Black Hole Information: Spacetime versus Quantum Mechanics

Joseph Polchinski

Strings 2014, Princeton
June 24, 2014

Black hole evaporation exposes an inconsistency between quantum mechanics and general relativity:

Hawking (1976): Information is lost. Quantum mechanics must be modified, replacing the S-matrix with a *$-matrix* that takes pure states to mixed states.

Black hole evaporation exposes an inconsistency between quantum mechanics and general relativity:

‘t Hooft, Susskind, BFSS, Maldacena, … (1993-97): Information is not lost, and QM is unmodified.

But spacetime is fundamentally nonlocal, *holographic*.

However, no single observer sees any nonlocality (black hole complementarity).

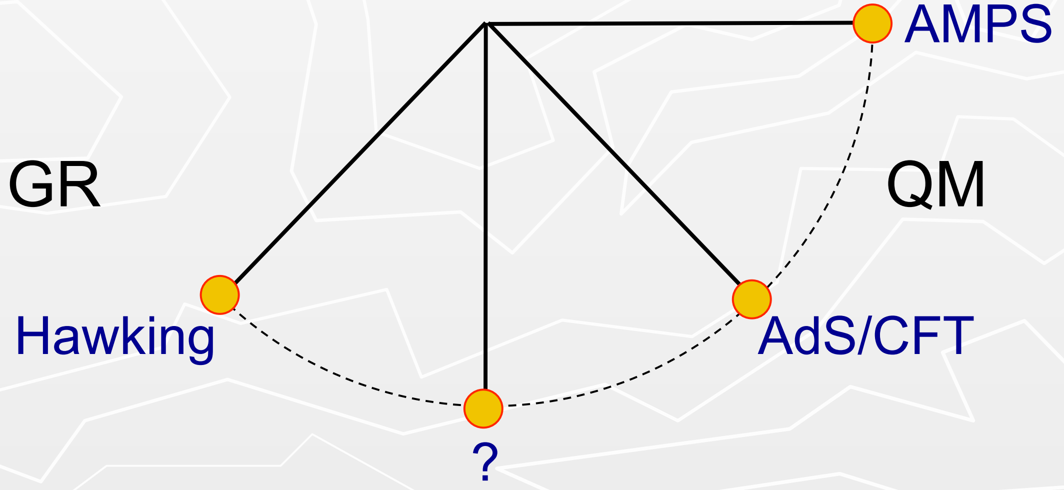
Black hole evaporation exposes an inconsistency between quantum mechanics and general relativity:

AMPS (2012): If QM is to be preserved, an infalling observer will see something radically different from what general relativity predicts, a *firewall* or perhaps just the end of space.

Black hole evaporation exposes an inconsistency between quantum mechanics and general relativity:

Most attempts to avoid the firewall modify QM, in new ways.

- Differ from Hawking: infalling vs. asymptotic observer.



The defenders of quantum mechanics:

Almheiri, Marolf, Polchinski, Sully 1207.3123

Almheiri, Marolf, Polchinski, Stanford, Sully 1304.6483

Marolf, Polchinski 1307.4706 and unpublished

Bousso 1207.5192, 1308.2665, 1308.3697

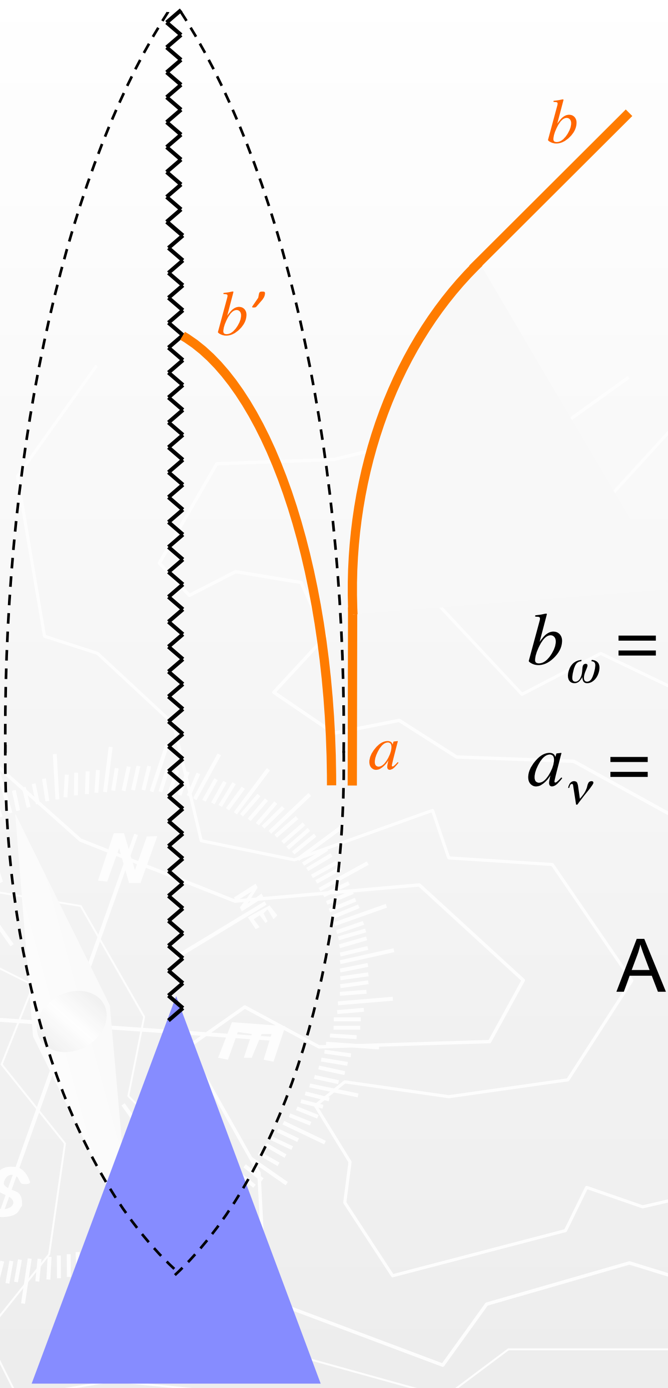
Harlow 1405.1995

Review:

Black hole evaporation

The Page curve and information loss

The AMPS argument



Hawking evaporation

b_ω: Outgoing Hawking modes

b'_ω: Interior Hawking modes

a_ν: Modes of infalling observer

$$b_\omega = A_{\omega\nu} a_\nu + B_{\omega\nu} a_\nu^\dagger$$

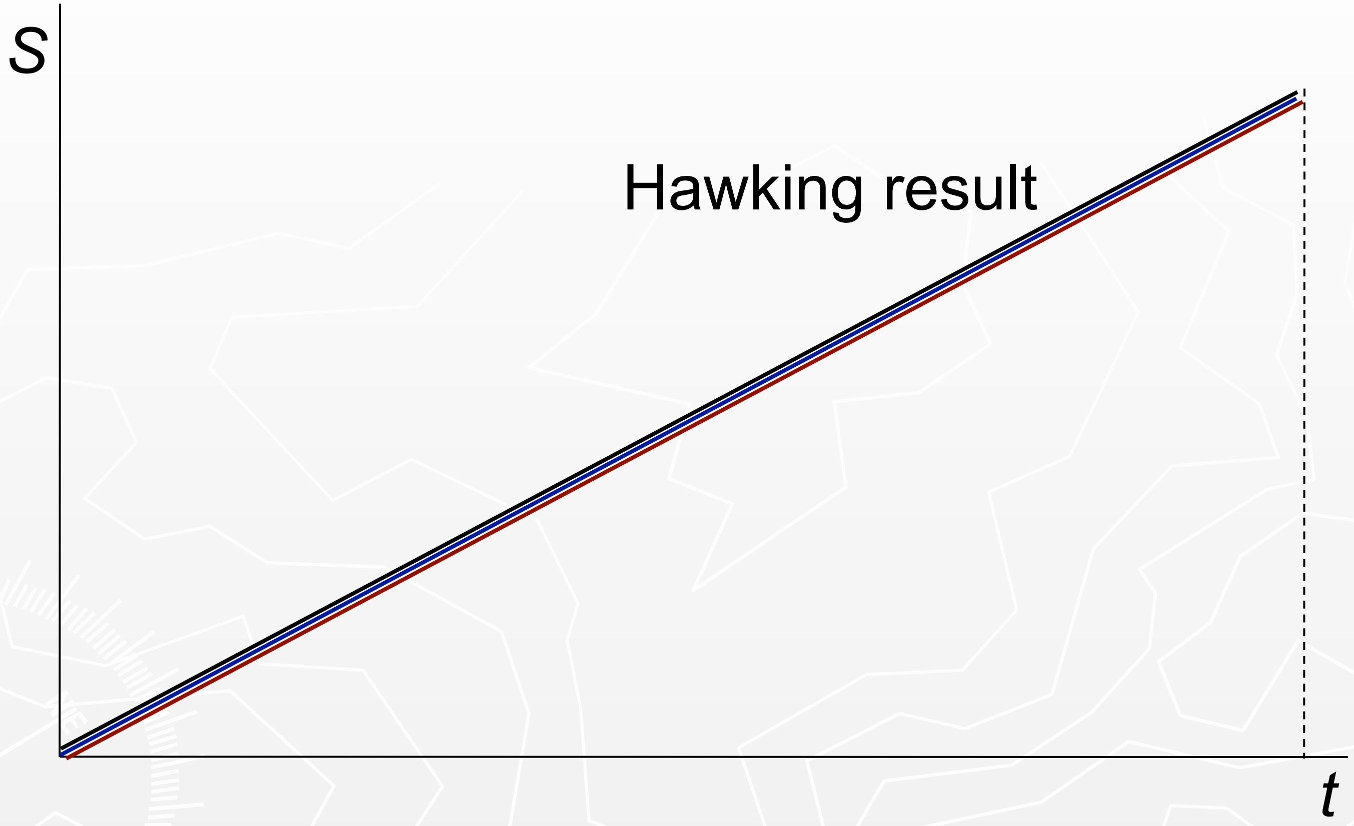
$$a_\nu = C_{\nu\omega} b_\omega + D_{\nu\omega} b_\omega^\dagger + E_{\nu\omega} b'_\omega + F_{\nu\omega} b'^\dagger_\omega$$

Adiabatic principle/no drama:

$a|\psi\rangle = 0$ so $b|\psi\rangle \neq 0$

→ Hawking radiation

The Page curve for an evaporating black hole:

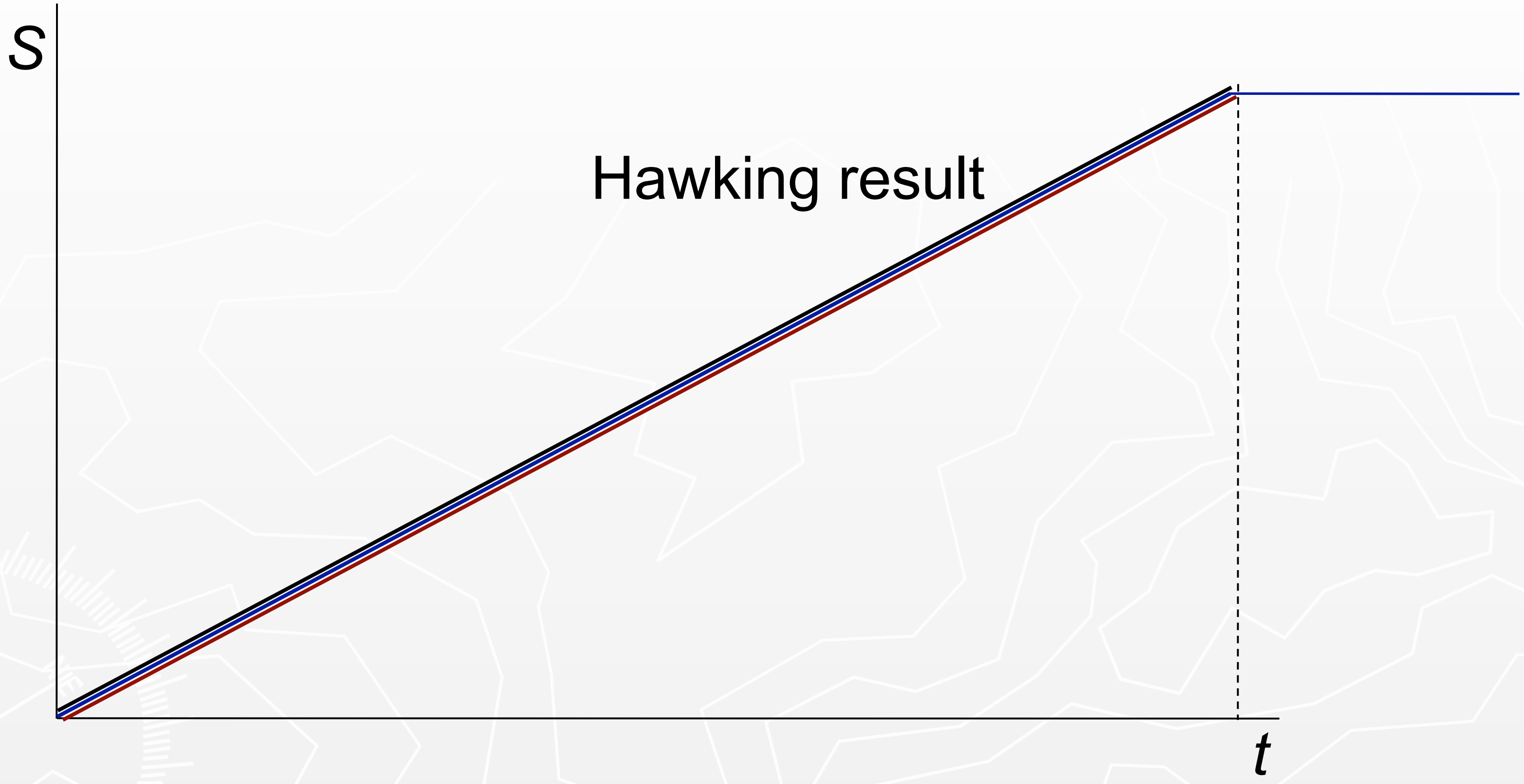


S = von Neumann entropy of the Hawking radiation

= entanglement entropy of radiation and black hole

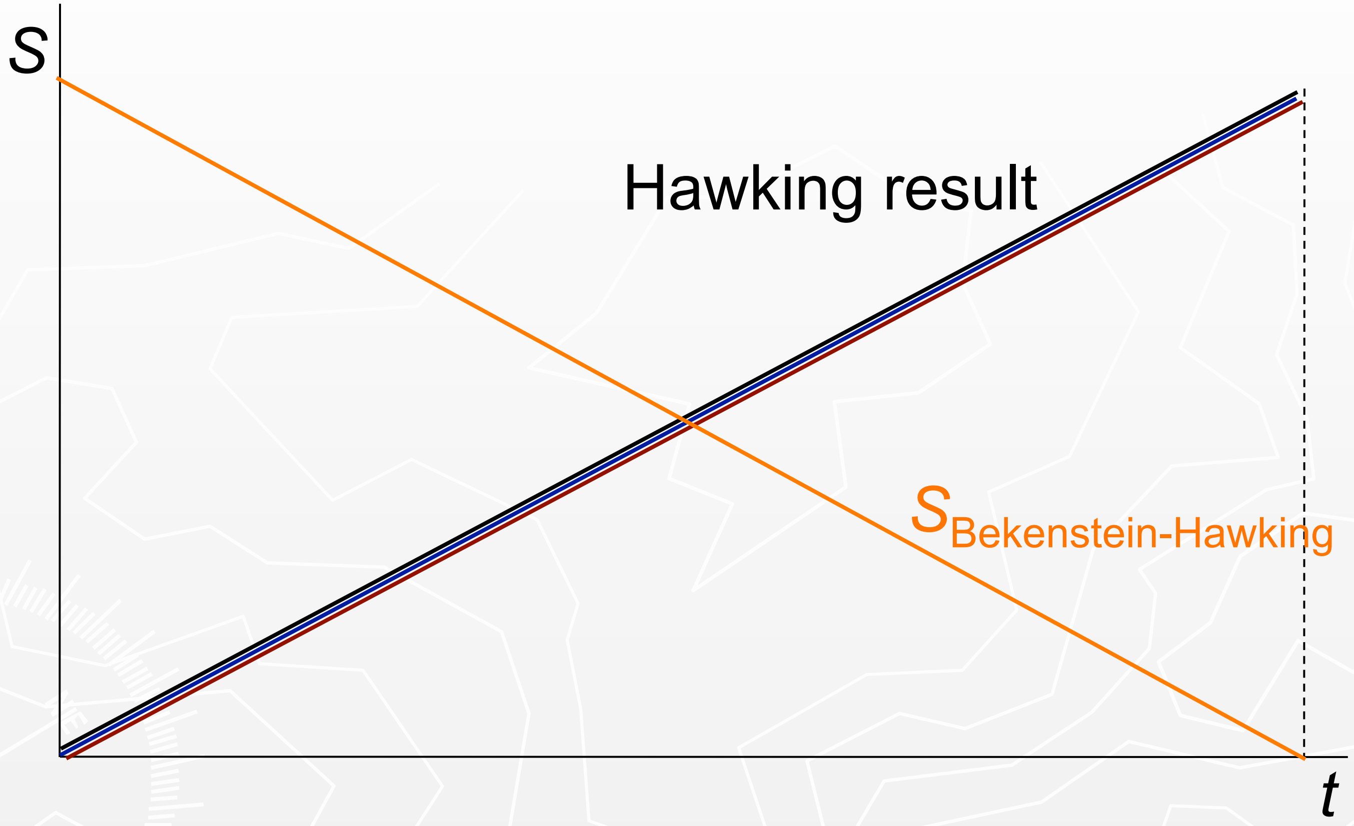
= von Neumann entropy of the black hole

The Page curve for an evaporating black hole:



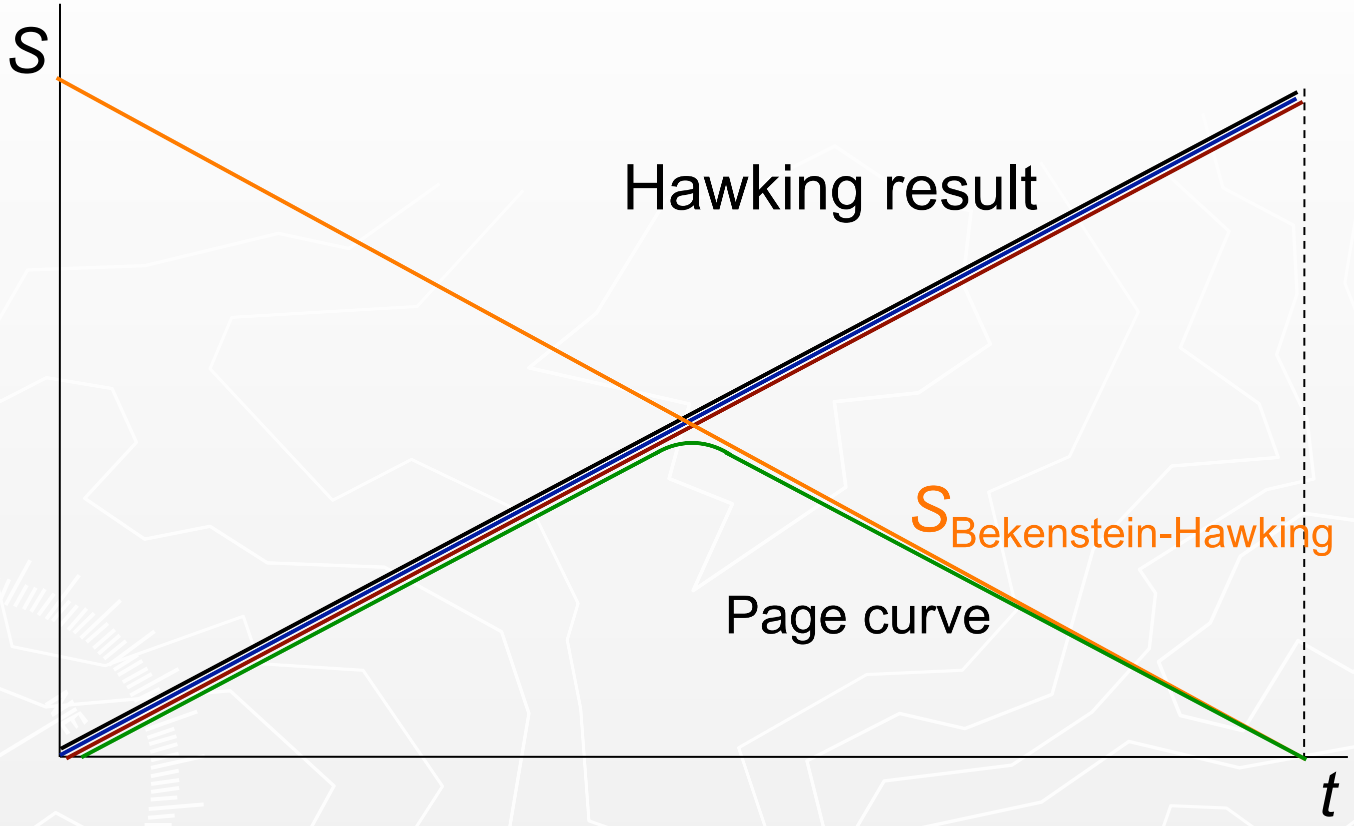
When the black hole has evaporated, all that is left is the Hawking radiation, in a mixed state.

The Page curve for an evaporating black hole:

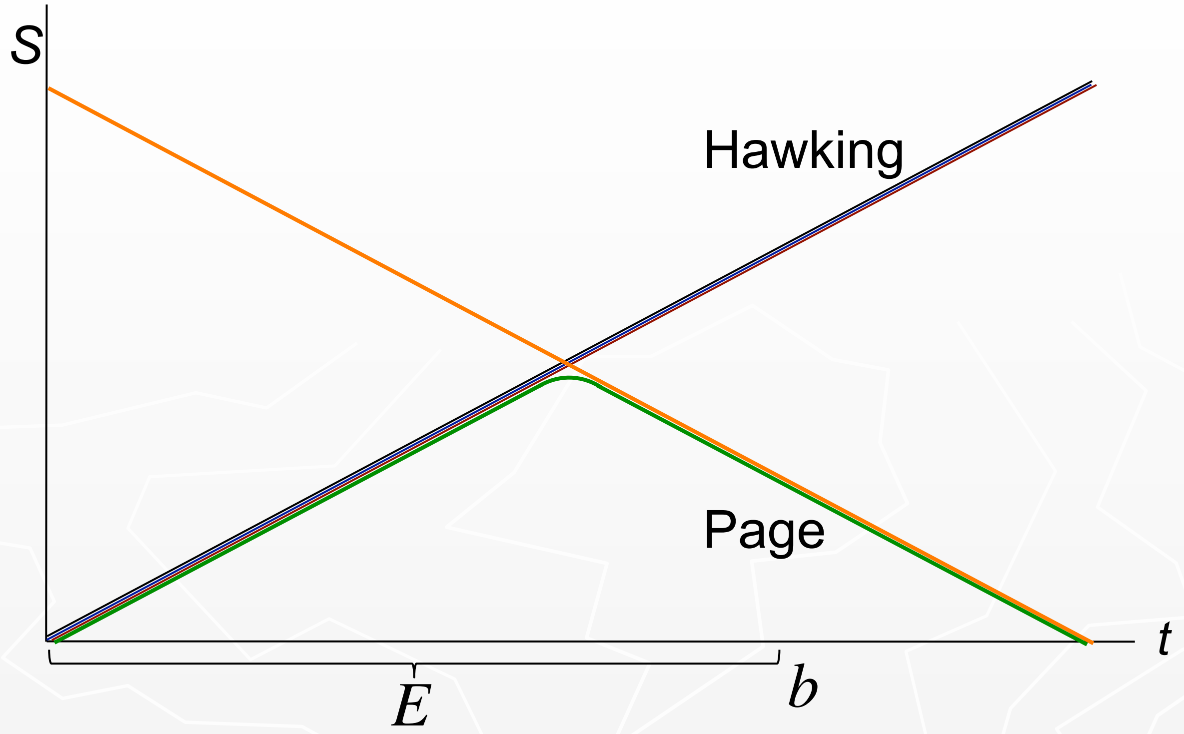


Around the midpoint, the fine-grained entropy of the black hole exceeds its course-grained entropy.

The Page curve for an evaporating black hole:



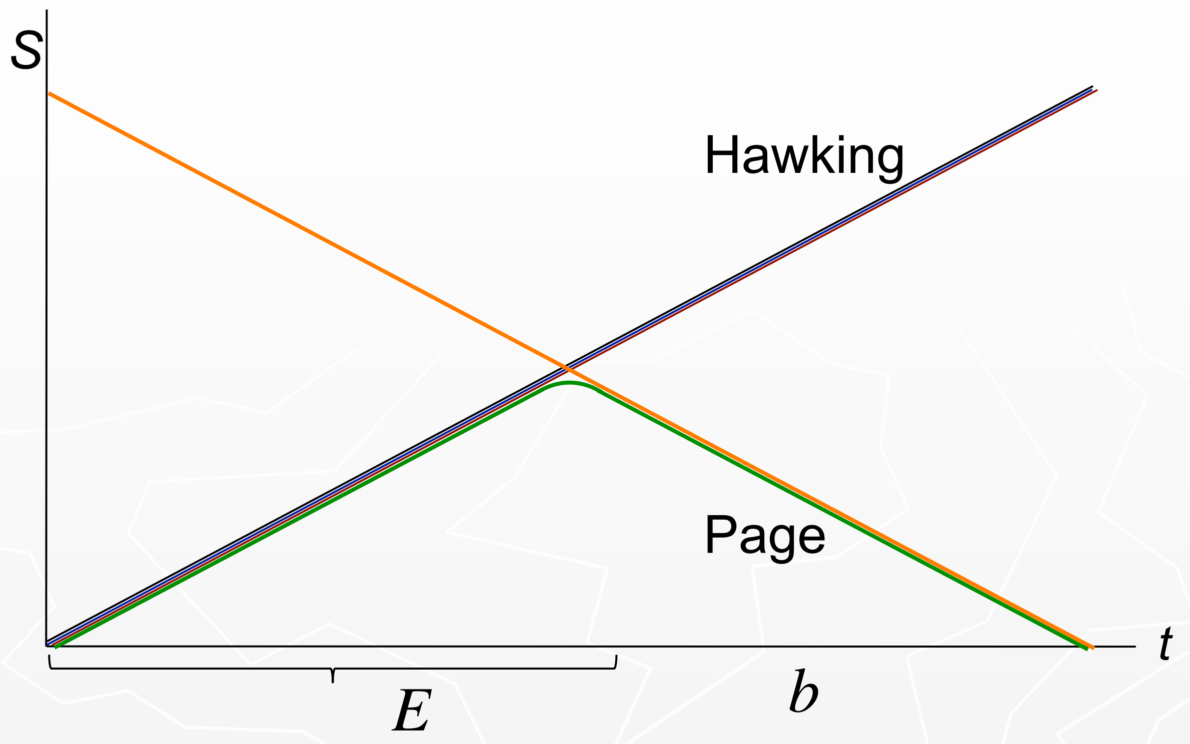
In order for the Hawking radiation to be pure, we must deviate from the Hawking calculation already around the midpoint: an $O(1)$ effect.



Strong subadditivity (Mathur 0909.1038):

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

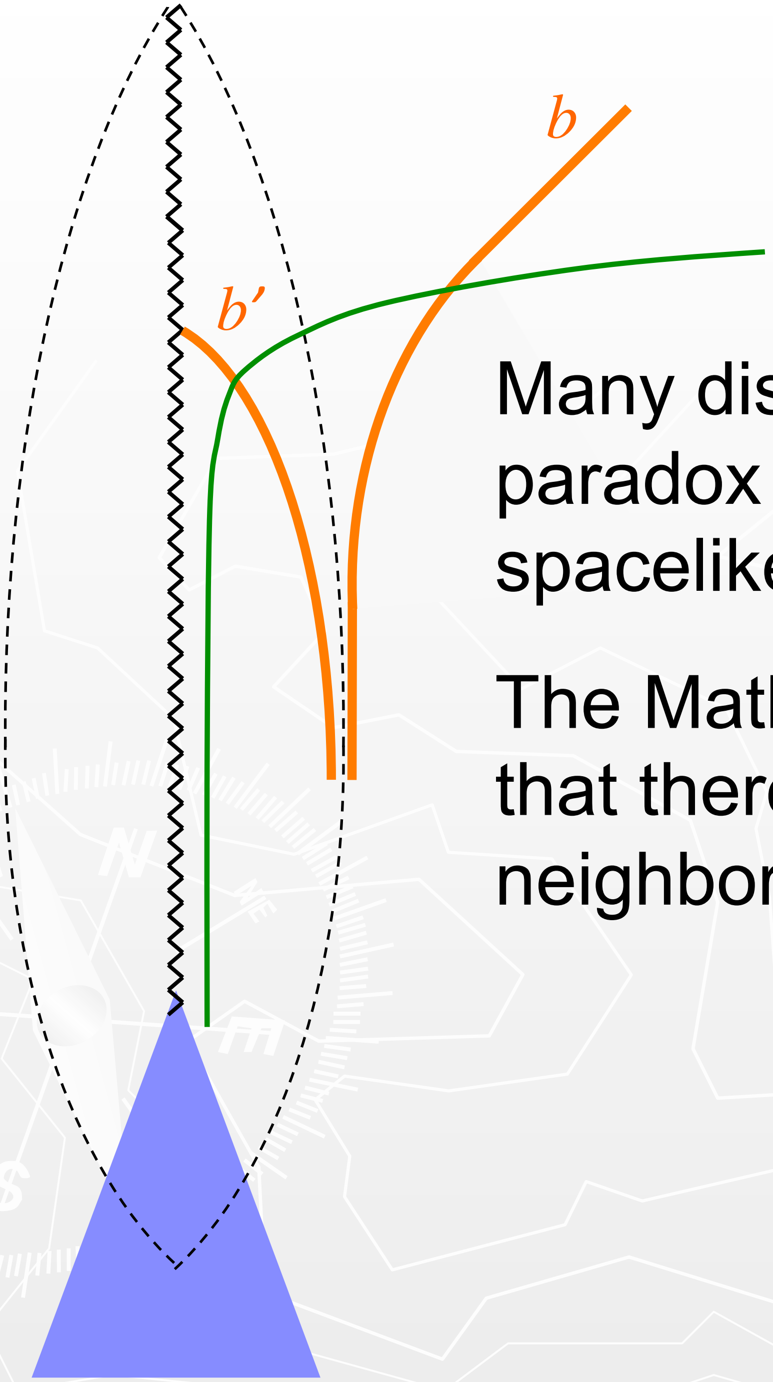
Here $S_{b'b} = 0 \rightarrow \quad S_{b'bE} = S_E \quad \rightarrow \quad S_{bE} \geq S_b + S_E \rightarrow$ Hawking, not Page.



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

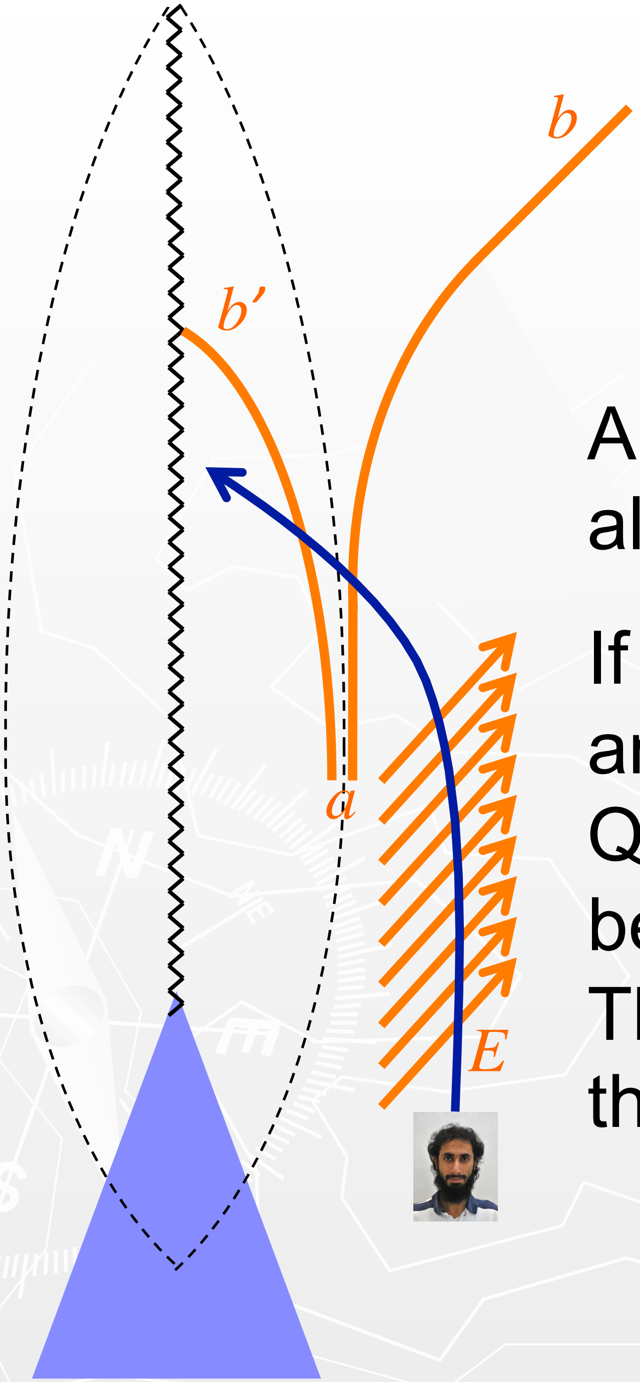
Here $S_{b'b} = 0 \rightarrow S_{b'bE} = S_E \rightarrow S_{bE} \geq S_b + S_E \rightarrow$ Hawking, not Page.

Aside: If chaotic, *E* need only be ½ + δ of the early photons (Hayden & Preskill 0708.4025)



Many discussions of the information paradox focus on the state on a long spacelike slice.

The Mathur argument emphasizes that there is already a problem in the neighborhood of the horizon.



AMPS: A single observer can see all of *E*, *b*, and *b*'

If the Hawking radiation is pure, and this observer obeys ordinary QM, then the entanglement between *b* and *b*' must be lost. Then the *a* modes are excited: the firewall.

Another argument: Put the black hole in a large box (AdS), so that it is stable. Typical high energy states look like black holes.

Consider a basis in which $N_i = b_i^\dagger b_i$ (and their CFT images) are diagonal, for some set of modes i.

N_i is thermal in the a-vacuum. The N_i eigenstates are therefore far from the a-vacuum: each a_i is excited with probability $O(½)$. So all these basis states have firewalls.

If there is a projection operator P onto states with firewalls, then $P \approx 1$ in this basis, and therefore in every basis.

In quantum mechanics such projection operators normally exist, e.g. for excitations above empty AdS, or outside a black hole.

Evidently if we are to avoid the firewall, we need different rules inside (or something else like nonlocal physics outside the black hole).

Is this a bug or a feature?

Ideas that modify quantum mechanics:

- State dependence (Papadodimas & Raju, Verlinde^2)
- EPR = ER (Maldacena & Susskind)
- Final state boundary condition at the black hole singularity (Horowitz & Maldacena; Preskill & Lloyd)
- Limits on quantum computation plus strong complementarity (Harlow & Hayden).

State dependence (P&R 1211.6767, 1310.6334, 1310.6335, V&V 1211.6913, 1306.0515, 1311.1137, ~Nomura, Varela, Weinberg 1207.6626, …, 1406.1505).

Consider a *typical* black hole state $|\psi_{\mathrm{typ}}>$. The distribution of the modes b_i is thermal:

$$|\psi_{\mathrm{typ}}> = Z^{-1/2}\left(|0>_B |\psi_{\mathrm{typ}},0>_{B*} + e^{-\beta\omega/2}\,|1>_B |\psi_{\mathrm{typ}},1>_{B*}\right)$$

where $B*$ is the complement to B. Compare

$$|0>_A = Z^{-1/2}\left(|0>_B |0>_{B'} + e^{-\beta\omega/2}\,|1>_B |1>_{B'}\right)$$

Thus identify the internal Hilbert space,

$$|n>_{B'} = |\psi_{\mathrm{typ}},n>_{B*}$$

With this interpretation, typical states are a-vacua: no firewall.

Key issue: given a black hole in some state $|\psi\rangle$, what reference state $|\psi_{\mathrm{typ}}\rangle$ do we use? A particular challenge is

$$|\psi\rangle = Z^{-1/2}(|0\rangle_B \, |\psi_{\mathrm{typ}},0\rangle_{B*} - e^{-\beta\omega/2} \, |1\rangle_B \, |\psi_{\mathrm{typ}},1\rangle_{B*}).$$

Is this an excitation of

$$|\psi\rangle = Z^{-1/2} (|0\rangle_B \, |\psi_{\mathrm{typ}},0\rangle_{B*} + e^{-\omega/2T} \, |1\rangle_B \, |\psi_{\mathrm{typ}},1\rangle_{B*}),$$

or is it a typical state in its own right, and therefore unexcited? (PR prescription later)

Given a reference state, P&R build interior operators

$$\tilde{A}_p = g^{mn} A_m e^{-\beta H/2} A_p^\dagger e^{-\beta H/2} |\psi_t\rangle\langle\psi_t| A_n^\dagger$$

from which they can construct projection operators $P(n_A)$ onto states of given excitation level for the infalling observer.

The issue is that when one specifies the reference state $\psi_t(\psi)$, these become nonlinear operators $P(n_A,\psi)$.

This state-dependence is a modification of the Born rule, and is different from normal notions of background-dependence.

Ordinary QM: The system is in a state $|\psi\rangle$. The probability of finding it in a given basis state $|i\rangle$ is

$$|\langle i|\psi\rangle|^2 = \langle\psi|P_i|\psi\rangle \,.$$

The probability of finding a given excitation is

$$\sum_{i\in S} |\langle i|\psi\rangle|^2 = \langle\psi|P_S|\psi\rangle \,,$$

where S is the set of all states with the given excitation and background.

`Background-dependence', the black hole or whatever is being excited, is all built into i and S.

P_S is a linear operator, which does not depend on $|\psi\rangle$. This is the Born rule, and one must modify it to $P_S(\psi)$ to avoid the firewall by this route.

More detailed issues:

The state

$$|\psi\rangle = Z^{-1/2}(|0\rangle_B\, |\psi_{\mathrm{typ}},0\rangle_{B*} - e^{-\beta\omega/2}\, |1\rangle_B\, |\psi_{\mathrm{typ}},1\rangle_{B*})$$

is not quite typical ($O(1/N^{\alpha})$). For any $|\psi\rangle$, can find U such that $U|\psi\rangle$ is an `equibrium state' (PR).

Then there are pairs of states $|\mathrm{vac}\rangle$ and $|\mathrm{exc}\rangle$ such that one is vacuum and one is excited, but

$$\langle \mathrm{vac}|\mathrm{exc}\rangle = 1 - O(1/N^{\alpha}).$$

(Possibly even 1 exactly: Harlow).

How to interpret

$$\alpha|vac\rangle + \beta|exc\rangle?$$

How to interpret

$$\{|vac\rangle\,|{+z}\rangle + |exc\rangle|{-z}\rangle\}/\sqrt{2},$$

where we have coupled this to a detector spin?

Problem: the interpretation is different if one writes this as

$$\{(|vac\rangle + |exc\rangle)|{+x}\rangle + (|vac\rangle - |exc\rangle)|{-x}\rangle\}/2!$$

Any framework that modifies QM has to be able to answer such questions. (Code subspaces [VV 1311.1137] don't help: $|vac\rangle$ and $|exc\rangle$ can't be in the same one.)

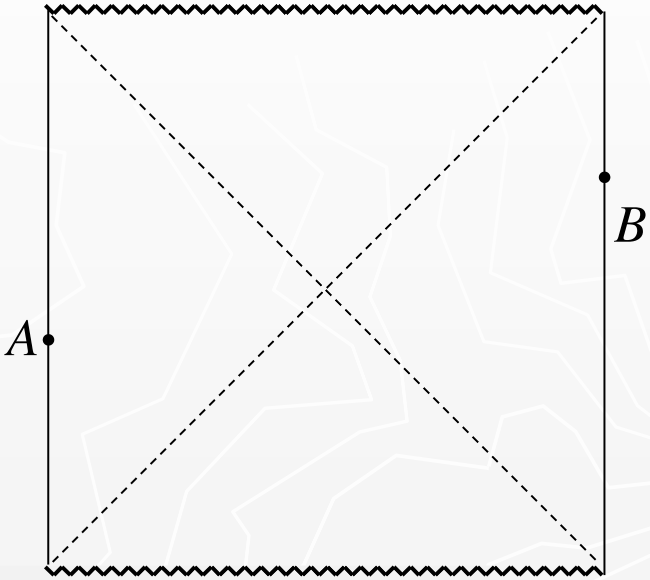
Nomura & Weinberg 1406.1505 similar but claim state-independence. Nonunitary evolution (v1).

Another issue (Bousso, Harlow): the equilibrium state prescription is designed for AdS black holes. It says nothing about evaporating black holes, where the Hawking radiation is far from equilibrium.

Possible alternative: that $|\psi_{\mathrm{typ}}>$ is determined by a dynamical evolution equation. Intuition: a black hole that has not been disturbed for a while should have a smooth horizon. Still modifies QM.

EPR = ER (Maldacena & Susskind 1306.0533):

Israel '76, Maldacena hep-th/0106112: two-sided AdS geometry in HH state calculates two-CFT correlators in thermofield state



$$\langle\psi|A_L(-t)B_R(t')|\psi\rangle$$

$$= \sum_{\alpha,\beta,\gamma,\delta} e^{-itE_{\delta\gamma}-it'E_{\alpha\beta}}\psi^*_{\delta\beta}\psi_{\gamma\alpha}A_{\gamma\delta}B_{\beta\alpha}$$

$$\psi_{\alpha\gamma} = Z^{-1/2}\delta_{\alpha\gamma}e^{-\beta E_\alpha/2}$$ (Energy eigenbasis)

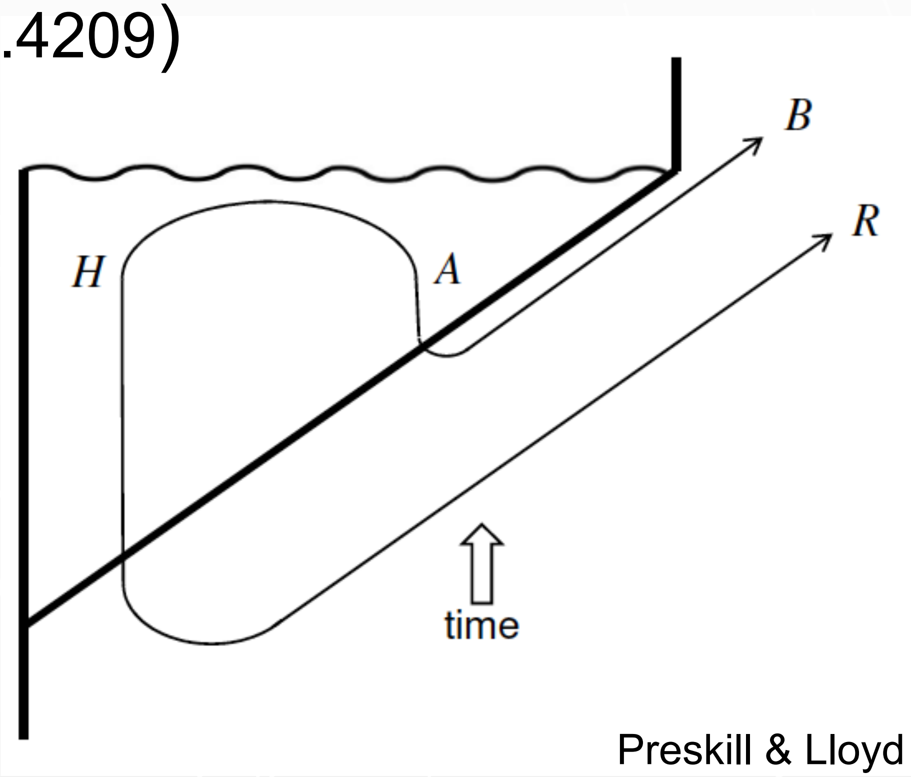
So ER → EPR.

Is the reverse true, are entangled systems in the TF state always connected by bridges, EPR → ER? Or does the interior depend on extra d.o.f. (Marolf & Wall, 1210.3590)?

Interpretation for more general states: what does an observer who jumps into one side see? If typical states are not to have firewalls, this reduces to PR. Additional problem: time-folds.

Susskind (1311.3335, 1311.7379, 1402.5674, +Stanford 1406.2678): Haar-typical states may have firewalls, but states of low complexity do not. Still nonlinear QM.

Final state boundary condition at the black hole singularity (Horowitz & Maldacena hep-th/0310281; Lloyd & Preskill 1308.4209)



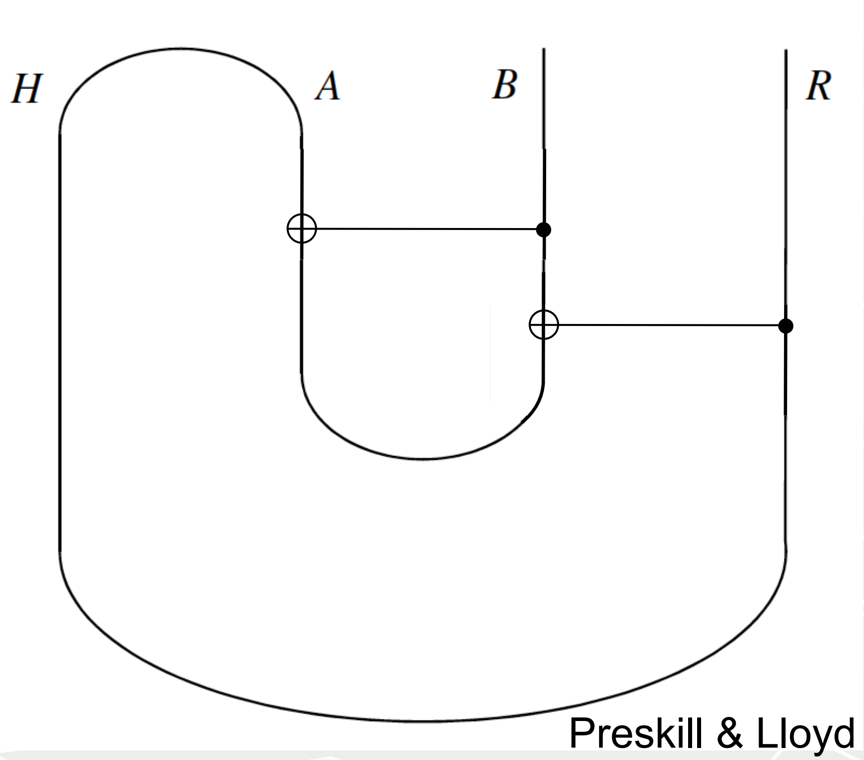
Preskill & Lloyd

Projecting on a final state at the singularity gives necessary entanglements.

Issues with final state:

- No probability interpretation in interior (Bousso & Stanford 1310.7457)
- Acausal behavior visible even *outside* the horizon (Lloyd & Preskill, 1308.4209v2, to appear).

Result of first measurement (outside the horizon) depends on whether later measurement is done.



Preskill & Lloyd

Limits on quantum computation (Harlow & Hayden 1301.4504): Perhaps there is not time to verify the *b*-*E* entanglement, in the first version of the paradox.

- Doesn’t apply to AdS black holes (AMPSS 1304.6483).
- Can be evaded by pre-computing (Oppenheim & Unruh 1401.1523).
- What would it mean – an uncertainty principle for the wavefunction?

Goal 1: a consistent scenario, either with or without firewalls.

Goal 2: a full theory of quantum gravity that gives rise to this scenario.

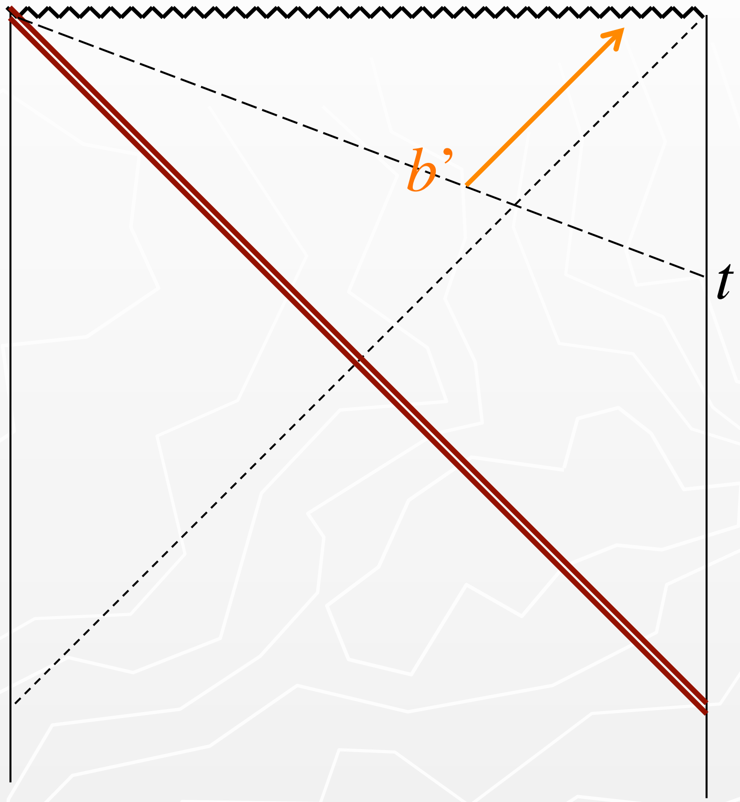
Bottom up versus top down.

Another lesson: the impotence of AdS/CFT.

Sharp (e.g. GKPW) dictionary only for asymptotics (including $t = \pm\infty$).

Must integrate the bulk to the boundary, e.g. with precursors.

But for inner Hawking modes, we hit either the singularity…

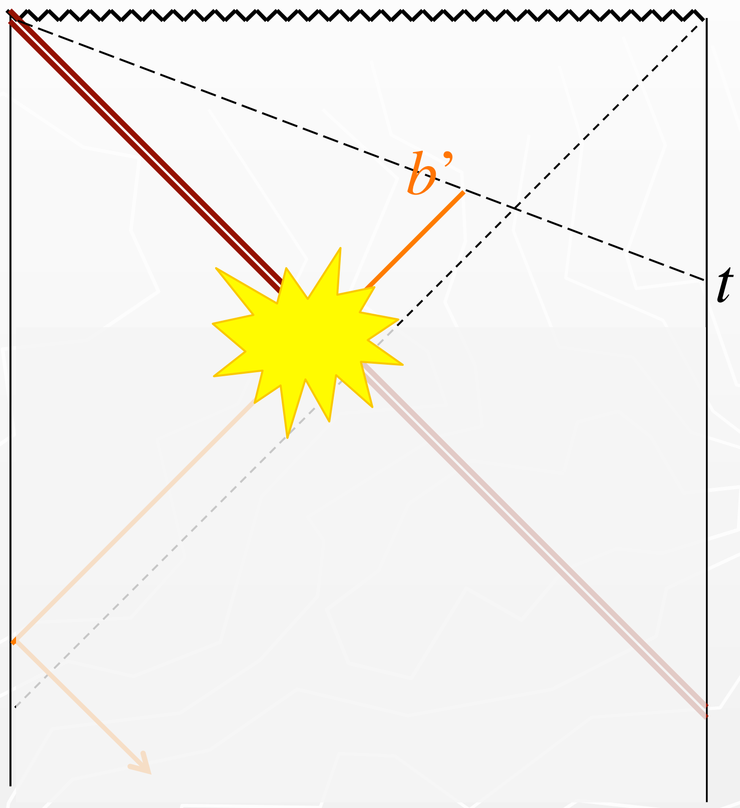


Another lesson: the impotence of AdS/CFT.

Sharp (e.g. GKPW) dictionary only for asymptotics (including $t = \pm\infty$).

Must integrate the bulk to the boundary, e.g. with precursors.

But for inner Hawking modes, we hit either the singularity or the collapsing star (trans-Planckian).



If we could construct b' then we could construct P, and there would be firewalls ($P \approx 1$, slide 17)

So, what to give up?

Purity of the Hawking radiation?

Absence of drama for the infalling observer?

EFT/locality outside the horizon?

Quantum mechanics for the infalling observer?

So, what to give up?

Purity of the Hawking radiation?

Absence of drama for the infalling observer?

EFT/locality outside the horizon?

Quantum mechanics for the infalling observer?

So, what to give up?

Purity of the Hawking radiation?

Absence of drama for the infalling observer?

EFT/locality outside the horizon?

Quantum mechanics for the infalling observer?

So, what to give up?

Purity of the Hawking radiation?

Absence of drama for the infalling observer?

EFT/locality outside the horizon?

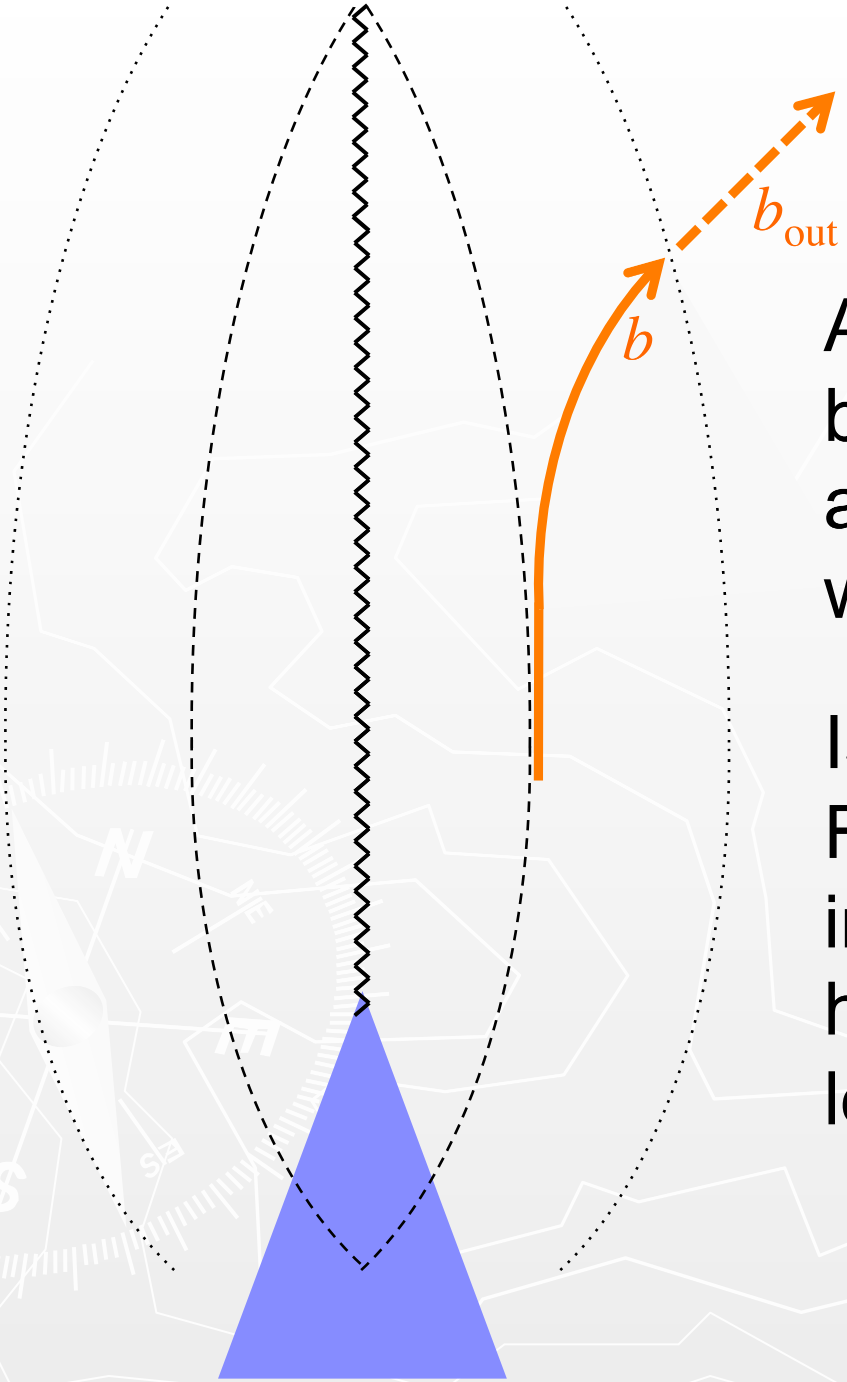
Quantum mechanics for the infalling observer?

EFT/locality outside the horizon?

Why shouldn’t nonlocality extend outside the horizon? (But it’s not a small effect).

E.g. `nonviolent nonlocality’ (Giddings 1108.2015, … ,1401.5804).

Example:



At $r = 2r_s$, original b teleports back into the black hole, and a new b_{out}, entangled with E, appears.

Issue: experiments at $r < 2r_s$. For example, one can pump information into the black hole without adding energy, leading to info loss.

So, what to give up?

Purity of the Hawking radiation?

Absence of drama for the infalling observer?

EFT/locality outside the horizon?

Quantum mechanics for the infalling observer?

How can firewalls form in a place that is not locally special?

The horizon is future-special, but it is also past-special (trans-Planckian effects). Maybe strings are sensitive to this (Silverstein 1402.1486):

Evidence for nonadiabaticity! How to understand from `nice-slice' point of view?

A comment on fuzzballs:

Fang Chen, Ben Michel, JP, Andrea Puhm, in prep.

Naïve geometry of 2-charge fuzzball:

$$ds^2_{\mathrm{IIB}} = \frac{1}{\sqrt{H_1 H_5}}(-dt^2 + R^2 dy^2) + \sqrt{H_1 H_5}\, dx^2 + \sqrt{\frac{H_1}{H_5}}\sqrt{V} dz_4^2$$

$$H_1 = 1 + \frac{Q_1}{r^2} \quad H_5 = 1 + \frac{Q_5}{r^2}$$

For y noncompact, this goes to AdS_3 x S^3 x T^4.
For y periodic, $r = 0$ becomes a cusp singularity.

According to the fuzzball program (e.g. Mathur review hep-th/0502050), this is not an acceptable string geometry, and must be replaced by fuzzball geometries.

$$
\begin{aligned}
ds^2_{\mathrm{IIB}} &= \frac{r^2}{Q}(-dt^2 + R^2 dy^2) + \frac{Q}{r^2}(dr^2 + r^2 d\Omega_3^2) + \sqrt{\frac{N_1}{N_5}}dz_4^2 \\
e^{2\phi_{\mathrm{IIB}}} &= g^2 \,.
\end{aligned}
$$

As $r \to 0$, y circle gets small: T-dual to IIA.

Then e^{ϕ} gets big: lift to M theory!

Then T^4 gets small: STS-dual to IIB!

Then curvature gets big and coupling gets small: go to free CFT dual.

(Martinec & Sasakian, hep-th/9901135.)

Towards decreasing r, lower energy:

IIB D1-D5

IIA D0-D4

M p-M5

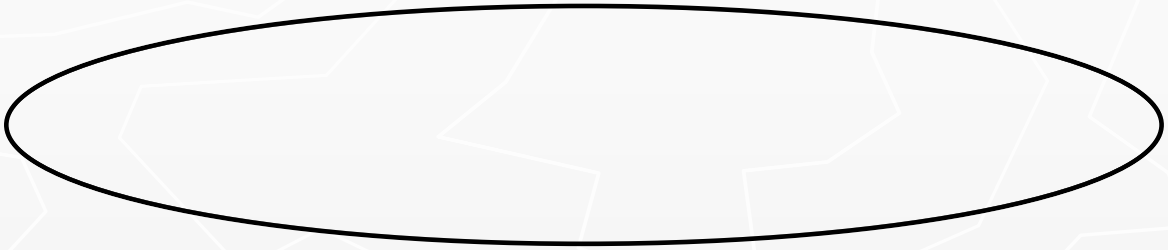
IIB' p-F1

long string CFT (Motl hep-th/9701025; Banks, Seiberg 9702187; Dijkgraaf, Verlinde, Verlinde 9703030)

Fuzzball geometries go over to naïve geometry at large r, typical size ~ crossover to free CFT.

$r_{\text{breakdown}} = r_{\text{fuzz}} = r_{\text{entropy}}$ (radius where area in Planck units equals microscopic entropy $N_1 N_5$).

Now look at states of nonzero J. Naïve geometry has a ring singularity (Elvang, Emparan, Mateos, Reall, hep-th/0407065, Balasubramanian, Kraus, Shigemori, hep-th/0508110).



Fuzzballs:

Now $\rho_{\text{fuzz}} = \rho_{\text{entropy}}$, but $\rho_{\text{breakdown}}$ can be larger or smaller. Lesson?