PHYSICAL REVIEW D VOLUME 14, NUMBER 10 15 NOVEMBER 1976

Breakdown of predictability in gravitational collapse*

S. W. Hawking†

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England and California Institute of Technology, Pasadena, California 91125

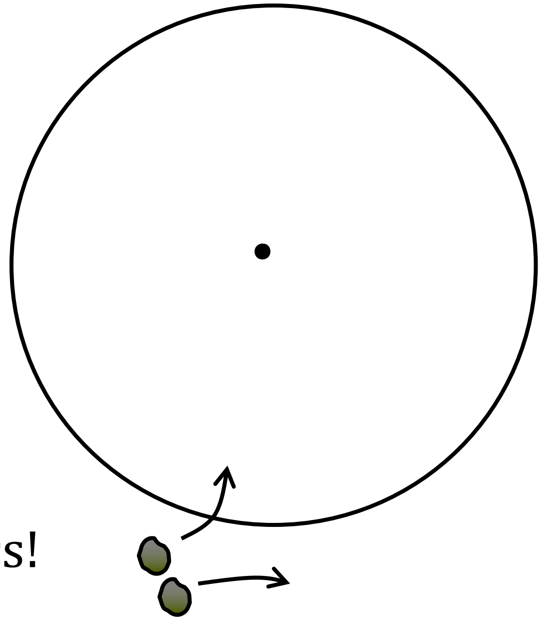
(Received 25 August 1975)

The principle of equivalence, which says that gravity couples to the energy-momentum tensor of matter, and the quantum-mechanical requirement that energy should be positive imply that gravity is always attractive. This leads to singularities in any reasonable theory of gravitation. A singularity is a place where the classical concepts of space and time break down as do all the known laws of physics because they are all formulated on a classical space-time background. In this paper it is claimed that this breakdown is not merely a result of our ignorance of the correct theory but that it represents a fundamental limitation to our ability to predict the future, a limitation that is analogous but additional to the limitation imposed by the normal quantum-mechanical uncertainty principle. The new limitation arises because general relativity allows the causal structure of space-time to be very different from that of Minkowski space. The interaction region can be bounded not only by an initial surface on which data are given and a final surface on which measurements are made but also a "hidden surface" about which the observer has only limited information such as the mass, angular momentum, and charge. Concerning this hidden surface one has a "principle of ignorance": The surface emits with equal probability all configurations of particles compatible with the observers limited knowledge. It is shown that the ignorance principle holds for the quantum-mechanical evaporation of black holes: The black hole creates particles in pairs, with one particle always falling into the hole and the other possibly escaping to infinity. Because part of the information about the state of the system is lost down the hole, the final situation is represented by a density matrix rather than a pure quantum state. This means there is no S matrix for the process of black-hole formation and evaporation. Instead one has to introduce a new operator, called the superscattering operator, which maps density matrices describing the initial situation to density matrices describing the final situation.

16. The Black Hole Information Loss Paradox

1. Set-Up

- Hawking's (1975) result: Black holes emit radiation.
  - *This is a process of evaporation!*
    - *Emission involves the absorption of a negative mass antiparticle.*
    - *A decrease in mass entails a decrease in surface area of event horizon, $A \sim M$.*

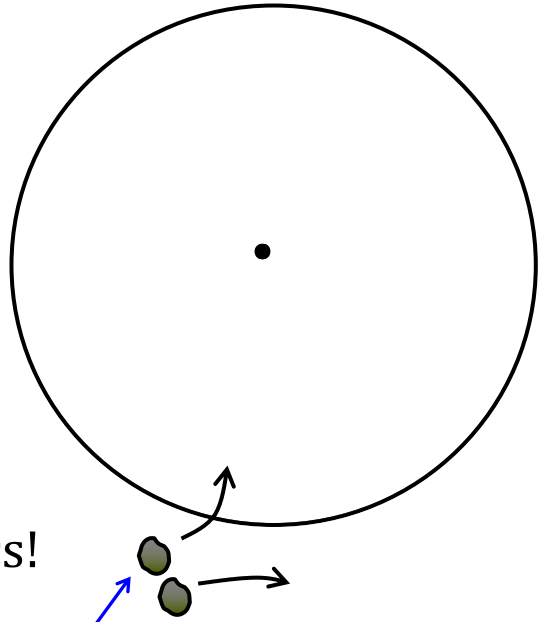


- *Evaporating black hole*: A test-bed for theoretical physics!
  - *statistical mechanics*
  - *thermodynamics*
  - *quantum field theory*
  - *general relativity*

- *Are these fundamental theories in physics consistent with each other?*
- *Can a consistent account of an evaporating black hole be given?*

1. Set-Up

- Hawking's (1975) result: Black holes emit radiation.
  - *This is a process of evaporation!*
    - *Emission involves the absorption of a negative mass antiparticle.*
    - *A decrease in mass entails a decrease in surface area of event horizon, $A \sim M$.*



- *Evaporating black hole*: A test-bed for theoretical physics!
- *Question*: What happens to the information encoded in a black hole as it evaporates?
  - *Does it escape into the environment?*
  - *Is it lost forever?*

Hawking's (1976) intuition: "...part of the information about the state of the system is lost down the hole..."

Information Loss Paradox, Naive Version

A burning chunk of coal emits (heat) radiation and "evaporates".

The information encoded in the chunk escapes in the heat radiation and diffuses into the environment.

- *But*: Unlike a chunk of coal, a black hole has an event horizon: information encoded in its interior cannot escape!

What happens to this information once the black hole has completely evaporated?

Why this is naive:

(a) What is the sense of "information" being appealed to?

(b) Why should it be concerning that "information" gets "lost"?

Let's try to be a bit more precise...

(a) *What is the sense of "information"?*

- *Suppose*: "info encoded in a state" means "the degree to which the state is mixed".
- *Then*: "info gets lost" means "a state evolves from a less mixed state to a more mixed state".

(b) *Why is "information loss" concerning?*

- *Because*: In quantum mechanics, the Schrödinger dynamics is *unitary*.
- *Which means*: It cannot transform a less mixed state to a more mixed state.

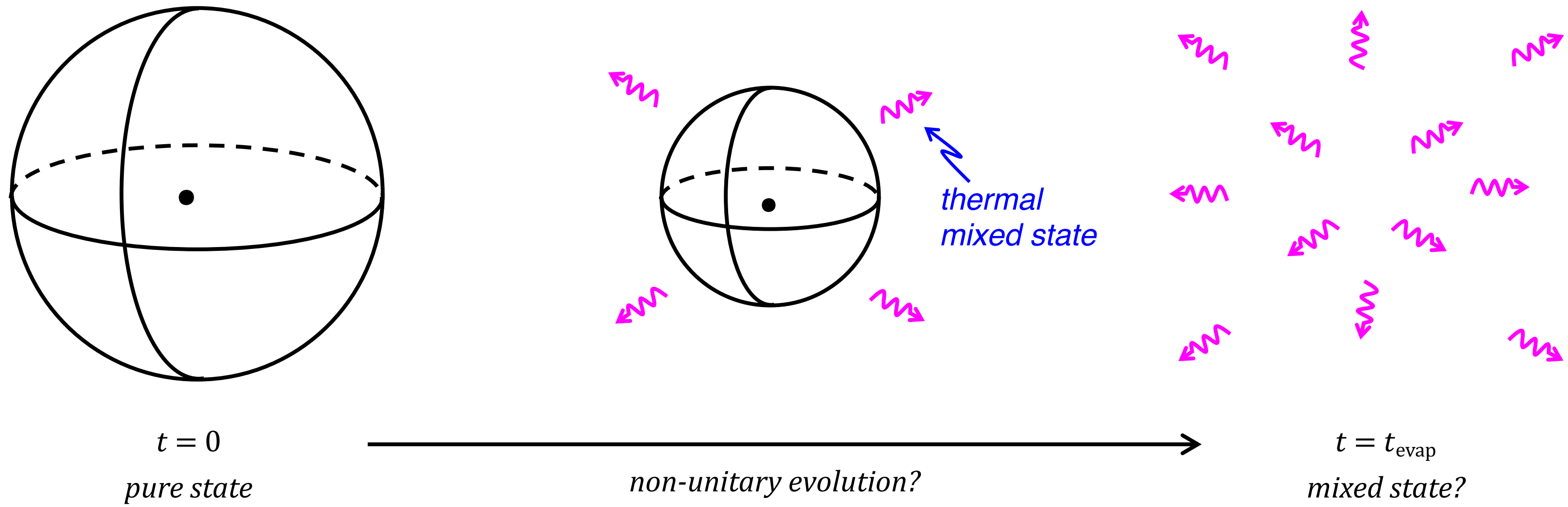
Lingering Concerns:

- Does the process of black hole evaporation involve a transition from a quantum mechanical less mixed state to a quantum mechanical more mixed state?
- And if so, doesn't the Projection Postulate allow this? Isn't information lost during measurements (Landauer's Principle)? Why should this be concerning?

Task: Formulate the "paradox" in a way that doesn't refer to the concept of "information"...

2. The Evaporation Time Paradox

Claim: Complete black hole evaporation is a non-unitary process.



- According to Hawking's (1975) analysis, radiation modes are in thermal *mixed* states at all times.
  - So at $t = t_{\text{evap}}$, the composite system (cloud of radiation) consists of subsystems (radiation modes) that are all in mixed states.
  - *But*: The state of a composite system can be pure when all of its subsystems are in mixed states...

How can we argue that, in this case, the final state must be mixed?

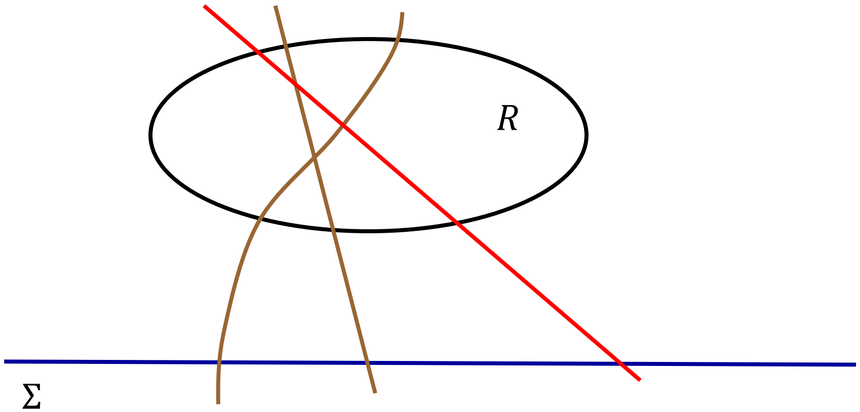
Argument

(1) For the state of a physical system in a background spacetime to change via a unitary transformation, the spacetime must be *globally hyperbolic*.

Def. A *globally hyperbolic spacetime* is a spacetime that admits a *Cauchy surface*.

A *spacelike* surface Σ such that every *non-spacelike* worldline without endpoints intersects Σ exactly once.

- *Why is this important?* Cauchy surfaces serve as initial data surfaces and thus provide a basis for *determinism* in relativistic spacetimes.



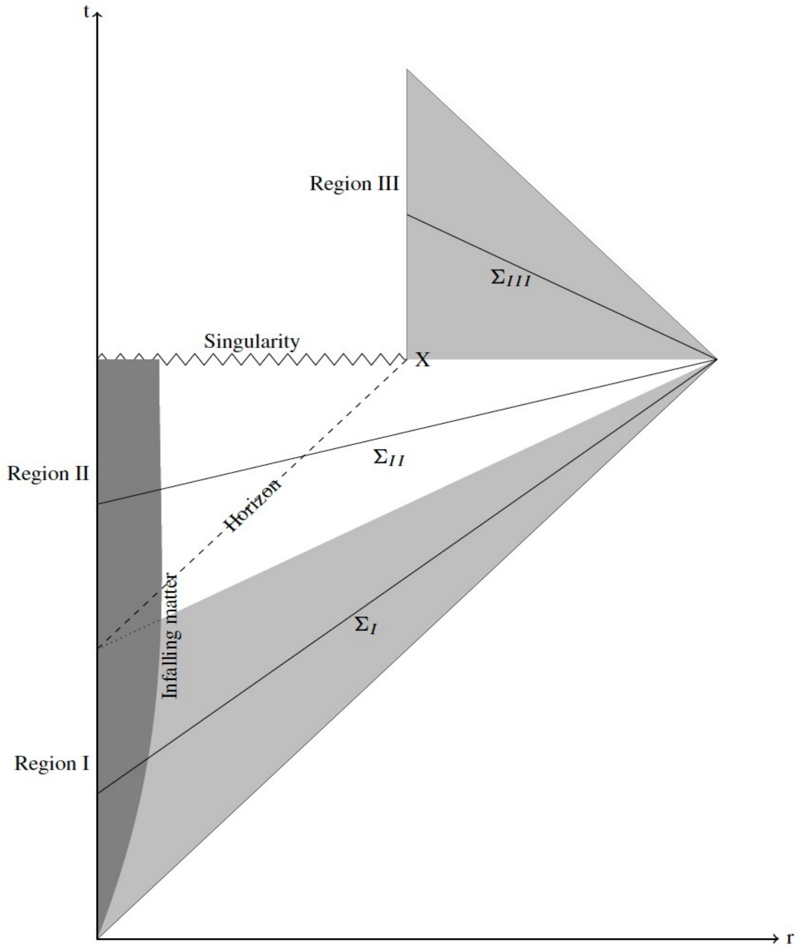
- *If Σ is Cauchy then all non-spacelike (causal) worldlines interacting in R must register on Σ.*
- *The data on Σ completely determines what goes on in R.*

Argument

(1) For the state of a physical system in a background spacetime to change via a unitary transformation, the spacetime must be *globally hyperbolic*.

(2) The spacetime of an evaporating black hole is not globally hyperbolic.

(3) Thus the state of an evaporating black hole cannot change via a unitary transformation.



- *Evaporating black hole spacetime* = (Region I) ∪ (Region II) ∪ (Region III)
- Σ_{I} *and* Σ_{II} *are Cauchy surfaces for* Region I *and* Region II *individually.*
- Σ_{III} *is not a Cauchy surface for* Region III*: X is a naked singularity.*
- Σ_{I}, Σ_{II}, Σ_{III} *are not Cauchy surfaces for the complete spacetime.*

A conflict between general relativity and the unitarity of quantum mechanics, after complete evaporation.

3. The Page Time Paradox

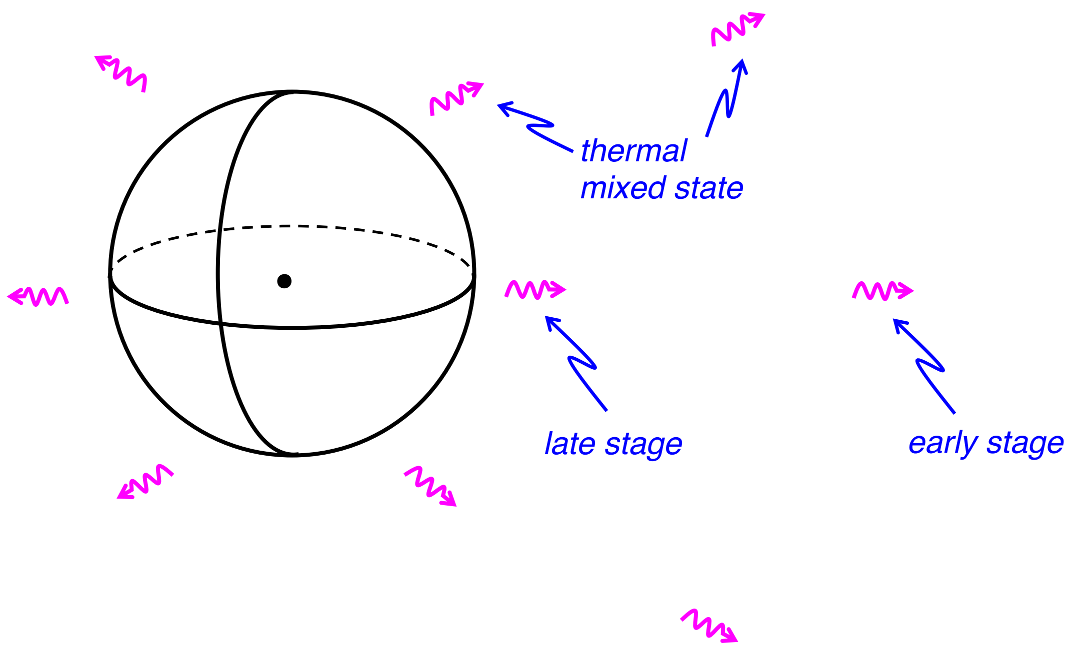
A conflict between quantum field theory and statistical mechanics, prior to complete evaporation.

(H1) Radiation modes are in thermal mixed states at all times.

One result of Hawking's (1975) quantum field theory analysis

(P) Late stage radiation modes are maximally entangled with early stage radiation modes.

Page's (1993) statistical mechanics result



Let's convince ourselves that (H1) *and* (P) *are contradictory...*

Def. (*Thermal state*). A **thermal density operator state** is a mixed state of energy eigenvector states $|E_i\rangle$ in thermal equilibrium at temperature T. It takes the form

$$\rho = Z^{-1}\sum_i e^{-\beta E_i}|E_i\rangle\langle E_i|$$

where $\beta = 1/T$, and $Z = \sum_i e^{-\beta E_i}$.

- *Recall*: A density operator state is associated with an ensemble of vector states $\{|\psi_i\rangle, p_i\}$, each with probablity p_i.
  - The ensemble of vector states associated with a thermal density operator state is $\{|E_i\rangle, p_i\}$ where the probabilities p_i are given by the classical canonical Gibbs distribution $\rho_c(x) = Z^{-1}e^{-\beta E(x)}$, where $Z = \int e^{-\beta E(x)}dx$.

Recall: A classical canonical Gibbs distribution characterizes a classical composite closed system consisting of a subsystem in thermal equilibrium with a heat bath: constant temperature, varying energy.

Def. (*Thermal state*). A **thermal density operator state** is a mixed state of energy eigenvector states $|E_i\rangle$ in thermal equilibrium at temperature T. It takes the form

$$\rho = Z^{-1}\sum_i e^{-\beta E_i}|E_i\rangle\langle E_i|$$

where $\beta = 1/T$, and $Z = \sum_i e^{-\beta E_i}$.

- *Note*: A thermal density operator state can also be expressed by

$$\rho = Z^{-1}e^{-\beta H}$$

for $Z = \text{Tr}(e^{-\beta H})$ and Hamiltonian operator H satisfying $H|E_j\rangle = E_j|E_j\rangle$.

Motivation: $e^{-\beta H}$ and $\sum_i e^{-\beta E_i}|E_i\rangle\langle E_i|$ do the same thing to an arbitrary energy eigenvector state $|E_j\rangle$:

$$e^{-\beta H}|E_j\rangle = [\text{I} + \beta H + ½(\beta H)^2 + \cdots]|E_j\rangle$$

$e^x = 1 + x + \frac{x^2}{2!} + \cdots$

$$= [\text{I} + \beta E_j + ½(\beta E_j)^2 + \cdots]|E_j\rangle$$

$$= e^{-\beta E_j}|E_j\rangle$$

$$\sum_i e^{-\beta E_i}|E_i\rangle\langle E_i|E_j\rangle = e^{-\beta E_j}|E_j\rangle$$

Claim NTSE (*No Thermal State Entanglement*). A thermal state cannot be entangled with another thermal state.

(*<u>More precisely</u>*: If A and B are systems in thermal states at the same temperature, then if the composite system AB is in a thermal state at the same temperature, then AB cannot be in an entangled state.)

<u>Proof</u>. Let AB be a composite system with subsystems A, B, all in thermal states at temperature β^{-1}. Then:

(i) $\rho_A = {Z_A}^{-1} e^{-\beta H_A}, \;\; \rho_B = {Z_B}^{-1} e^{-\beta H_B}$

(ii) $\rho_{AB} = {Z_{AB}}^{-1} e^{-\beta(H_A + H_B + H_{AB})} \sim \rho_A \otimes \rho_B \otimes e^{-\beta H_{AB}}$ ← H_{AB} encodes possible interactions between A and B

(iii) $\rho_A = \mathrm{Tr}_B \rho_{AB}, \;\; \rho_B = \mathrm{Tr}_A \rho_{AB}$

- The combination of (i), (ii), and (iii) entails $H_{AB} = 0$, which means $\rho_{AB} = \rho_A \otimes \rho_B$; so ρ_{AB} is a product (non-entangled) state.
- *<u>So</u>*: If AB, A, and B are all in thermal states at the same temperature, then AB is in a product state, and hence A and B cannot be entangled.

So:

(H1) Radiation modes are in thermal mixed states at all times.

(P) Late stage radiation modes are maximally entangled with early stage radiation modes.

(NTSE) A thermal state cannot be entangled with another thermal state.

(NTSE) *entails* (H1) *and* (P) *cannot both be correct!*

- *Moreover*: The conflict occurs *before* complete evaporation at t_{evap}.

  It occurs as soon as there is a distinction between "late stage" and "early stage" radiation!

- *So*: There must be a time, call it the "Page time" t_{Page}, $0 < t_{\text{Page}} < t_{\text{evap}}$, afterwhich the quantum field theory prediction (H1) conflicts with the statistical mechanical prediction (P).

Let's look more closely at Page's statistical mechanical prediction (P)...

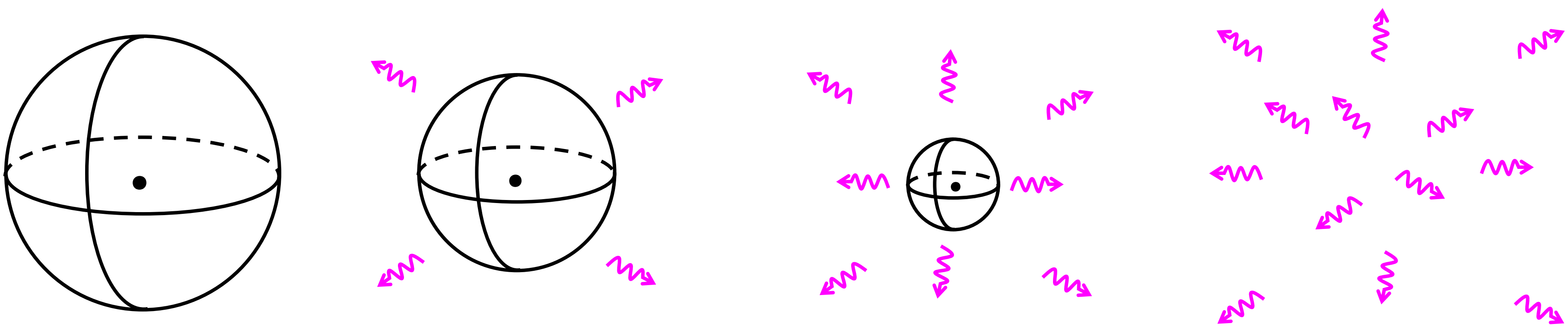
Assumption 1: An evaporating black hole can be described by an ensemble of states with microcanonical Gibbs distribution $\rho_{\rm mc} = 1/\Omega$.

Recall: The microcanonical Gibbs entropy, call it $S_{\rm mc}$, is given by $S_{\rm mc} = \ln\Omega$, where Ω = #micro-states with a given energy.

Assumption 2: An evaporating black hole can be described as a quantum mechanical bipartite system BR consisting of black hole B and Hawking radiation R with states in $\mathcal{H}_{\rm BR} = \mathcal{H}_{\rm B} \otimes \mathcal{H}_{\rm R}$.

- *The number of states in $\mathcal{H}$ is given by its dimension $|\mathcal{H}|$.*
- *So: $S^{\rm B}_{\rm mc} = \ln|\mathcal{H}_{\rm B}|$ and $S^{\rm R}_{\rm mc} = \ln|\mathcal{H}_{\rm R}|$*

Relation between $S^{\rm B}_{\rm mc}$ and $S^{\rm R}_{\rm mc}$



$t = 0$, $S^{\rm BR}_{\rm mc} = S^{\rm B}_{\rm mc}$ | $S^{\rm R}_{\rm mc} \ll S^{\rm B}_{\rm mc}$ | $S^{\rm R}_{\rm mc} \gg S^{\rm B}_{\rm mc}$ | $t = t_{\rm evap}$, $S^{\rm BR}_{\rm mc} = S^{\rm R}_{\rm mc}$

Let's look more closely at Page's statistical mechanical prediction (P)...

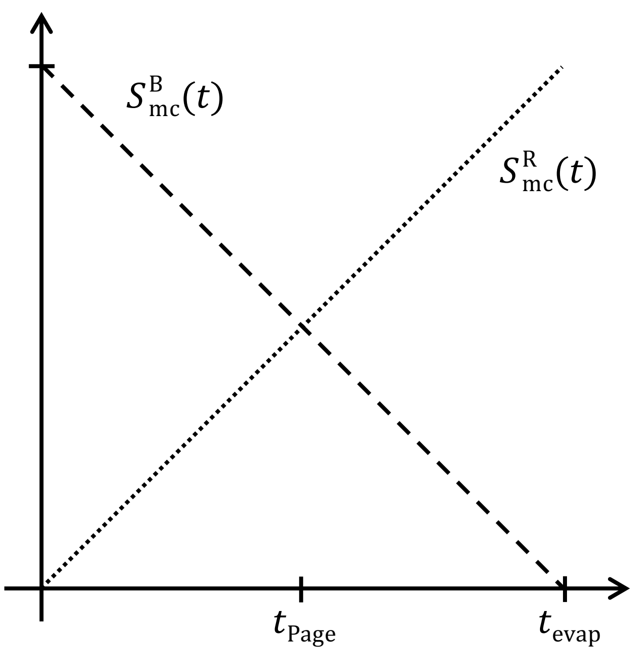
Assumption 1: An evaporating black hole can be described by an ensemble of states with microcanonical Gibbs distribution $\rho_{\mathrm{mc}} = 1/\Omega$.

Recall: The microcanonical Gibbs entropy, call it S_{mc}, is given by $S_{\mathrm{mc}} = \ln \Omega$, where Ω = #micro-states with a given energy.

Assumption 2: An evaporating black hole can be described as a quantum mechanical bipartite system BR consisting of black hole B and Hawking radiation R with states in $\mathcal{H}_{\mathrm{BR}} = \mathcal{H}_{\mathrm{B}} \otimes \mathcal{H}_{\mathrm{R}}$.

- *The number of states in $\mathcal{H}$ is given by its dimension $|\mathcal{H}|$.*
- *So: $S^{\mathrm{B}}_{\mathrm{mc}} = \ln|\mathcal{H}_{\mathrm{B}}|$ and $S^{\mathrm{R}}_{\mathrm{mc}} = \ln|\mathcal{H}_{\mathrm{R}}|$*

Relation between $S^{\mathrm{B}}_{\mathrm{mc}}$ and $S^{\mathrm{R}}_{\mathrm{mc}}$



Page time t_{Page} is half-way point of evaporation at which $S^{\mathrm{B}}_{\mathrm{mc}}(t) = S^{\mathrm{R}}_{\mathrm{mc}}(t)$.

- *Now*: Let's determine how the entanglement entropy $S_{\mathrm{vN}}(\rho_{\mathrm{R}})$, or $S_{\mathrm{vN}}^{\mathrm{R}}$, of the radiation changes with time.

Claim 1: *If* the composite initial state of BR is pure, then $S_{\mathrm{vN}}^{\mathrm{R}}(t) = S_{\mathrm{vN}}^{\mathrm{B}}(t)$ for all t.

Claim 2: For all t,
(a) $S_{\mathrm{vN}}^{\mathrm{R}}(t) \leq S_{\mathrm{mc}}^{\mathrm{R}}(t)$
(b) $S_{\mathrm{vN}}^{\mathrm{B}}(t) \leq S_{\mathrm{mc}}^{\mathrm{B}}(t)$

Recall: The maximum value of $S_{\mathrm{vN}}(\rho)$ for a density operator state ρ on an n-dim Hilbert space $\mathcal{H}$ is $\ln n$, or $\ln|\mathcal{H}|$, and this is $S_{\mathrm{mc}}(\rho)$!

Claim 3: A randomly chosen pure state in a product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ is likely to be very close to maximally entangled as long as $|\mathcal{H}_A| \ll |\mathcal{H}_B|$.

Proven by Page (1993).

(i) For $0 < t < t_{\mathrm{Page}}$, $|\mathcal{H}_{\mathrm{R}}| \ll |\mathcal{H}_{\mathrm{B}}|$.

- *By Claim 3*: $S_{\mathrm{vN}}^{\mathrm{R}}(t)$ must be very close to its maximum value.

  Substates of a max. entangled state are max. mixed

- *By Claims 1 & 2*: The max value of $S_{\mathrm{vN}}^{\mathrm{R}}(t)$ is either $S_{\mathrm{mc}}^{\mathrm{R}}(t)$ or $S_{\mathrm{mc}}^{\mathrm{B}}(t)$.
- *By Claim 2a*: It can't be $S_{\mathrm{mc}}^{\mathrm{B}}(t)$, since $S_{\mathrm{mc}}^{\mathrm{B}}(t) \gg S_{\mathrm{mc}}^{\mathrm{R}}(t)$.
- *So*: The max value of $S_{\mathrm{vN}}^{\mathrm{R}}(t)$ must be $S_{\mathrm{mc}}^{\mathrm{R}}(t)$.
- *So*: $S_{\mathrm{vN}}^{\mathrm{R}}(t) \approx S_{\mathrm{mc}}^{\mathrm{R}}(t)$.

- *Now*: Let's determine how the entanglement entropy $S_{\mathrm{vN}}(\rho_{\mathrm{R}})$, or $S_{\mathrm{vN}}^{\mathrm{R}}$, of the radiation changes with time.

Claim 1: *If* the composite initial state of BR is pure, then $S_{\mathrm{vN}}^{\mathrm{R}}(t) = S_{\mathrm{vN}}^{\mathrm{B}}(t)$ for all t.

Claim 2: For all t,
(a) $S_{\mathrm{vN}}^{\mathrm{R}}(t) \leq S_{\mathrm{mc}}^{\mathrm{R}}(t)$
(b) $S_{\mathrm{vN}}^{\mathrm{B}}(t) \leq S_{\mathrm{mc}}^{\mathrm{B}}(t)$

Recall: The maximum value of $S_{\mathrm{vN}}(\rho)$ for a density operator state ρ on an n-dim Hilbert space $\mathcal{H}$ is $\ln n$, or $\ln|\mathcal{H}|$, and this is $S_{\mathrm{mc}}(\rho)$!

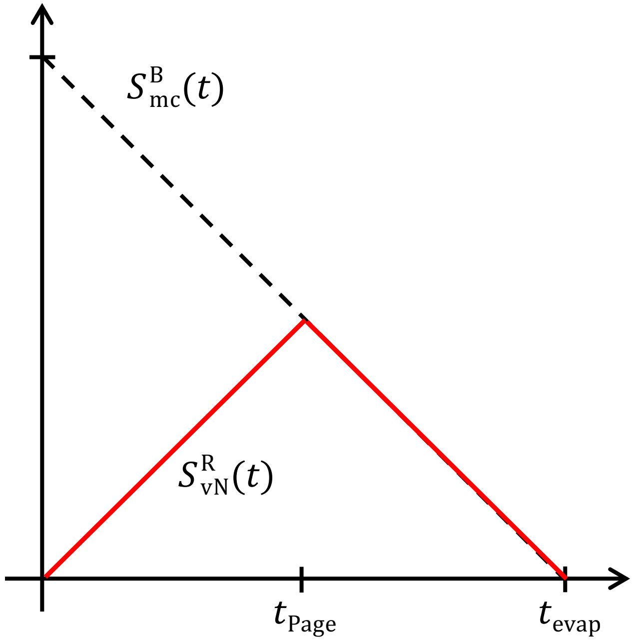
Claim 3: A randomly chosen pure state in a product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ is likely to be very close to maximally entangled as long as $|\mathcal{H}_A| \ll |\mathcal{H}_B|$.

Proven by Page (1993).

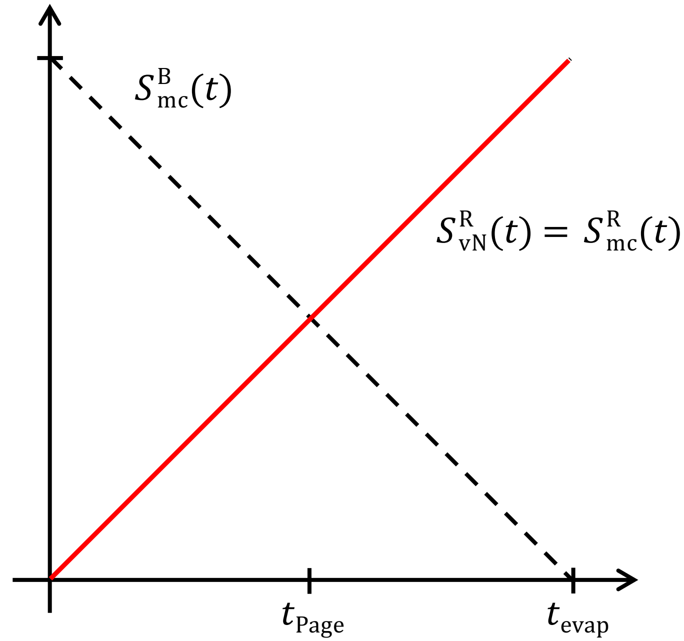
(ii) For $t_{\mathrm{Page}} < t < t_{\mathrm{evap}}$, $|\mathcal{H}_{\mathrm{B}}| \ll |\mathcal{H}_{\mathrm{R}}|$.

- *By Claim 3*: $S_{\mathrm{vN}}^{\mathrm{B}}(t)$ must be very close to its maximum value. *Substates of a max. entangled state are max. mixed*
- *By Claims 1 & 2*: The max value of $S_{\mathrm{vN}}^{\mathrm{B}}(t)$ is either $S_{\mathrm{mc}}^{\mathrm{R}}(t)$ or $S_{\mathrm{mc}}^{\mathrm{B}}(t)$.
- *By Claim 2b*: It can't be $S_{\mathrm{mc}}^{\mathrm{R}}(t)$, since $S_{\mathrm{mc}}^{\mathrm{R}}(t) \gg S_{\mathrm{mc}}^{\mathrm{B}}(t)$.
- *So*: The max value of $S_{\mathrm{vN}}^{\mathrm{B}}(t)$ must be $S_{\mathrm{mc}}^{\mathrm{B}}(t)$.
- *So*: $S_{\mathrm{vN}}^{\mathrm{B}}(t) \approx S_{\mathrm{mc}}^{\mathrm{B}}(t)$.
- *By Claim 1*: $S_{\mathrm{vN}}^{\mathrm{R}}(t) \approx S_{\mathrm{mc}}^{\mathrm{B}}(t)$.

The "Page curve" for $S^{\mathrm{R}}_{\mathrm{vN}}(t)$



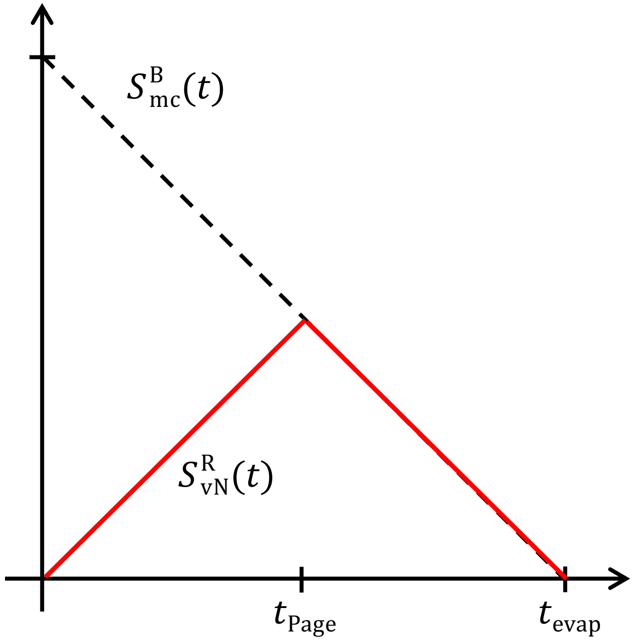
Hawking's prediction for $S^{\mathrm{R}}_{\mathrm{vN}}(t)$



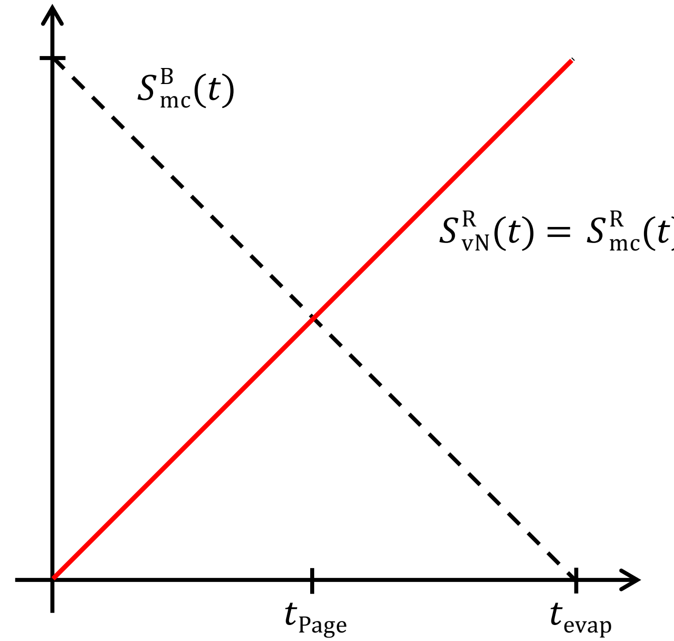
- *According to Page*: $S^{\mathrm{R}}_{\mathrm{vN}}(t)$ behaves like $S^{\mathrm{R}}_{\mathrm{mc}}(t)$ before t_{Page}, and like $S^{\mathrm{B}}_{\mathrm{mc}}(t)$ after t_{Page}.
  - Before t_{Page}, $S^{\mathrm{R}}_{\mathrm{vN}}(t)$ increases as more and more radiation modes entangled with black hole modes are emitted.
  - After t_{Page}, radiation modes continue to be emitted, but $S^{\mathrm{R}}_{\mathrm{vN}}(t)$ decreases, and *this can only be possible if these late stage radiation modes are now entangled with early stage radiation modes, and not black hole modes.* ← This is (P)!

- *According to Hawking*: Radiation is in thermal mixed states at all times, so $S^{\mathrm{R}}_{\mathrm{vN}}(t)$ should steadily increase as more and more radiation modes are emitted.
  - *Which entails*: $S^{\mathrm{R}}_{\mathrm{vN}}(t)$ behaves like $S^{\mathrm{R}}_{\mathrm{mc}}(t)$ at all times.

The "Page curve" for $S^{\mathrm{R}}_{\mathrm{vN}}(t)$



Hawking's prediction for $S^{\mathrm{R}}_{\mathrm{vN}}(t)$



- *Aside*: The Page curve entails that if the composite BR state starts out pure, the final completely evaporated BR state remains pure.
  - *Unitarity is preserved!*
  - *"Information" is not lost!*
- *How does "information" escape?*
  - By becoming encoded in the entanglement correlations between late stage and early stage radiation after t_{Page}? ← What explains this?

3. The Firewall Paradox

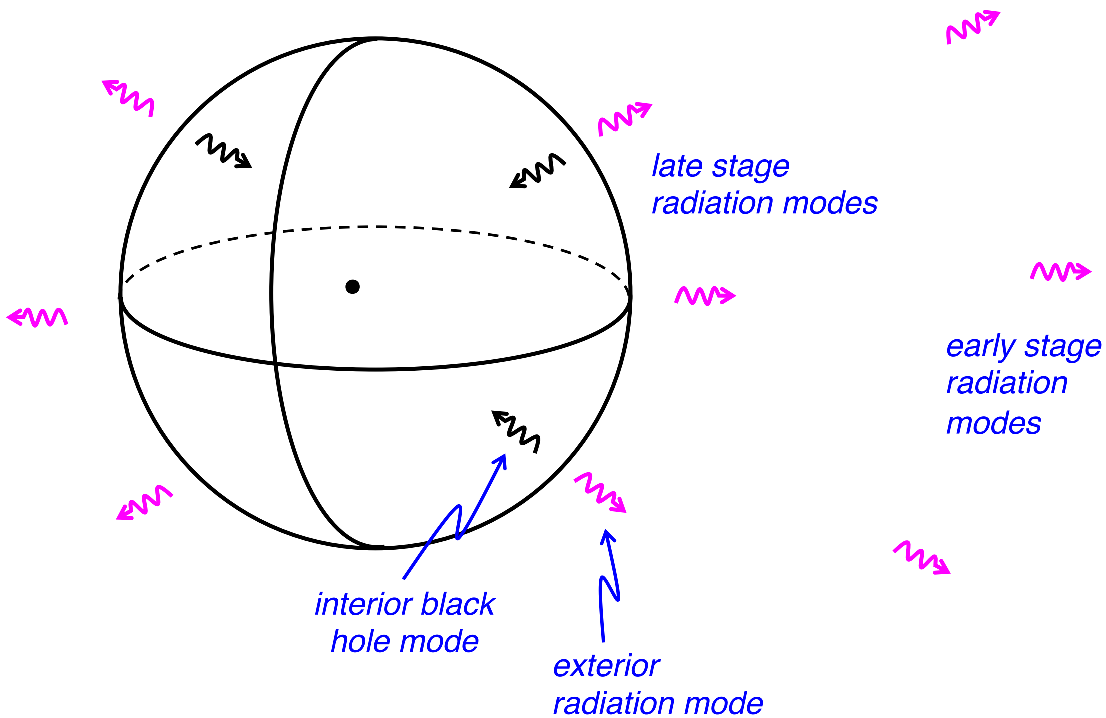
Almheiri, Marolf, Polchinski, Sully (2013) or "AMPS"

(H2) Radiation modes outside the event horizon are maximally entangled with black hole degrees of freedom inside the event horizon.

← *Another result of Hawking's (1975) quantum field theory analysis*

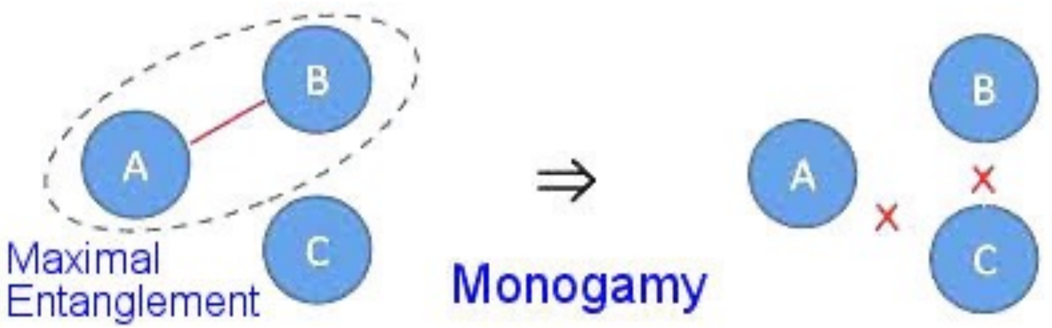
(P) Late stage radiation modes are maximally entangled with early stage radiation modes.

← *Page's (1993) statistical mechanics result*



Claim: (H2), (P), *and "entanglement monogamy" are contradictory...*

Entanglement Monogamy (EM): A physical system cannot be maximally entangled with two other systems.
(*More precisely*: If ρ_{ABC} is a pure density operator state of a tripartite system that consists of subsystems A, B, C, and ρ_{AB} is a pure maximally entangled density operator state of the joint subsystem of A and B, then C cannot be entangled with either A or B.)



Proof. Suppose ρ_{ABC} is a pure density operator state of a tripartite system ABC, and suppose ρ_{AB} is pure and maximally entangled. Now suppose C is entangled with either A or B.

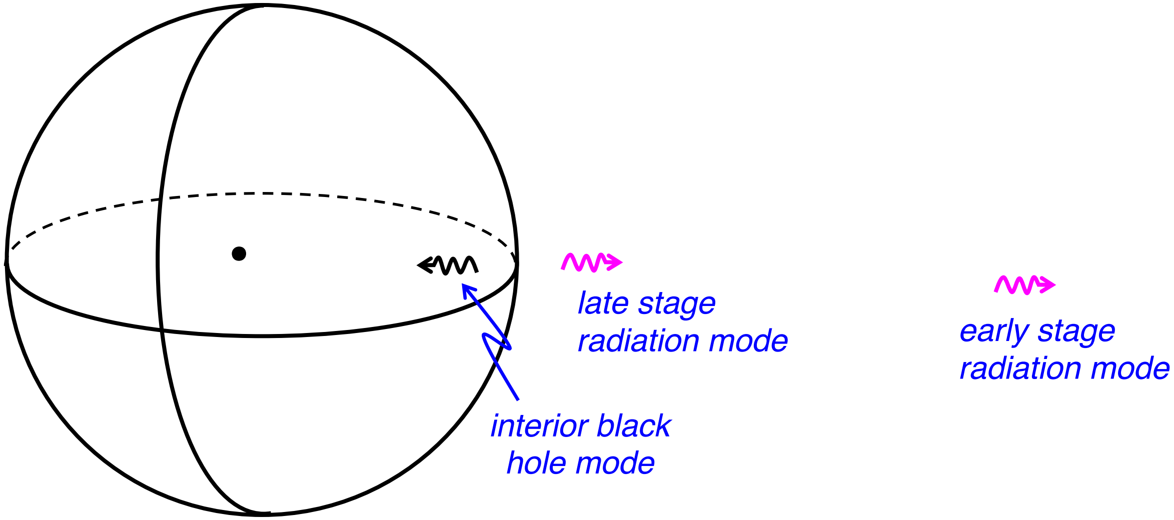
- *Then*: $\mathrm{Tr}_C \rho_{ABC}$ must be a mixed density operator state.
- *But*: ρ_{AB} is assumed to be pure.

Claim: (H2), (P), *and "entanglement monogamy" are contradictory...*

(a) For any late stage radiation mode just outside the event horizon, there is an interior mode maximally entangled with it. (H2)

(b) For any late stage radiation mode just outside the event horizon, there is an early stage radiation mode far from the horizon maximally entangled with it. (P)

(c) Thus, any late stage radiation mode is max entangled with both an interior mode and an early stage radiation mode. ← *A violation of* (EM)*!*



Claim: (H2), (P), *and "entanglement monogamy" are contradictory...*

(a) For any late stage radiation mode just outside the event horizon, there is an interior mode maximally entangled with it. (H2)

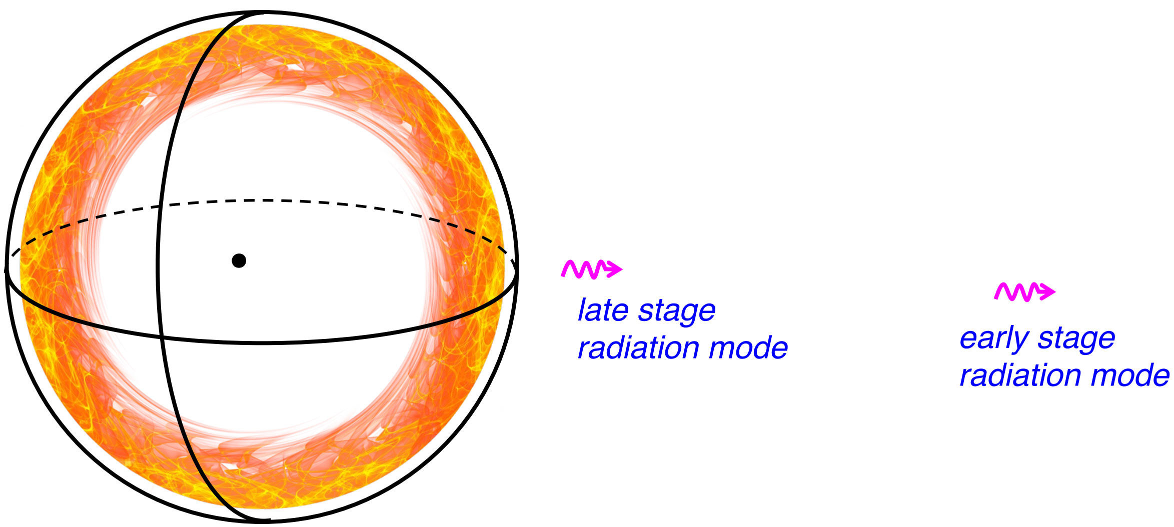
(b) For any late stage radiation mode just outside the event horizon, there is an early stage radiation mode far from the horizon maximally entangled with it. (P)

(c) Thus, any late stage radiation mode is max entangled with both an interior mode and an early stage radiation mode.

A violation of (EM)*!*

AMPS proposal: Deny (a).

- Suppose every late stage radiation mode corresponds to a high energy interior mode, with the aggregate of all the interior modes constituting a "firewall".



The firewall prevents late stage radiation modes from being entangled with interior modes!

Claim: (H2), (P), *and "entanglement monogamy" are contradictory...*

(a) For any late stage radiation mode just outside the event horizon, there is an interior mode maximally entangled with it. (H2)

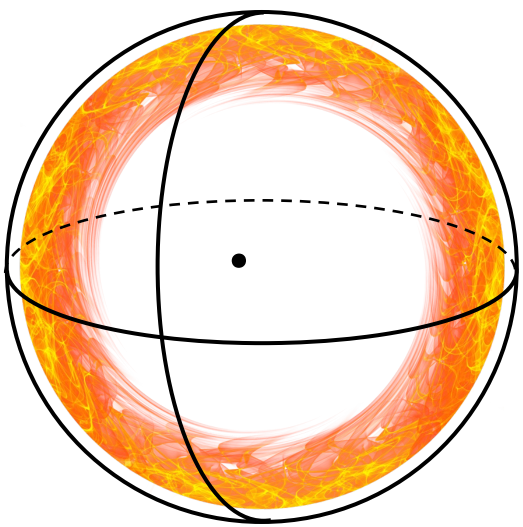
(b) For any late stage radiation mode just outside the event horizon, there is an early stage radiation mode far from the horizon maximally entangled with it. (P)

(c) Thus, any late stage radiation mode is max entangled with both an interior mode and an early stage radiation mode.

A violation of (EM)*!*

AMPS proposal: Deny (a).

- Suppose every late stage radiation mode corresponds to a high energy interior mode, with the aggregate of all the interior modes constituting a "firewall".



Is this even plausible?!?

- *Yes! According to QFT in curved spacetime, the energy momentum tensor of a quantum field on one side of an event horizon diverges (becomes infinite) with respect to field modes on the other side.*

But: Maybe just an indication that QFT in curved spacetimes is an incomplete mash-up?