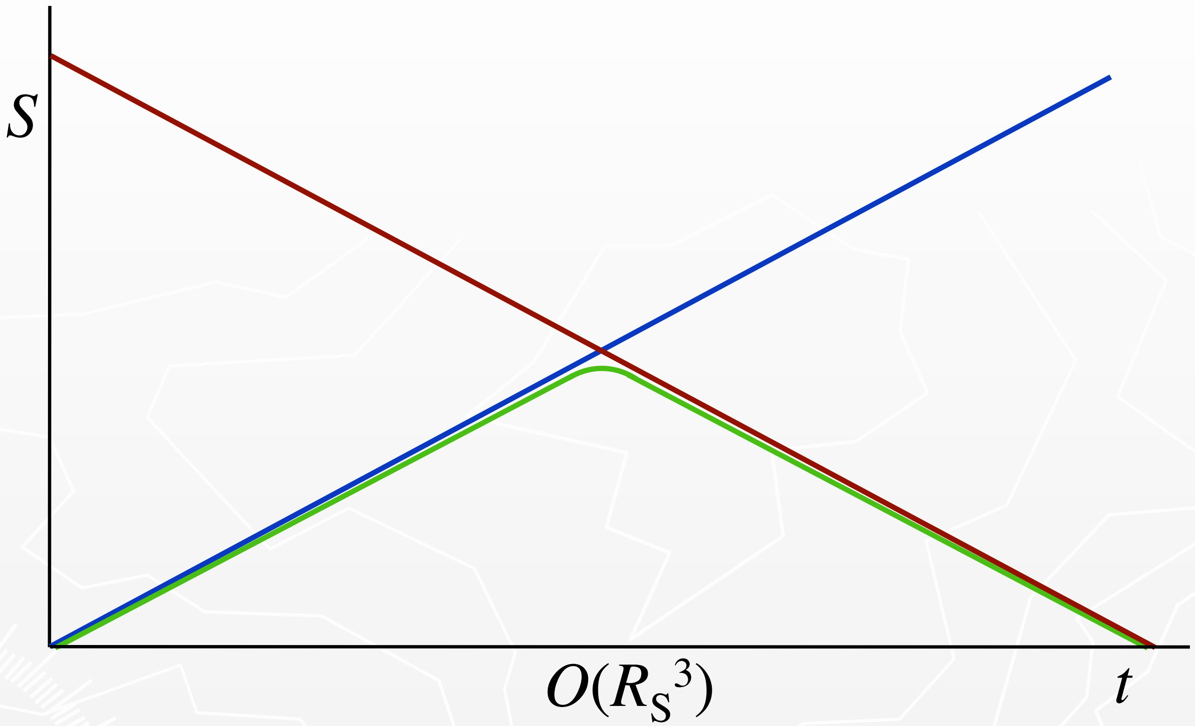
Counterintuition: for an ordinary burning object, the apparent evolution from a pure to mixed state is a result of coarse-graining, not an exact statement. Perhaps Hawking has missed subtle phases.

But the entanglement is not sensitive to small corrections: to turn $|0', 0\rangle + e^{-\omega/kT}|1', 1\rangle$ into a pure state would need O(1) admixture of $|0', 1\rangle, \quad |1', 0\rangle$ at each step (cf. Mathur 0909.1038, also Giddings 1108.2015).



The difference is that an ordinary burning object does not have a horizon.

The Page argument (hep-th/9306083):

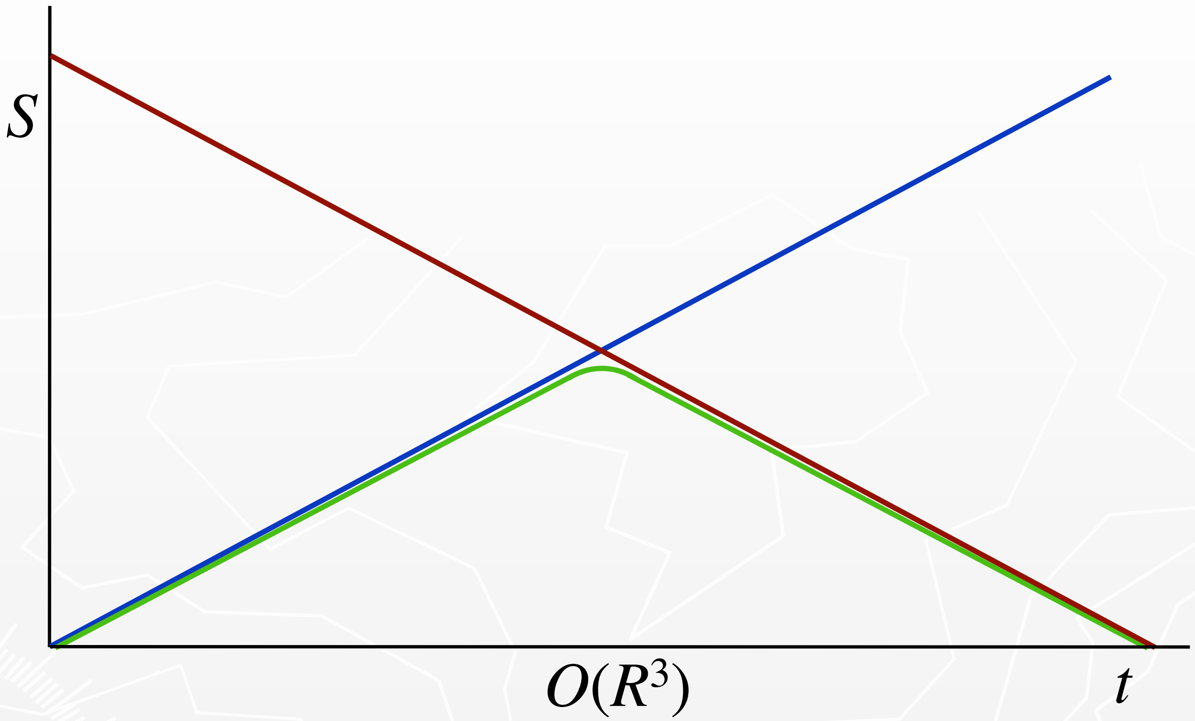


Blue: (S_{vN} of radiation) = ($S_{\mathrm{ent.}}$ of rad. + BH) ≤ (S_{vN} of BH), according to Hawking ’76.

Red: S_{coarse} (= $S_{\mathrm{Bekenstein\text{-}Hawking}}$) of evaporating black hole.

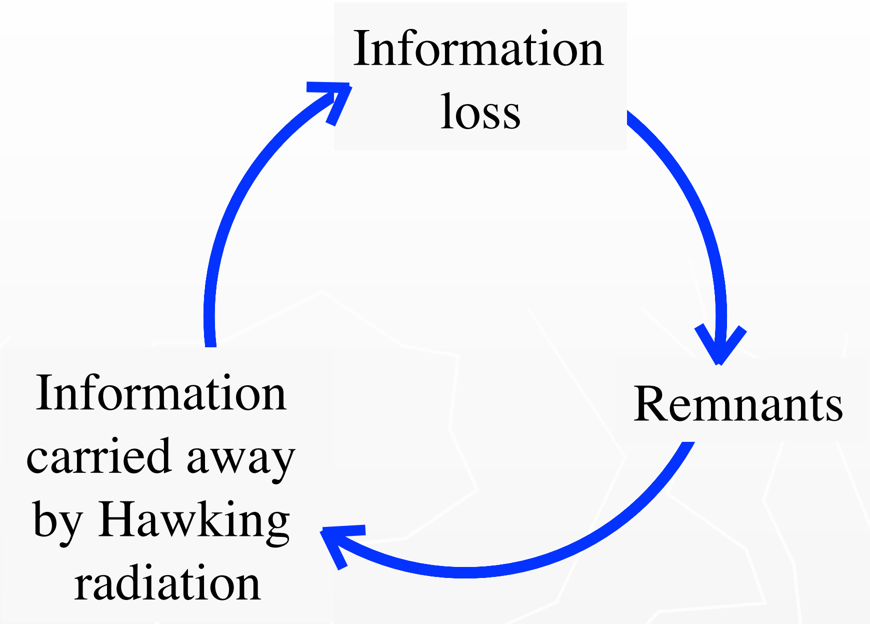
Deviations from Hawking must begin when the black hole is still large (green).

Asides: $R_S = M$ in Planck units, remnants, A vs. $A^{3/4}$



Old vs. young black holes. Hayden and Preskill 0708.4025: a message thrown into an old black hole can be read almost immediately from the Hawking radiation, if we have collected the first half+$\varepsilon$of the radiation. Early radiation purifies the late radiation, $|\psi\rangle = \Sigma_i\, |E_i\rangle |F_i\rangle$

Going around in circles (1976-97):



Information loss

Remnants

Information carried away by Hawking radiation

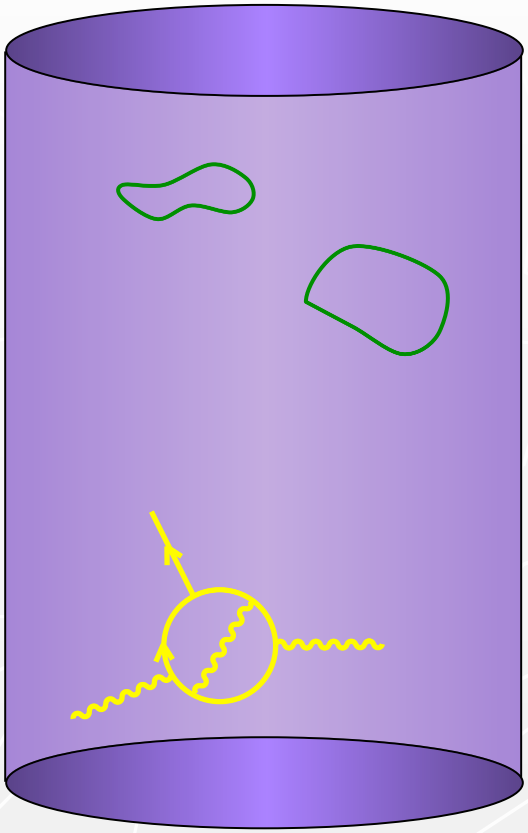
1. Information loss. Not like ordinary thermal objects. Large violations of energy conservation.
2. Remnants. Not like ordinary thermal objects. Out of control virtual effects.
3. Unitarity of Hawking radiation: Like ordinary thermal objects, but requires large violation of locality.

AdS/CFT duality:

Quantum gravity (actually string theory) in an anti-de Sitter box.

=

A quantum field theory of gauge fields, fermions, and scalars living on the *surface* of the box.

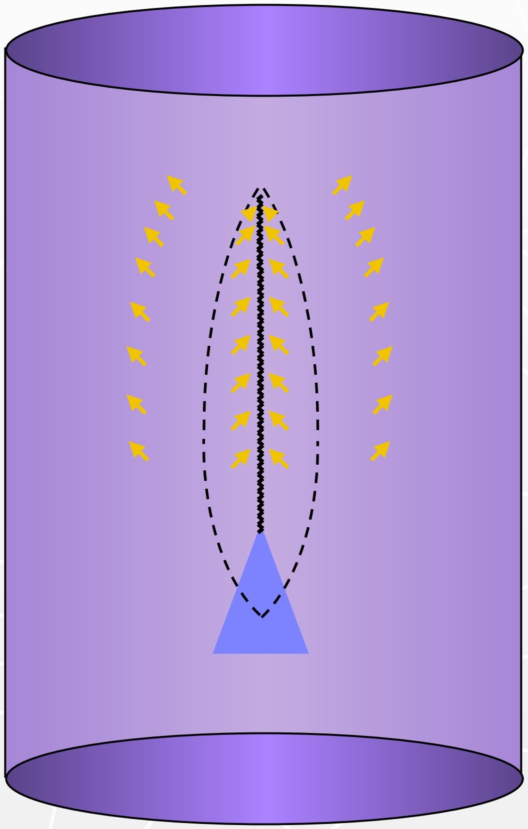


What does this say about the information paradox?

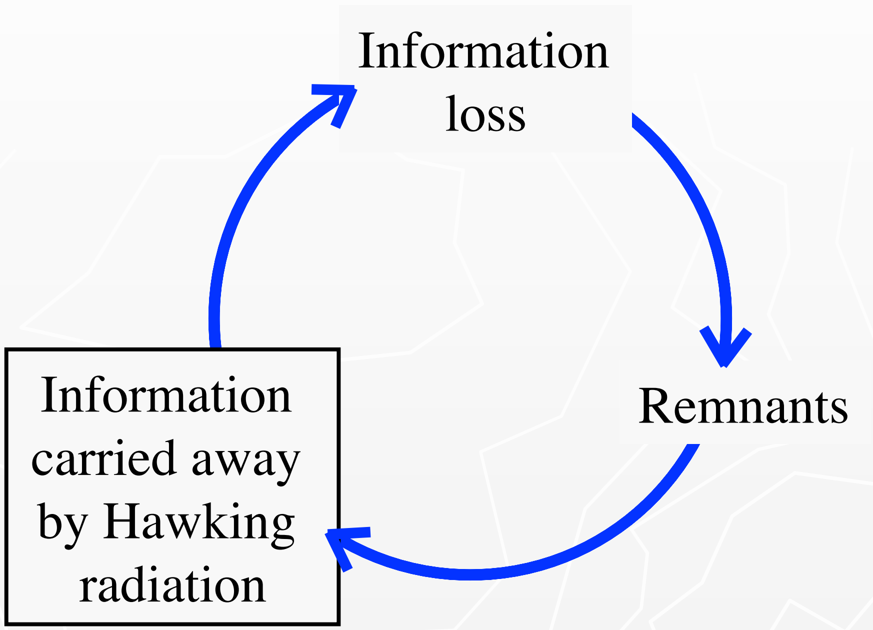
We can consider the Hawking experiment in an AdS box. Since the dual quantum field theory is described by ordinary QM, pure states must evolve to pure states.

Moreover, as anticipated, the dual description is highly nonlocal, and holographic.

Black hole = thermal state of CFT.



The winner!



Information loss

Remnants

Information carried away by Hawking radiation

What we would still like to understand:

How does the nonlocality appear?

How do we extend the holographic construction of gravity to other kinds of spacetime? The black hole interior is much like a cosmology, so it would be useful to understand how this is represented in AdS/CFT.

Forms of nonlocality:

1. Explicit nonlocal interaction.

2. Black hole complementarity: a new relativity principle ('t Hooft/Preskill/Susskind '93). Infalling observer sees bit inside, asymptotic observer sees same bit outside, no one sees both (would violate linearity of quantum mechanics).

   External observer perceives the horizon as a dynamical membrane which re-radiates radiation, while infalling observer passes through freely.

Postulates of Black Hole Complementarity (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.

Some roots of this work:

Mathur’s theorem, that purity implies that the horizon cannot be `information-free.’ I.e. *O*(1) breakdown of QFT in curved spacetime.

Giddings’s models of nonlocal interactions extending well outside the horizon.

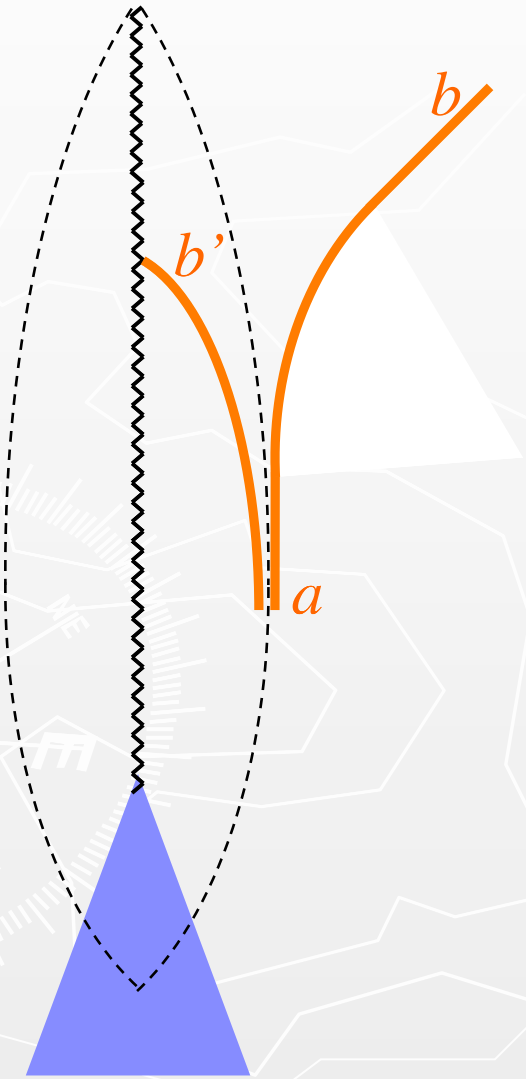
These were in the context of non-complementary `bit models,’ but imposing complementarity does not help. We have extended their arguments, and would claim even stronger conclusions.

Also: S. Braunstein 0907.1190

Consequences of No Drama + EFT

$$b = Aa + Ba^\dagger$$

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



Creation/annihilation operators:

a: Inertial observer near horizon

b: Outgoing Hawking modes

b': Ingoing Hawking modes

Adiabatic principle/no drama:

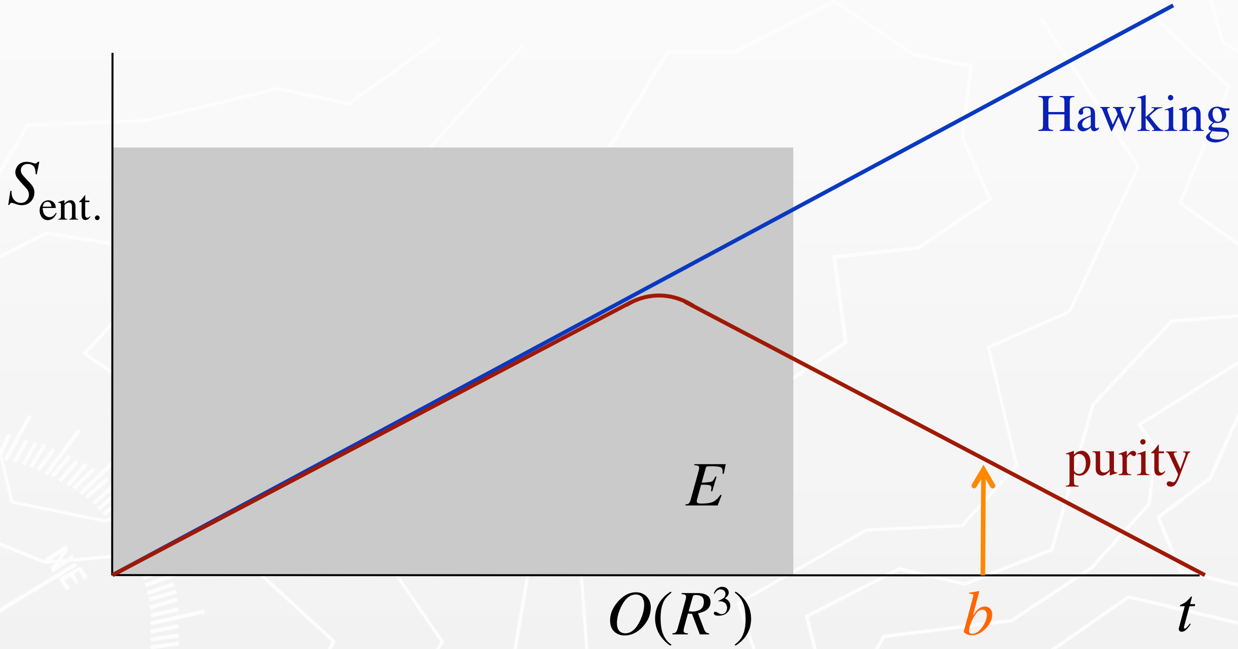
$$a|\psi\rangle = 0 \quad \text{so} \quad b|\psi\rangle \neq 0$$

This implies:

- Hawking radiation
- *b* and *b'* 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):



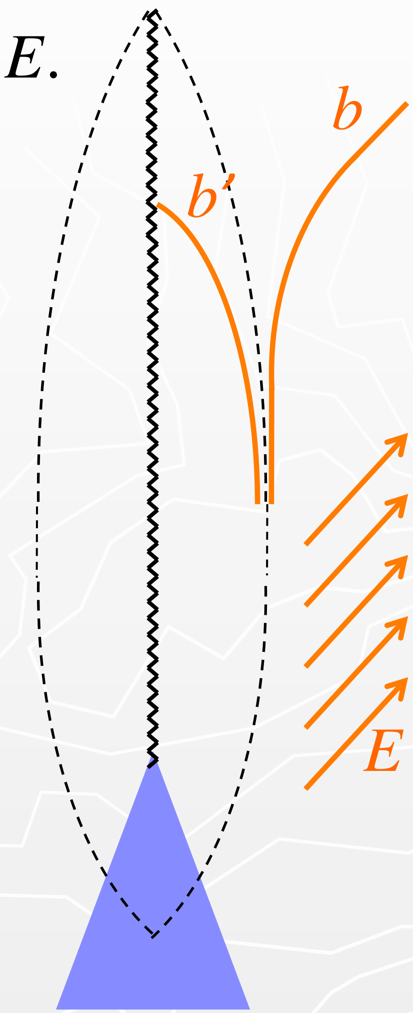
Consider the `early' Hawking radiation E, to somewhat *past* the turnover point. The state of a later Hawking mode is fully entangled with E (that is, b together with some subsystem b_E of E are in a pure state).

A contradiction:

Purity: *b* is entangled with the early radiation *E*.

No drama: *b* + *b'* are in a pure state.

EFT: These are the same *b*.



Quantum mechanics doesn't allow this!

e.g. $(|00> + |11>)\ |E>$

vs. $|00>\ |E_1> + |11>\ |E_2>$

Moreover, a single observer can interaction with all three subsystems: the early radiation, *b*, and *b'*.

Another way to state the problem:

Strong subadditivity requires (Mathur)

$$S_{b'b} + S_{bE} \geq S_b + S_{b'bE}$$

No drama → $S_{b'b} = 0$ and so also $S_{b'bE} = S_E$. Then

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

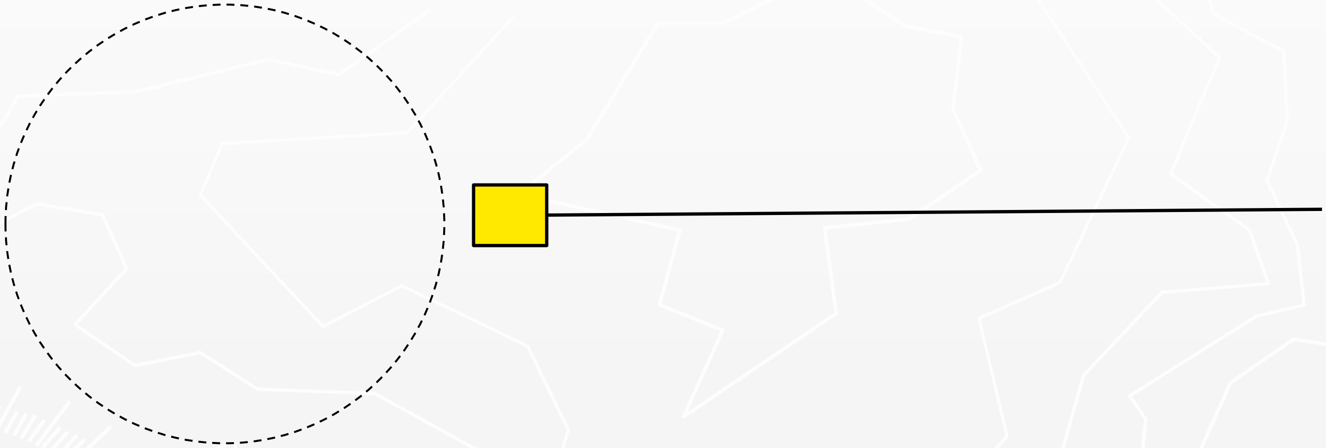
(no entanglement between *b* and *E*).
The Page curve plus ignoring gray body factors gives

$$S_{bE} = S_E - S_b$$

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*.

What to give up?

Unitarity? Not consistent with AdS/CFT.

Absence of drama?

EFT outside the horizon?

Other implicit assumption (e.g. quantum mechanics)?

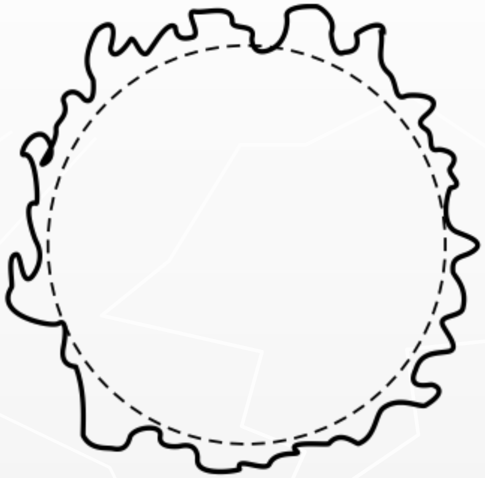
Giving up ‘No Drama’?

How bad is it - what energy excitations, and how many?

Energy is limited only by the assumed cutoff on EFT.

The first argument only applies to low angular momenta, due to a centrifugal barrier, but the mining argument applies to all *L*: the infalling observer encounters a Planck-scale *firewall*. A radical conclusion.

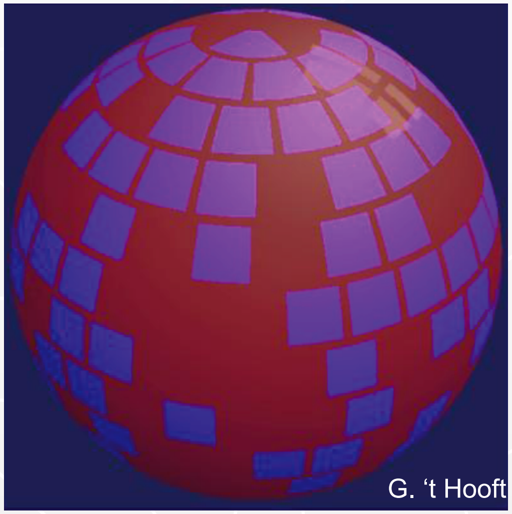
Is the firewall just the fuzzball?

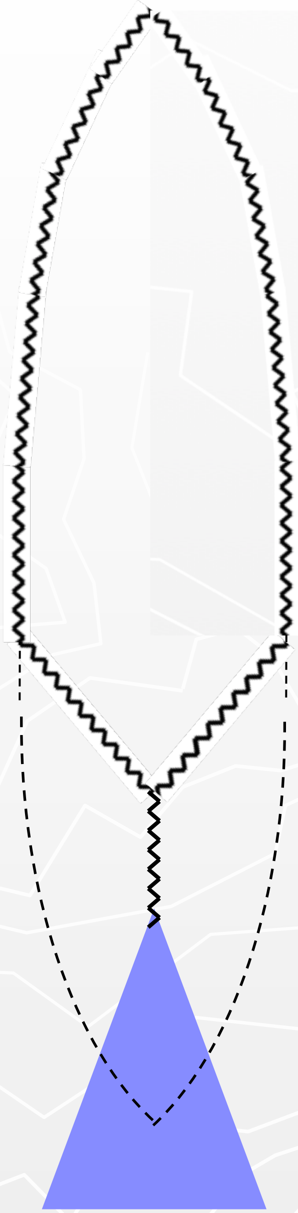


Scenario (Mathur): branes tunnel out to Schwarzschild radius (large number of configurations offsets small amplitude).

Extension of singularity (Susskind 1210.2098):

Idea: black hole interior is constructed from entanglement of bits on horizon,





After Page time, no self-entanglement so no interior. Singlurity expands out to horizon.

Giving up EFT outside the horizon?

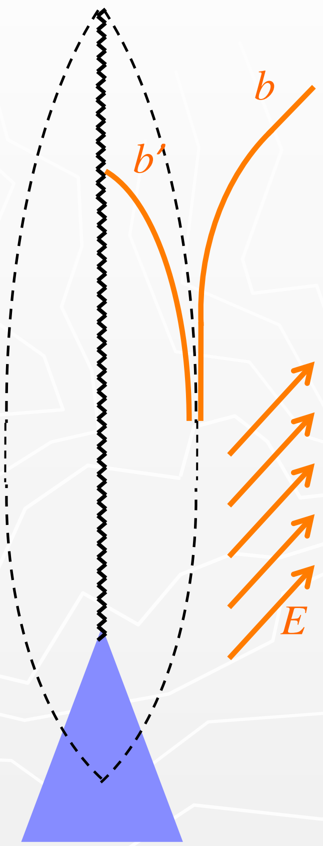
It is usually expected that effects that restore purity are small, $O(e^{-S})$, or involve $O(S)$ particles, or involve $O(e^{S})$ time scales. Here, the 2-point function of b changes by an amount of order 1 over a time of order $R \ln S$.

$A = E$? (Srednicki, Bousso, Nomura, Papadapoulos & Raju, Verlinde[2] ...)

A = interior, E = early radiation

Is there one large Hilbert space, which contains both b' and E? In this case quantum computation is not necessary: if e is a simple bit from the early radiation, $[e, b] = O(1)$, so *any* measurement of E will create a firewall.

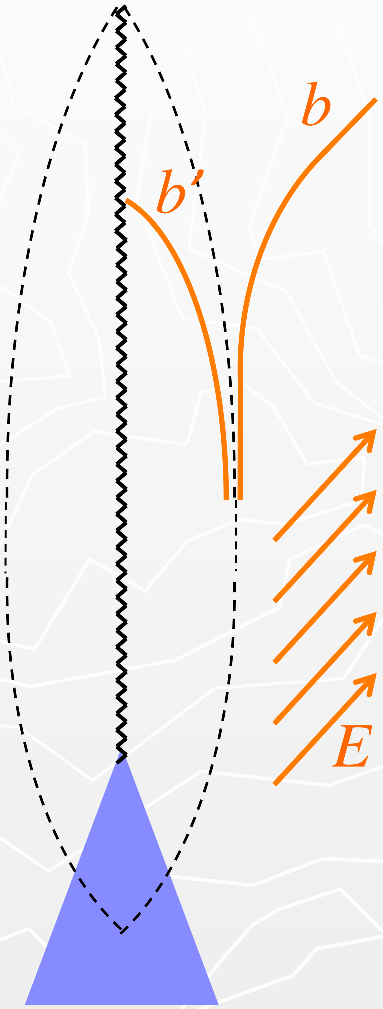
Alternative: 'strong complementarity,' e.g. Banks and Fischler There is no single Hilbert space (e.g. that of the CFT) in which the interior Hilbert space can be embedded. Not yet well-formulated.



Giving up EFT outside the horizon?

Problem: a single observer can see all of *b*, *b'*, *E*, so does not have a consistent quantum mechanics.

Possible out (Harlow & Hayden 1301.4504): it is not possible to do the quantum computation necessary to measure the entanglement between *b* and *E*, before falling into the black hole (?!). (Counterargument from AdS/CFT).

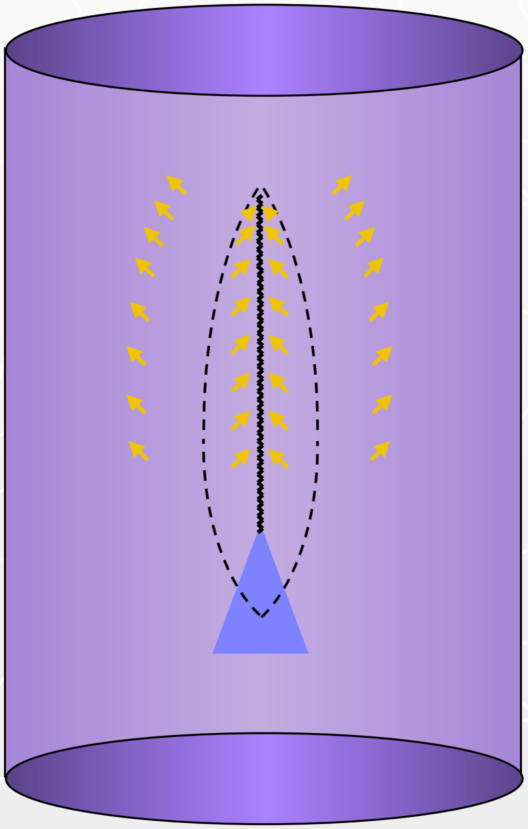


Old problem: construct fields in interior of AdS, and interior of black hole, in terms of CFT operators.

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

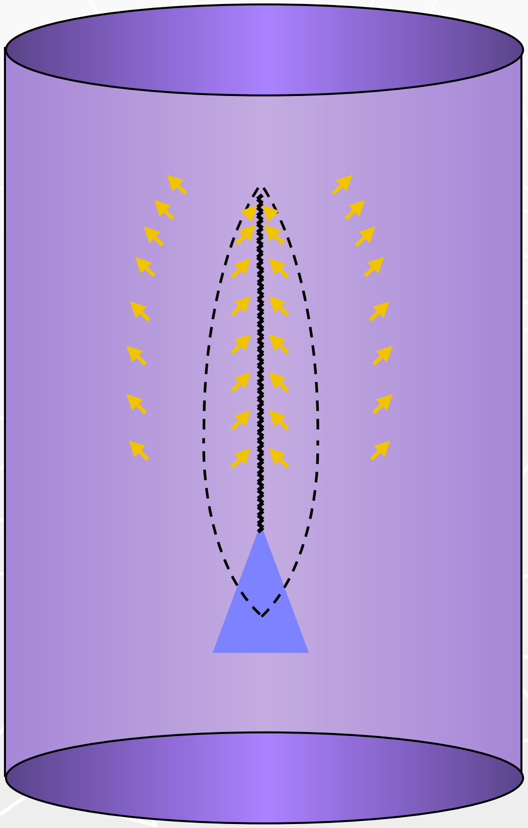
$$O(x) = \lim_{z\to 0} z^{\Delta}\phi(x,z)$$

How to extend into the bulk, and behind the horizon? One approach: integrate inward. This breaks down at the black hole horizon due to large blue shift, except for *very* young black hole.



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

PR,VV: we know that fields behind the horizon are entangled with fields *b* outside. Use this to identify them, since we already know the field operators outside!



Problem: entanglement depends on state, must know which states map to infalling vacuum =(?) *code subspace* of VV. But we still need a dynamical theory…

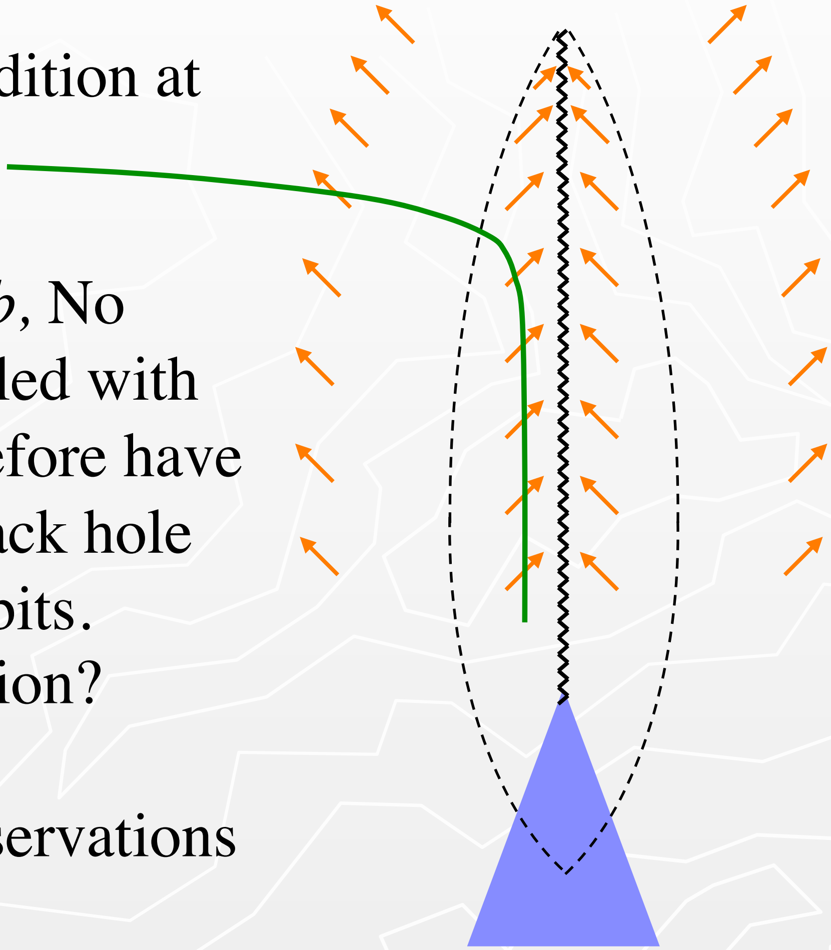
Before the Page time: *b* is entangled with black hole degrees of freedom, OK, but after the Page time it is entangled with the earlier Hawking radiation.

Black hole final state proposal (Horowitz & Maldacena hep-th/0310281, applied by Preskill & Kitaev, AMPS):

Impose final state boundary condition at black hole singularity.

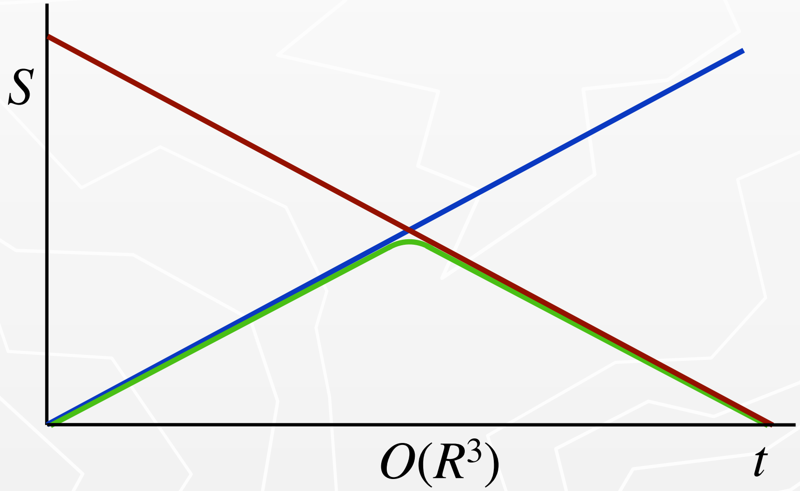
After an N-bit black hole emits b, No Drama requires that b be entangled with the black hole, which must therefore have N+1 bits. Purity requires the black hole Hilbert space to have only N−1 bits. Modify QM by adding a projection?

Is there a sensible theory for observations of infalling observer?



Nonlocal transport (Giddings 1211.7070):

Difficult to avoid processes leading to $S_{\mathrm{vN}} > S_{\mathrm{BH}}$ (AMPS, AMPSS).



Open questions

- If firewalls exist, when do they form?

Entanglement argument gives upper bound R^3, but most black hole properties are expected to come to equilibrium in a much shorter time, e.g. the light-crossing time R or the fast-scrambling time $R \ln R$. Susskind argues for R^3, interior is produced by self-entanglement of horizon.

We need a dynamical theory of the firewall.

If $R \ln R$, we must distinguish :

very young: $t < R \ln R$,
middle aged: $R \ln R < t <$ half-life
old: half-life $< t$

Open questions

- Are there any observational effects for black holes?

The argument is consistent with the exterior being exactly as in the usual picture, except perhaps for very subtle quantum effects. But who knows?

Heuristically, firewall = stretched horizon, except that you can’t fall through it.

Open questions

- Are there any consequences for cosmology?

Are cosmological horizons like black hole horizons? Is there a version of the information problem?

If we just carry over the black hole result, our current cosmological horizon is very young, but the horizon during inflation may have been middle-aged, depending on number of *e*-foldings vs. $\ln(R/l_P)$.

Most important, this may give us a new lever on applying holography to cosmology.

Conclusions

- Trivial resolution? Looking unlikely.

- I still trust AdS/CFT, so keep unitarity.

- If locality is emergent, no reason subtle nonlocalities shouldn't extend to finite distance. But it is hard to make a scenario.

- Attempts to construct the interior using AdS/CFT always assume that we understand bulk dynamics, maybe there is no interior…