Black Holes: Entropy Bounds, Complementarity, and Firewalls

Raphael Bousso

Berkeley Center for Theoretical Physics
University of California, Berkeley

CERN Winter School, February 4–8, 2013

DISCLAIMER

These slides were prepared for lecturing, not for publication. They are made available upon request by the organizers. There are figures missing that were drawn on the blackboard. No attempt is made at providing references systematically, nor at conveying a full or unbiased history of the subject; many important contributions and contributors go unmentioned. Please don't be offended if this includes you.

Black holes

- Gravitational collapse of stars greater than about 1.3 solar masses leads to the formation of a black hole. (For smaller masses the star may stabilize as a white dwarf or neutron star.)
- There is indirect observational evidence both for stellar mass ($\sim 10^1 M_\odot$) and supermassive ($10^6 - 10^{10} M_\odot$) black holes.
- There appears to be no obstruction in principle to the formation of black holes of any mass above the Planck mass, $M_P \sim 10^{-5}$g, e.g., by sufficiently energetic collisions.

Schwarzschild Metric

- In practice the formation of a black hole is messy, but in simple models it can be understood analytically.
- Oppenheimer-Snyder collapse of a uniform ball of dust: Use closed FRW universe, $ds^2 = -dT^2 + a^2(T)[d\chi^2 + \sin^2\chi d\Omega_2^2]$; keep region up to some comoving radius; use Birkhoff's theorem. Outside the dust ball, the geometry is given by the vacuum solution found by Karl Schwarzschild:

$$ds^2 = -(1 - \frac{r_S}{r})dt^2 + \frac{1}{1 - \frac{r_S}{r}}dr^2 + r^2 d\Omega^2 \ .$$

- The mass as measured at infinity is $M = r_S/2G$. By energy conservation, the Schwarzschild radius r_S is set by the mass of the dust ball.

Kruskal Extension

The Schwarzschild metric breaks down at $r = r_S$ but this is a coordinate singularity $\rightarrow$ Kruskal extension

Kruskal–Szekeres coordinates are defined, from the Schwarzschild coordinates (t, r, θ, ϕ), by replacing *t* and *r* by a new time coordinate V and a new spatial coordinate U:

$$V = \left(\frac{r}{2GM} - 1\right)^{1/2} e^{r/4GM} \sinh\left(\frac{t}{4GM}\right)$$

$$U = \left(\frac{r}{2GM} - 1\right)^{1/2} e^{r/4GM} \cosh\left(\frac{t}{4GM}\right)$$

for the exterior region $r > 2GM$, and:

$$V = \left(1 - \frac{r}{2GM}\right)^{1/2} e^{r/4GM} \cosh\left(\frac{t}{4GM}\right)$$

$$U = \left(1 - \frac{r}{2GM}\right)^{1/2} e^{r/4GM} \sinh\left(\frac{t}{4GM}\right)$$

for the interior region $0 < r < 2GM$. Note *GM* is the gravitational constant multiplied with the Schwarzschild mass parameter, and this article is using units where *c* = 1.

Kruskal Extension

The Schwarzschild metric breaks down at $r = r_S$ but this is a coordinate singularity $\longrightarrow$ Kruskal extension

It follows that the Schwarzschild r, in terms of Kruskal-Szekeres coordinates, is implicitly given by:

$$V^2 - U^2 = \left(1 - \frac{r}{2GM}\right) e^{r/2GM}$$

In these new coordinates the metric of the Schwarzschild black hole manifold is given by

$$ds^2 = \frac{32G^3M^3}{r} e^{-r/2GM}(-dV^2 + dU^2) + r^2 d\Omega^2,$$

written using the (− + + +) metric signature convention and where the angular component of the metric (the line element of the 2-sphere) is:

$$d\Omega^2 \stackrel{\text{def}}{=} d\theta^2 + \sin^2\theta \, d\phi^2$$

The location of the event horizon ($r = 2GM$) in these coordinates is given by $V = \pm U$. No

Key Features

- There is a curvature singularity at $r = 0$, where GR breaks down.
- At late times the proper size of the dust ball approaches zero; in particular, it will be smaller than r_S.
- No signals can travel from $r < r_S$ to infinity.
- From the outside, therefore, the spacetime appears to end at $r = r_S$ and is entirely described by the Schwarzschild solution.
- The Hawking-Penrose singularity theorems and the no-hair theorem guarantee that more realistic collapse processes will share these features.

Geography

- Singularity ($r = 0$)
- Interior ($r < r_S$):
  - Collapsing matter
  - Vacuum
- Horizon($r = r_S$)
- Exterior ($r > r_S$)
  - Near-horizon region ("zone", $r - r_S \lesssim r_s$)
  - Far region ($r - r_S \gg r_S$)

Amnesia

- Quasi-normal modes decay exponentially with characteristic time-scale r_S
- After a time of order $r_S \log r_S$ (the "scrambling time"), the the black hole is indistinguishable from its no-hair state
- After the scrambling time, even infalling observers cannot distinguish the solution from vacuum Kerr
- Signals must be sent within a time $r_S \exp(-t/R_S)$ after horizon crossing, in order to be received by an observer who jumps in at time t, before that observer hits the singularity [HOMEWORK or see hep-th/9308100]

The Black Hole, Messing With Our Little Brains

In the early 1970s, black holes were found to behave according to laws that were *exactly* analogous to the laws of thermodynamics, with $T \leftrightarrow \kappa \sim r_S^{-1}$, $S \leftrightarrow A$, $E \leftrightarrow M = r_S/2G$:

- 0th Law: The temperature of a body in equilibrium is constant $\leftrightarrow$ κ constant over horizon
- 1st Law: $dS = dE/T \leftrightarrow dA/8\pi = dM/\kappa$
- 2nd Law: $dA \geq 0$ (Hawking's area theorem)

Take this seriously? No. Otherwise, black holes would have to radiate, which is impossible [Bardeen, Carter, Hawking 1973]

Bekenstein's Argument

LETTERE AL NUOVO CIMENTO VOL. 4, N. 15 12 Agosto 1972

Black Holes and the Second Law (*).

J. D. BEKENSTEIN (**)

Joseph Henry Laboratories, Princeton University - Princeton, N. J.

(ricevuto il 22 Maggio 1972)

Bekenstein's Argument

Black-hole physics seems to provide at least two ways in which the second law of thermodynamics may be transcended or violated:

a) Let an observer drop or lower a package of entropy into a black hole; the entropy of the exterior world decreases. Furthermore, from an exterior observer's point of view a black hole in equilibrium has only three degrees of freedom: mass, charge and angular momentum ([1]). Thus, once the black hole has settled down to equilibrium, there is no way for the observer to determine its interior entropy. Therefore, he cannot exclude the possibility that the total entropy of the universe may have decreased in the process. It is in this sense that the second law appears to be transcended ([2]).

Bekenstein's Argument

We state the second law as follows: « Common entropy plus black-hole entropy never decreases. » By common entropy we mean entropy in the black-hole's exterior. By black-hole entropy $S_{\text{b.h.}}$ we mean

$$S_{\text{b.h.}} = \eta k L_{\text{P}}^{-2} A \,, \tag{1}$$

where A is the area of the black hole (4) in question, L_{P} is the Planck length $(\hbar G/c^3)^{\frac{1}{2}} = 1.6 \cdot 10^{-33}$ cm, k is Boltzmann's constant $1.38 \cdot 10^{-16}$ erg/°K and η is a constant number of order unity.

Bekenstein's Argument

The introduction of a black-hole entropy is necessitated by our first example. Without it the second law is definitely transcended. With black-hole entropy the second law becomes a well-defined statement susceptible to verification by an exterior observer. The choice of the area of a black hole as a measure of its entropy is motivated by the results of CHRISTODOULOU ([5]) and of HAWKING ([6]) that the area of a black hole never decreases, and that it increases for all but a very special class of black-hole transformations ([5]). The area of a black hole appears to be the only one of its properties having this entropylike behavior which is so essential if the second law as we have stated it is to hold when entropy goes down a black hole.

By choosing $S_{b.h.}$ proportional to A rather than to some monotonically increasing function of A, we ensure that the total black-hole entropy of a system of black holes (the sum of individual $S_{b.h.}$) depends only on the total horizon area—also a nondecreasing quantity which is of fundamental importance in Hawking's work ([6]).

Bekenstein's Argument

We conclude on a speculative note. One sees from (1) that the natural unit of area of a black hole is the Planck length squared. One cannot help wondering about the possible connection between this feature and the expected quantum structure of space-time at a scale of the order of the Planck length ([12]).

Lessons*

- Ask the right question in the right way (complementarity is implicit!)
- To find the beautiful subtle answer, you first have to reject the obvious dumb one
- Know which objections to ignore for now, even though they seem devastating and you have no good answer ($T^{-1} = dS/dE$)

Lessons*

- Ask the right question in the right way (complementarity is implicit!)
- To find the beautiful subtle answer, you first have to reject the obvious dumb one
- Know which objections to ignore for now, even though they seem devastating and you have no good answer ($T^{-1} = dS/dE$)

*Warning to grad students (and self): this almost never works.

Bekenstein, again

- Throw an object with entropy S into a black hole
- By the "no-hair theorem", the final result will be a (larger) black hole, with no classical attributes other than mass, charge, and angular momentum
- This state would appear to have no or negligible entropy, independently of S
- So we have a process in which $dS < 0$
- The Second Law of Thermodynamics appears to be violated!

Bekenstein, again

- In order to rescue the Second Law, Bekenstein proposed that black holes themselves carry entropy
- Hawking (1971) had already proven the "area law", which states that black hole horizon area never decreases in any process:

$$dA \geq 0$$

- So the horizon area seemed like a natural candidate for black hole entropy. On dimensional grounds, the entropy would have to be of order the horizon area in Planck units:

$$S_{\mathrm{BH}} \sim A$$

Bekenstein, again

- In order to rescue the Second Law, Bekenstein proposed that black holes themselves carry entropy
- Hawking (1971) had already proven the "area law", which states that black hole horizon area never decreases in any process:

$$dA \geq 0$$

- So the horizon area seemed like a natural candidate for black hole entropy. On dimensional grounds, the entropy would have to be of order the horizon area in Planck units:

$$S_{\mathrm{BH}} \sim A$$

- (This was later confirmed, and the factor $1/4$ determined, by Hawking's calculation of the temperature and the relation $dS = dE/T$)

The Generalized Second Law

- With black holes carrying entropy, it is no longer obvious that the total entropy decreases when a matter system is thrown into a black hole
- Bekenstein proposed that a Generalized Second Law of Thermodynamics remains valid in processes involving the loss of matter into black holes
- The GSL states that $dS_{\text{total}} \geq 0$, where

$$S_{\text{total}} = S_{\text{BH}} + S_{\text{matter}}$$

Is the GSL true?

- However, it is not obvious that the GSL actually holds!
- The question is whether the black hole horizon area increases by enough to compensate for the lost matter entropy
- If the initial and final black hole area differ by ΔA, is it true that

$$S_{\text{matter}} \leq \Delta A/4 \ ?$$

- Note that this would have to hold for all types of matter and all ways of converting the matter entropy into black hole entropy!

Testing the GSL

- Let's do a few checks to see if the GSL might be true
- There are two basic processes we can consider:
- Dropping a matter system to an existing black hole, and
- Creating a new black hole by compressing a matter system or adding mass to it

Testing the GSL

- Let's do a few checks to see if the GSL might be true
- There are two basic processes we can consider:
- Dropping a matter system to an existing black hole, and
- Creating a new black hole by compressing a matter system or adding mass to it
- Let's consider an example of the second type

Testing the GSL

- Spherical box of radius R, filled with radiation at temperature at temperature T, which we slowly increase
- Let Z be the effective number of massless particle species
- $S \equiv S_{\text{matter}} \approx ZR^3T^3$, so the entropy increases arbitrarily?!
- However, the box cannot be stable if its mass, $M \approx ZR^3T^4$, exceeds the mass of a black hole of the same radius, $M \approx R$.
- A black hole must form when $T \approx Z^{-1/4}R^{-1/2}$. Just before this point, the matter entropy is

$$S \approx Z^{1/4}A^{3/4}$$

Testing the GSL

- After the black hole forms, the matter entropy is gone and the total entropy is given by the black hole horizon area, $S = A/4$.
- This is indeed larger than the initial entropy, $Z^{1/4}A^{3/4}$, as long as $A \gtrsim Z$, which is just the statement that the black hole is approximately a classical object.
- (We require this in any case since we wish to work in a setting where classical gravity is a good description.)
- So in this example the GSL is satisfied

Entropy bounds from the GSL

- In more realistic examples, such as the formation of black holes by the gravitational collapse of a star, the GSL is upheld with even more room to spare
- As our confidence in the GSL grows, it is tempting to turn the logic around and assume the GSL to be true
- Then we can derive a bound on the entropy of arbitrary matter systems, namely

$$S_{\text{matter}} \leq \Delta A/4 \ ,$$

where ΔA is the increase in horizon entropy when the matter system is converted into or added to a black hole

Spherical entropy bound

- For example, consider an arbitrary spherical matter system of mass m that fits within a sphere of area $A \sim R^2$.
- We could presumably collapse a shell of mass $R/G - m$ around this system to convert it into a black hole, also of area A
- The GSL implies that $S_{\text{matter}} \leq A/4$, i.e., that the entropy of any matter system is less than the area of the smallest sphere that encloses it

Spherical entropy bound

- For example, consider an arbitrary spherical matter system of mass m that fits within a sphere of area $A \sim R^2$.
- We could presumably collapse a shell of mass $R/G - m$ around this system to convert it into a black hole, also of area A
- The GSL implies that $S_{\text{matter}} \leq A/4$, i.e., that the entropy of any matter system is less than the area of the smallest sphere that encloses it
- In this sense the world is like a hologram!
- The amount of information needed to fully specify the quantum state in a spherical region fits on its boundary, at a density of order one qubit per Planck area.
- Local QFT is hugely redundant; there are only $\exp(A/4)$ states

Bekenstein bound

- A tighter bound results from a cleverer process:
- Slowly lower the matter system into a very large black hole, to minimize ΔA
- This decreases the energy of the system at infinity, by a redshift factor, before it is dropped in
- The mass added to the black hole is nonzero, however, because the system has finite size
- After some algebra (see hep-th/0203101), one finds

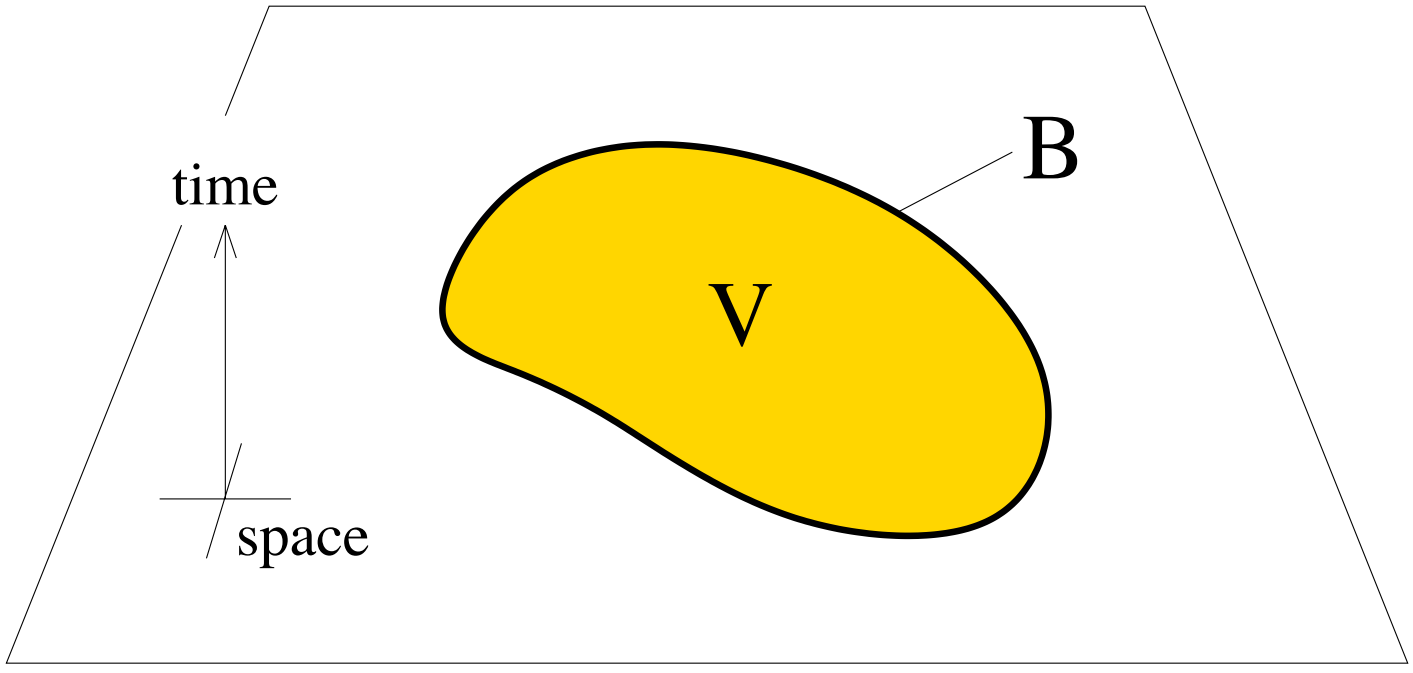
$$S \leq 2\pi MR/\hbar$$

- We will return to this bound later but focus for now on the holographic bound, $S \leq A/4G\hbar$

Limitations

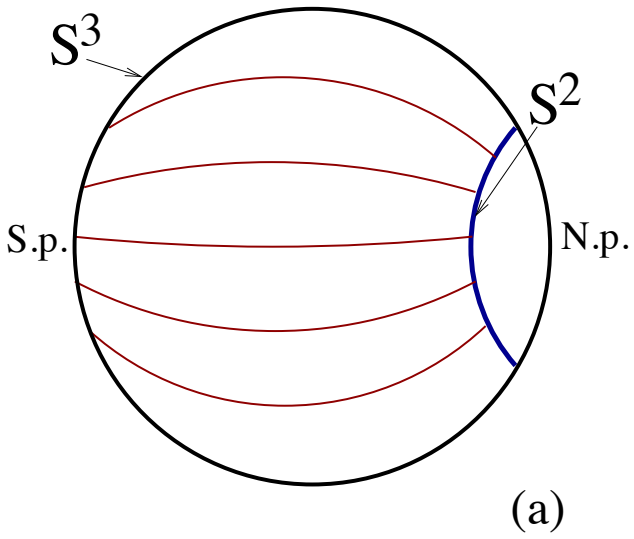
- The derivation of the above bounds from the GSL is somewhat handwaving
- E.g., what if some mass is shed before the black hole forms? It is difficult to treat gravitational collapse processes exactly except in overly idealized limits
- Moreover, the derivation implicitly assumes that we are dealing with a matter system that has weak self-gravity ($M \ll R$).
- Hence, it does not imply that $S \leq A/4$ for all matter systems.
- Will shortly see that indeed, the bound does not hold for some matter systems, if S is naively defined as the entropy "enclosed" by the surface

Spacelike Entropy Bound



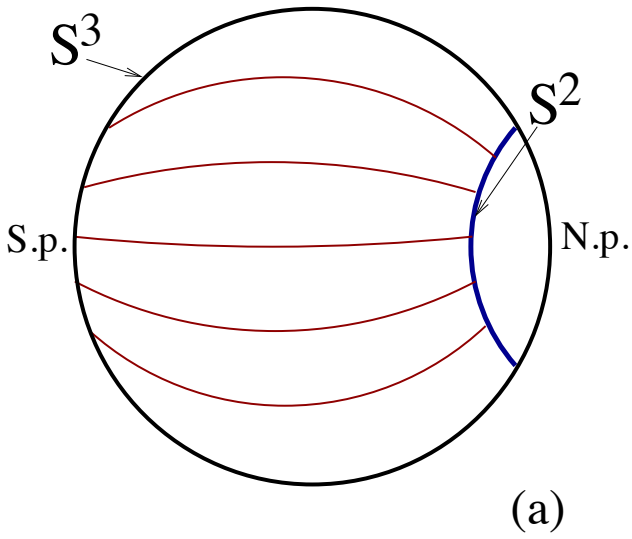
- SEB: $S[V] \leq A[B]/4$, for any 3-dimensional volume V
- I will now give four counterexamples to this bound

(1) Closed universe



- Let V be almost all of a closed three-dimensional space, except for a small region bounded by a tiny sphere B.
- The SEB should apply, $S[V] \leq A[B]/4$, but we can choose $S[V]$ arbitrarily large, and $A[B]$ arbitrarily small

(1) Closed universe



(a)

- Let V be almost all of a closed three-dimensional space, except for a small region bounded by a tiny sphere B.
- The SEB should apply, $S[V] \leq A[B]/4$, but we can choose $S[V]$ arbitrarily large, and $A[B]$ arbitrarily small
- (This type of "arbitrarily bad" violation can be found for any proposed entropy bound other than the CEB; all our counter-examples to the SEB will be of this type.)

(2) Flat FRW universe

$$ds^2 = -dt^2 + a(t)^2(dx^2 + dy^2 + dz^2)$$

- (E.g., with radiation, $a(t) \sim t^{1/2}$ and the physical entropy density is $\sigma \sim t^{-3/2}$)
- Consider a volume of physical radius R at fixed time t:

$$V \sim R^3 \; ; A[B] \sim R^2$$

$$S[V] \sim \sigma R^3$$

- In large volumes of space ($R \gtrsim \sigma^{-1}$), the SEB is violated
- $S/A \to \infty$ as $R \to \infty$

(3) Collapsing star

- Consider a collapsing star (idealize as spherical dust cloud)
- Its initial entropy S_0 can be arbitrarily large
- Let V be the volume occupied by the star just before it crunches to a singularity
- (This is well after it crosses its own Schwarzschild radius, so gravity is dominant and the surface of the star is trapped)
- From collapse solutions we know that $A[B] \to 0$ in this limit
- From the (ordinary) Second Law, we know that $S[V] \geq S_0$
- So we can arrange $S[V] > A[B]/4$ and, indeed, $S/A \to \infty$

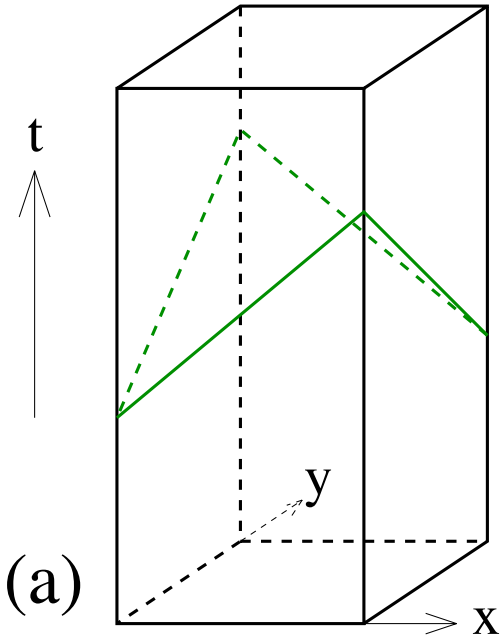
Give up?

- Perhaps there exists no general entropy bound of the form $S \leq A/4$, which holds for arbitrary regions?
- Instead try to characterize spatial regions that are in some sense sufficiently small, such that the SEB always holds for all of these “special” regions?
- E.g., interior of apparent horizon in FRW, interior of particle horizon, interior of Hubble horizon, etc.?

Give up?

- Perhaps there exists no general entropy bound of the form $S \leq A/4$, which holds for arbitrary regions?
- Instead try to characterize spatial regions that are in some sense sufficiently small, such that the SEB always holds for all of these "special" regions?
- E.g., interior of apparent horizon in FRW, interior of particle horizon, interior of Hubble horizon, etc.?
- Not well-defined beyond highly symmetric solutions
- Counterexamples have been found to all of these proposals, so
- Retreating from generality doesn't help!
- The notion of a "sufficiently small spatial region" conflicts with general covariance! (See next counterexample.)

(4) Nearly null boundaries

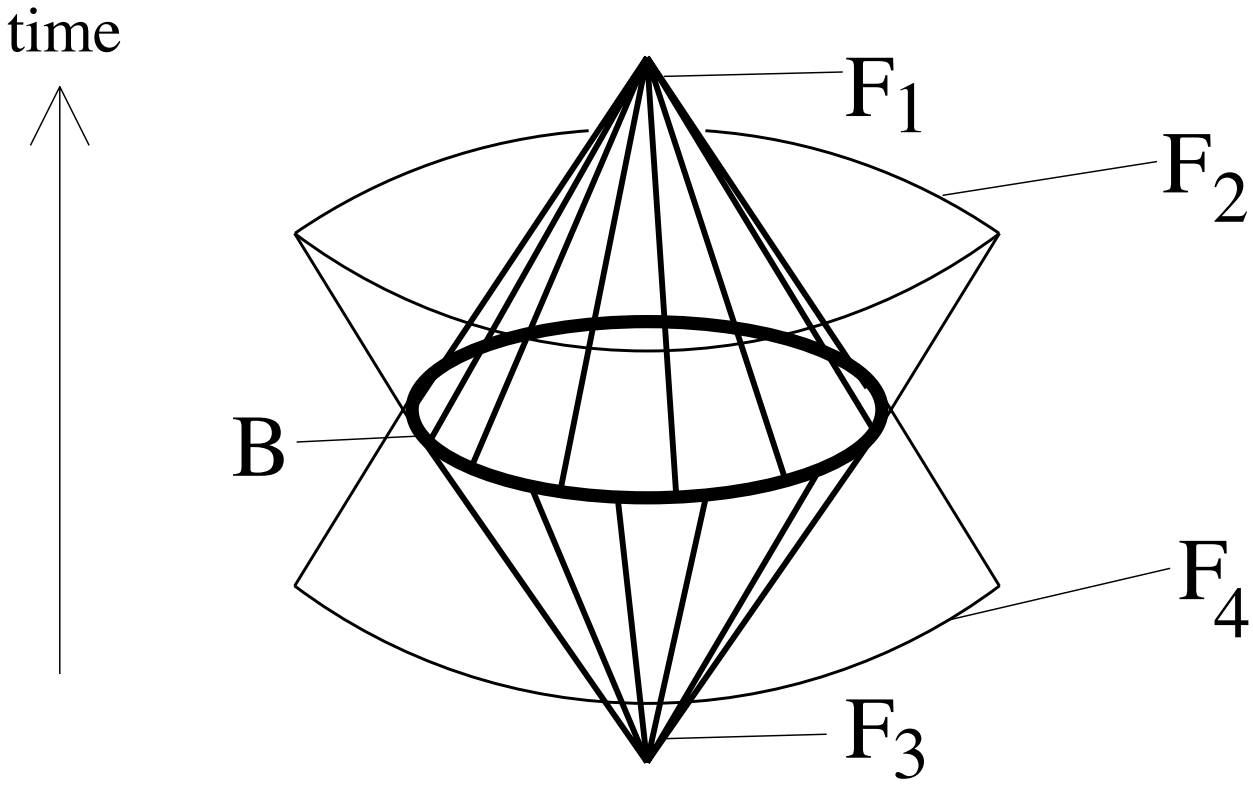


- Consider an ordinary matter system of constant entropy S
- Choose V such that B is Lorentz-contracted everywhere
- In the null limit $A[B] \to 0$, so again,
- the SEB is violated

What is the holographic principle?

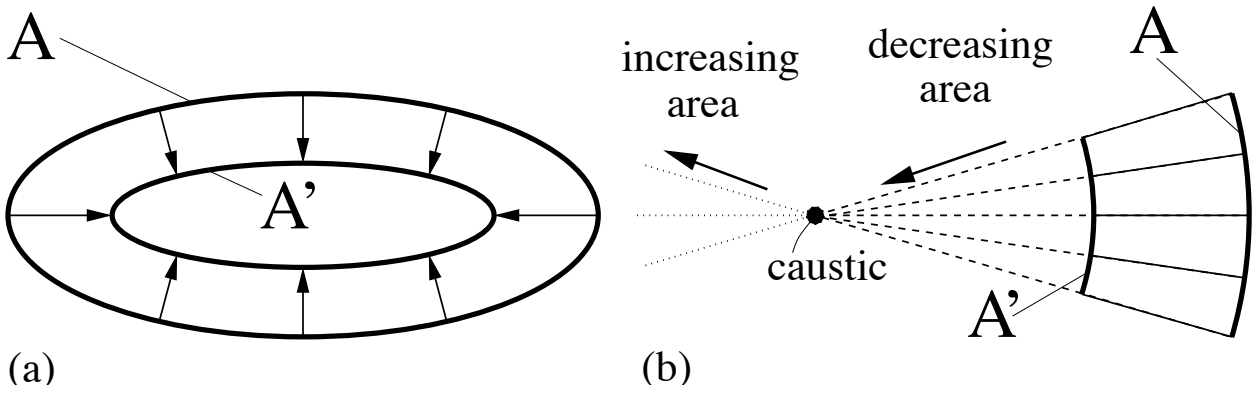
- In its most general form, the holographic principle is a relation between the geometry and information content of spacetime
- This relation manifests itself in the Covariant Entropy Bound
- The CEB reduces to the types of statements we were trying to make before using black holes, in regimes where they are valid
- But the CEB is more general; it remains valid for arbitrary surfaces in arbitrary spacetimes.
- I will first state the bound, and then revisit its definition more carefully.

Covariant Entropy Bound



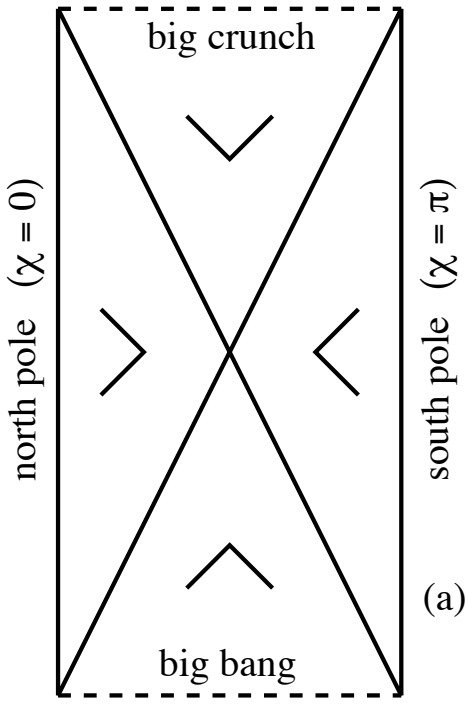
For any two-dimensional surface B of area A, one can construct lightlike hypersurfaces called light-sheets. The total matter entropy on any light-sheet is less than $A/4$ in Planck units: $S \leq A/4G\hbar$.

Covariant Entropy Bound



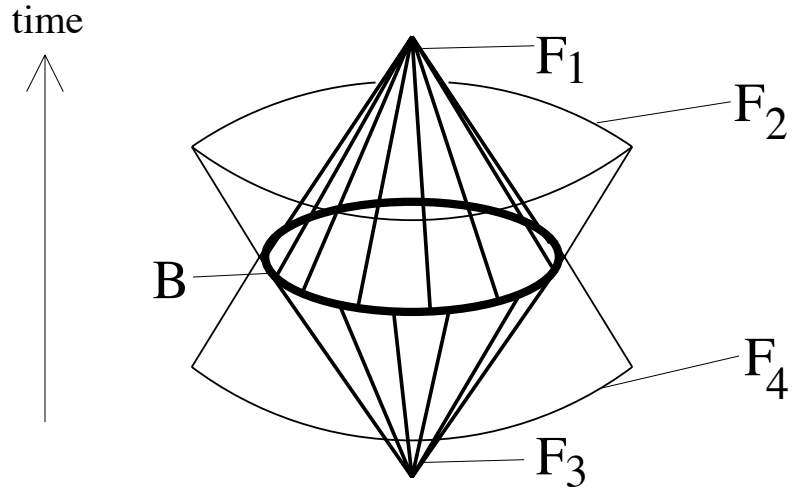
A light-sheet is generated by nonexpanding light-rays orthogonal to the initial surface *B*. Out of the 4 null directions orthogonal to *B*, at least 2 will have this property.

Covariant Entropy Bound



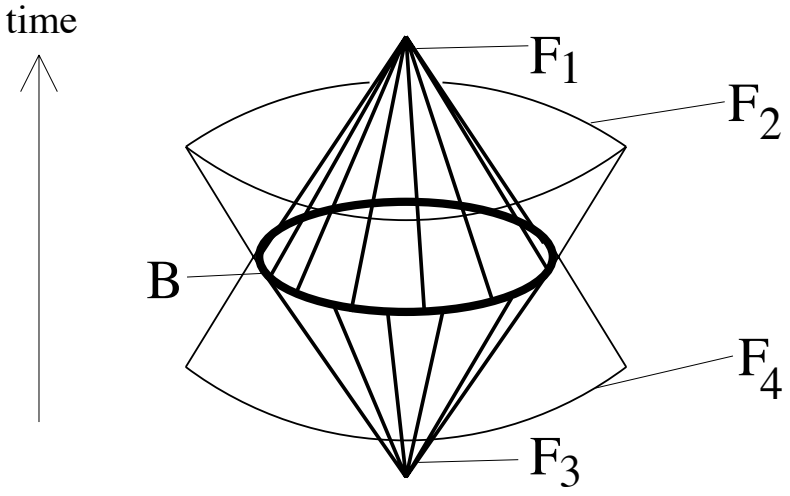
- If B is closed and "normal", the light-sheet directions will coincide with our intuitive notion of the "interior" of B
- But if B is trapped (anti-trapped) the light-sheets go only to the future (to the past).

What is the holographic principle?



- The CEB is completely general: it appears to hold for arbitary physically realistic matter systems and arbitrary surfaces in any spacetime that solves Einstein's equation
- The CEB can be checked case by case; no counterexamples are known
- But it seems like a conspiracy every time.
  The Origin of the CEB is not known!

What is the holographic principle?



- The CEB is completely general: it appears to hold for arbitary physically realistic matter systems and arbitrary surfaces in any spacetime that solves Einstein's equation
- The CEB can be checked case by case; no counterexamples are known
- But it seems like a conspiracy every time.
  The Origin of the CEB is not known!
- This is similar to the "accident" that inertial mass is equal to gravitational charge

What is the holographic principle?

- Solution: Elevate this to a principle and demand a theory in which it could be no other way!
- Equivalence Principle $\rightarrow$ General Relativity
- Holographic Principle $\rightarrow$

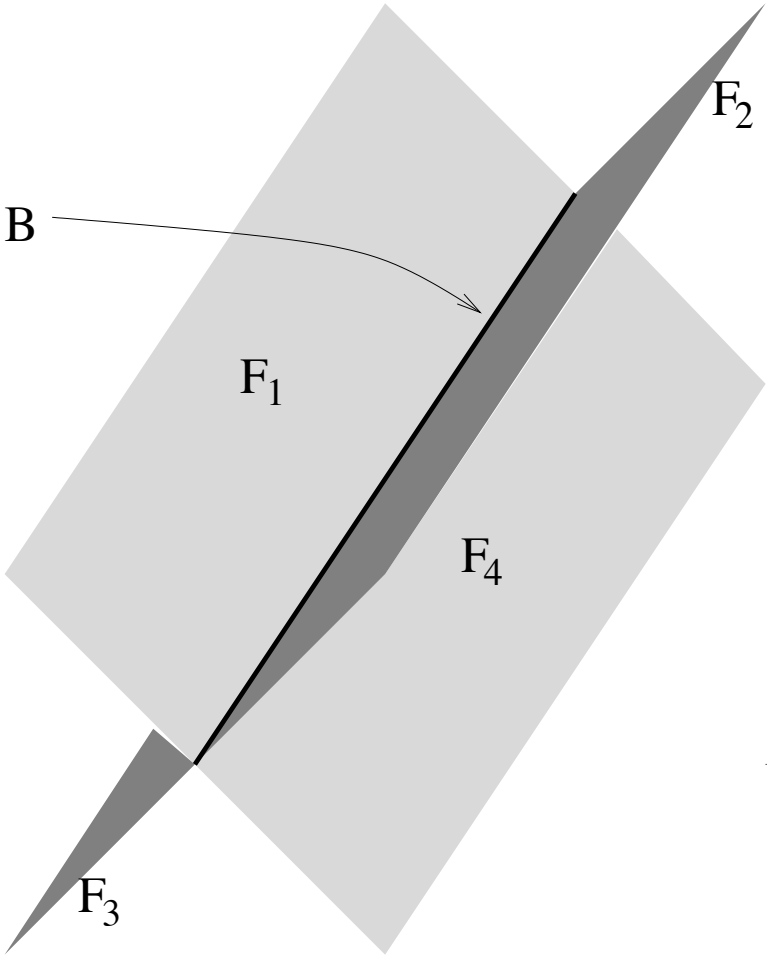
What is the holographic principle?

- Solution: Elevate this to a principle and demand a theory in which it could be no other way!
- Equivalence Principle $\rightarrow$ General Relativity
- Holographic Principle $\rightarrow$ Quantum Gravity
- Because the CEB involves both the quantum states of matter and the geometry of spacetimes, any theory that makes the CEB manifest must be a theory of everything, i.e., quantum gravity theory that also specifies the matter content. (Example of a candidate: string theory.)
- Someone please explain how the CEB and locality fit together!

What is entropy?

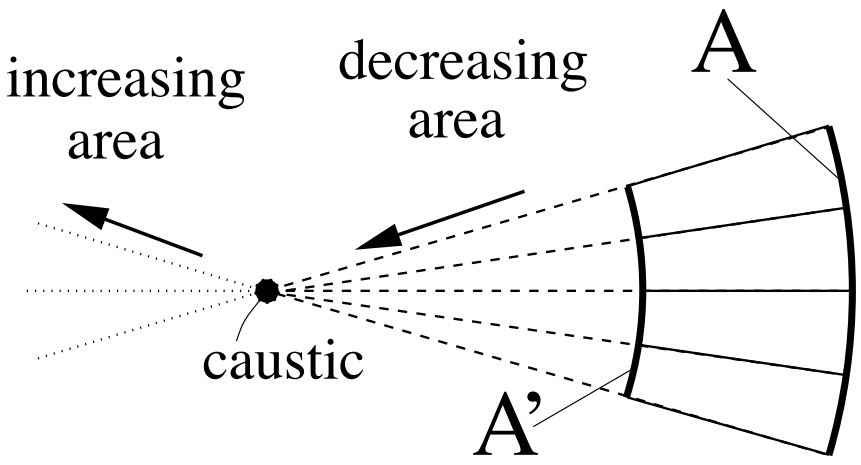
- Entropy is the (log of the) number of independent quantum states compatible with some set of macroscopic data (volume, energy, pressure, temperature, etc.)
- The relevant boundary condition for our purposes is that the matter system should fit on a light-sheet of a surface of area A (roughly, that it fits within a sphere of that area)

Null geodesic congruences



- Any 2D spatial surface B bounds four (2+1D) null hypersurfaces
- Each is generated by a congruence of null geodesics $\perp B$

Expansion of a null congruence



- $\theta \equiv \widehat{\nabla_a k^a}$, where k^a is the affine tangent vector field to the congruence (see Wald)
- In terms of an infinitesimal area element $\mathcal{A}$ spanned by nearby light-rays,

$$\theta = \frac{d\mathcal{A}/d\lambda}{\mathcal{A}}$$

- $\theta < 0 \leftrightarrow$ contraction;
- $\theta \to -\infty \leftrightarrow$ caustic ("focal point", "conjugate point to B")

Light-sheets

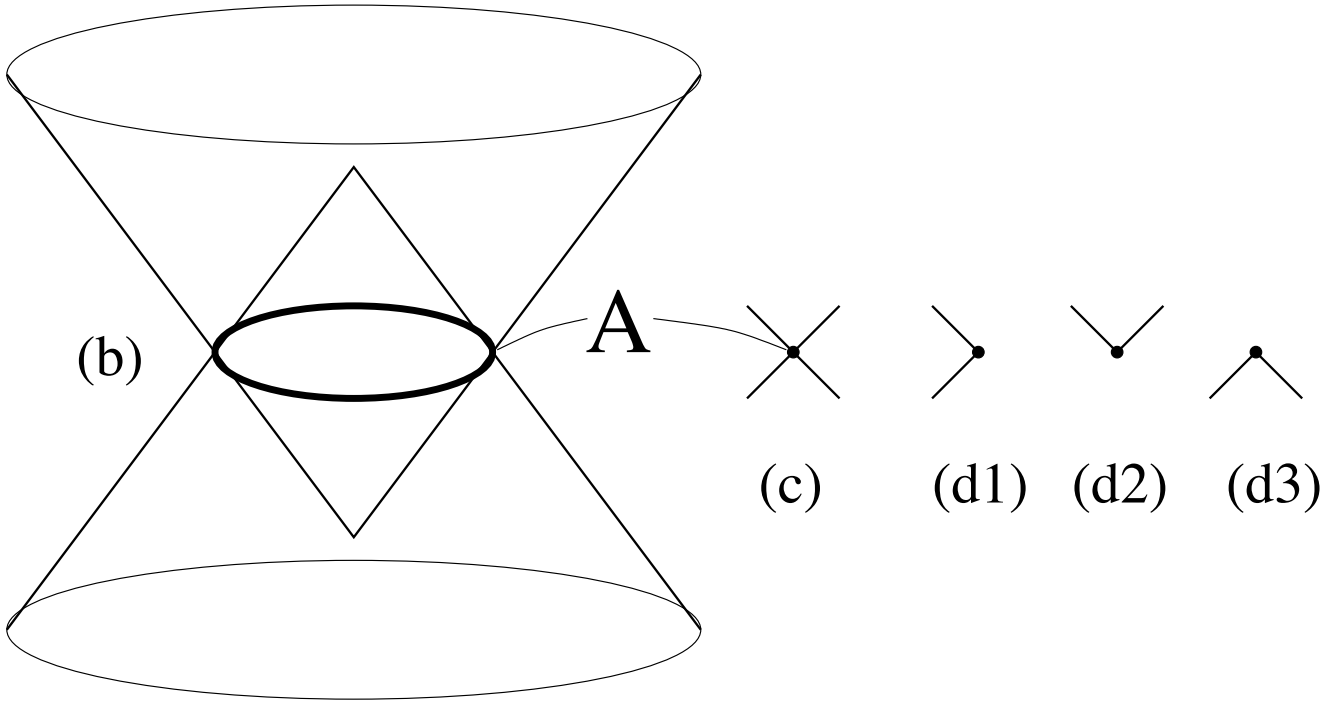
- A light-sheet of B is a null hypersurface $L \perp B$ with boundary B and $\theta \leq 0$ everywhere on L
- Note: Assuming the null energy condition ($T_{ab}k^a k^b \geq 0$) holds,
  - there are at least two null directions away from B for which $\theta \leq 0$ initially
  - $d\theta/d\lambda \leq -\theta^2/2$, so a caustic is reached in finite affine time
- If we think of generating L by following null geodesics away from B, we must stop as soon as θ becomes positive
- In particular, we must stop at any caustic

Covariant Entropy Bound

The total matter entropy on any light-sheet of B is bounded by the area of B:

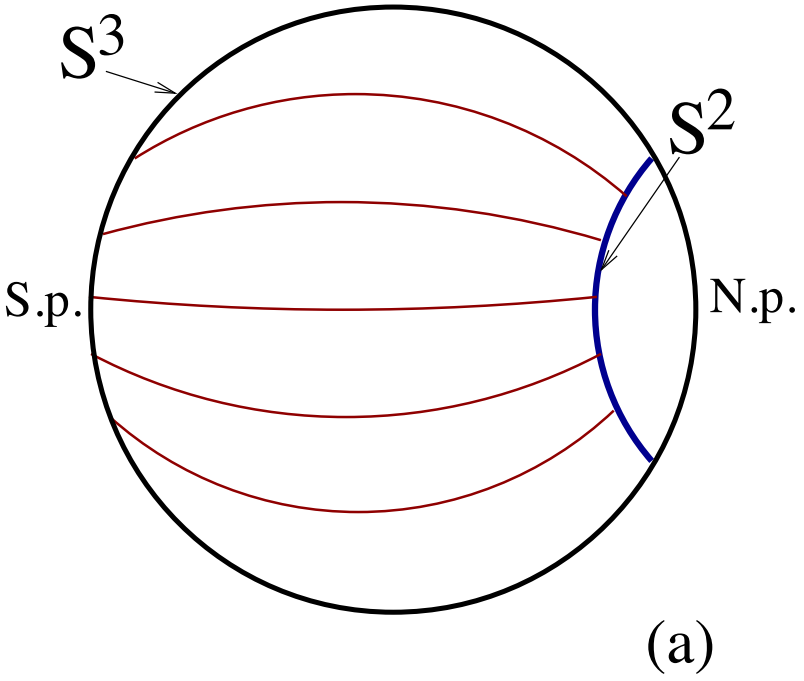
$$S[L(B)] \leq A[B]/4G\hbar$$

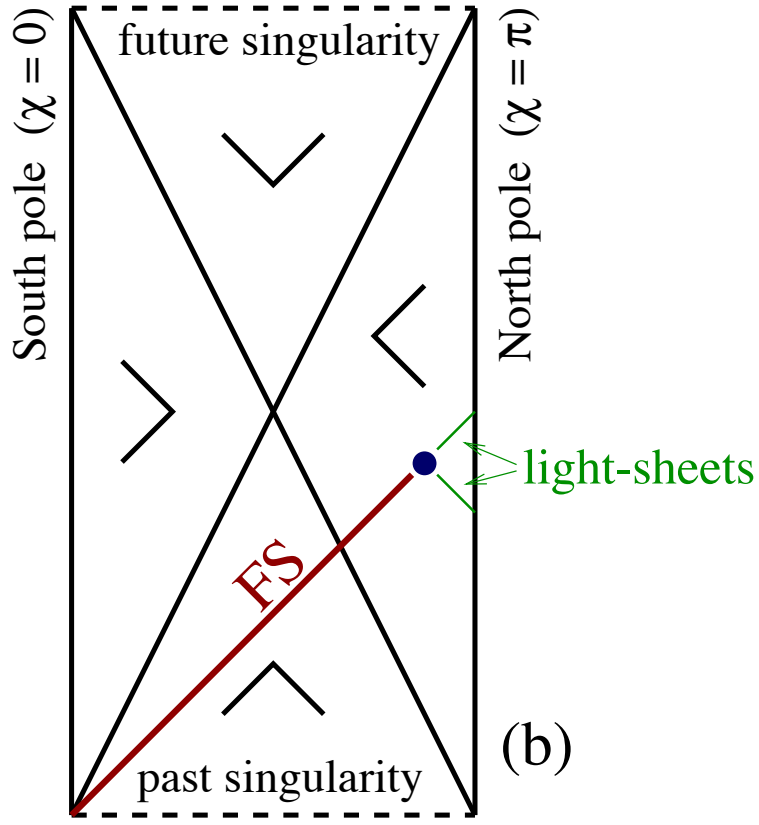
Allowed light-sheet directions



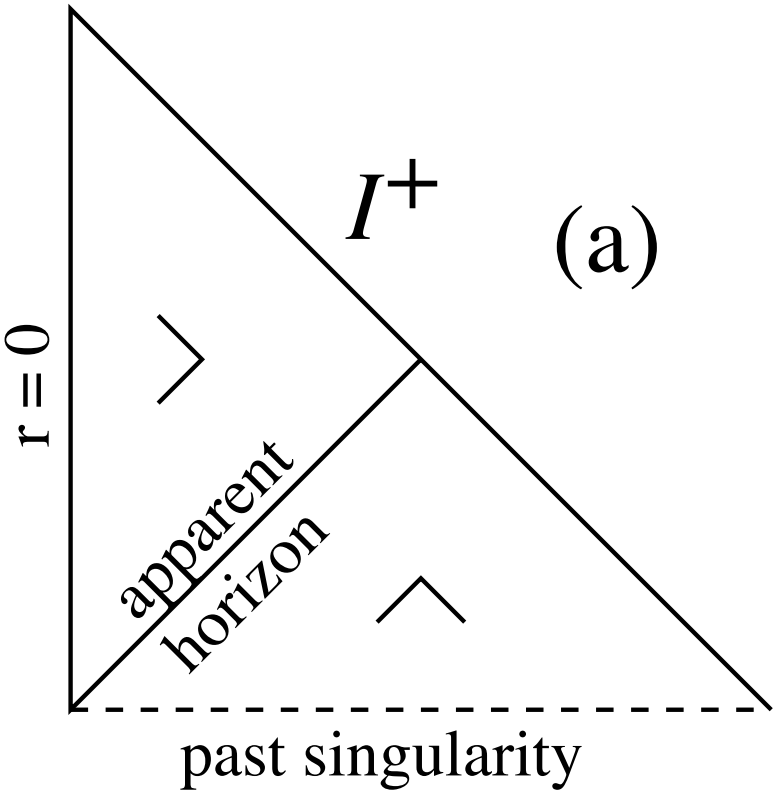
- Often we consider spherically symmetric spacetimes
- In a Penrose diagram, a sphere is represented by a point
- The allowed light-sheet directions can be represented by wedges
- This notation will be useful as we analyze examples

(1) Closed universe



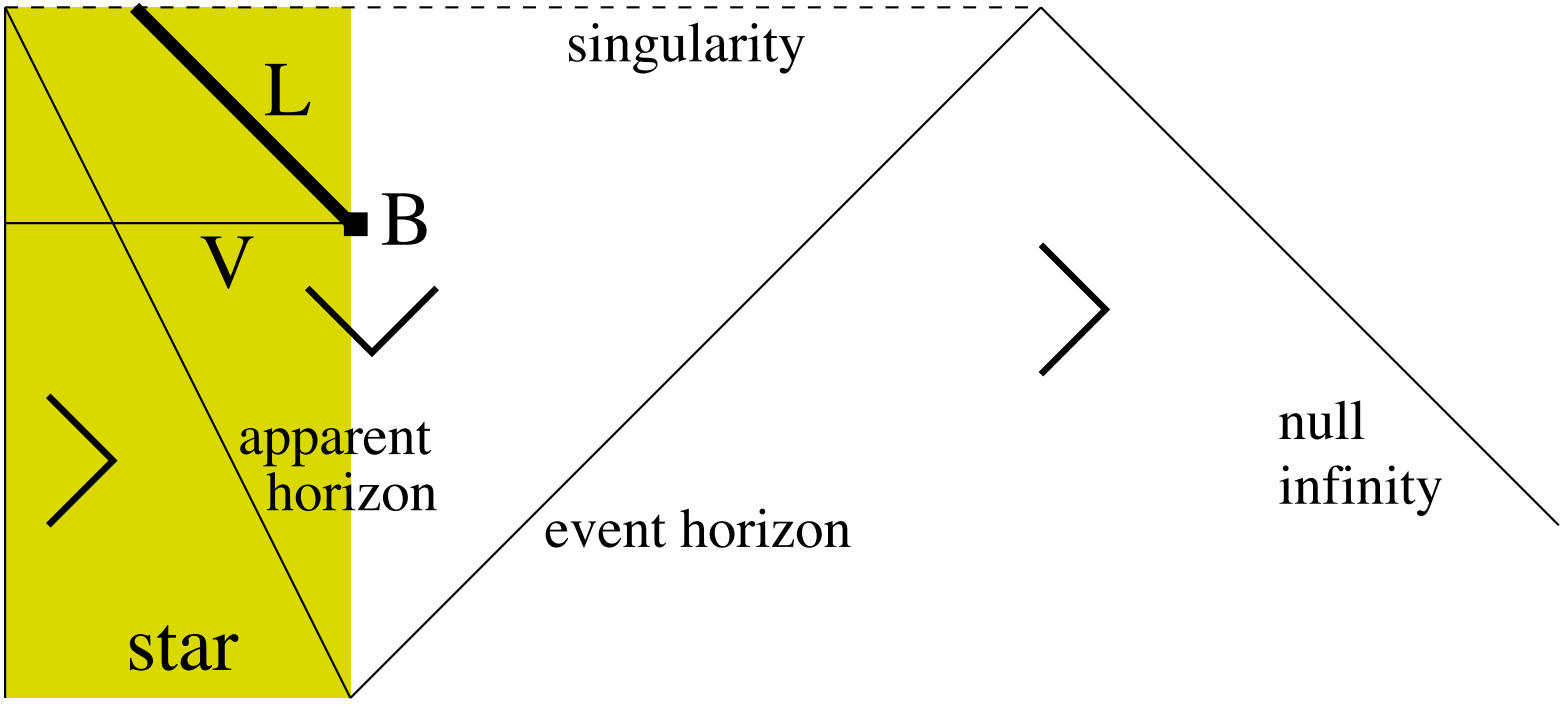


(2) Flat FRW universe



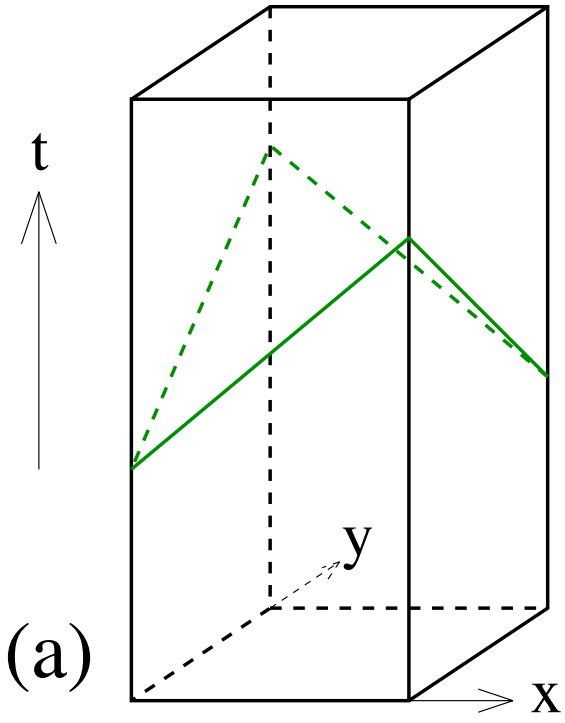
- Sufficiently large spheres at fixed time t are anti-trapped
- Only past-directed light-sheets are allowed
- The entropy on these light-sheets grows only like R^2

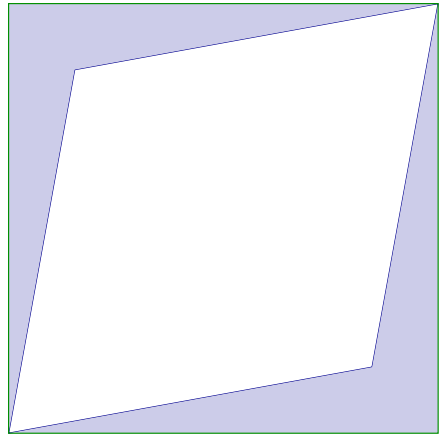
(3) Collapsing star



- At late times the surface of the star is trapped
- Only future-directed light-sheets exist
- They do not contain all of the star

(4) Nearly null boundaries

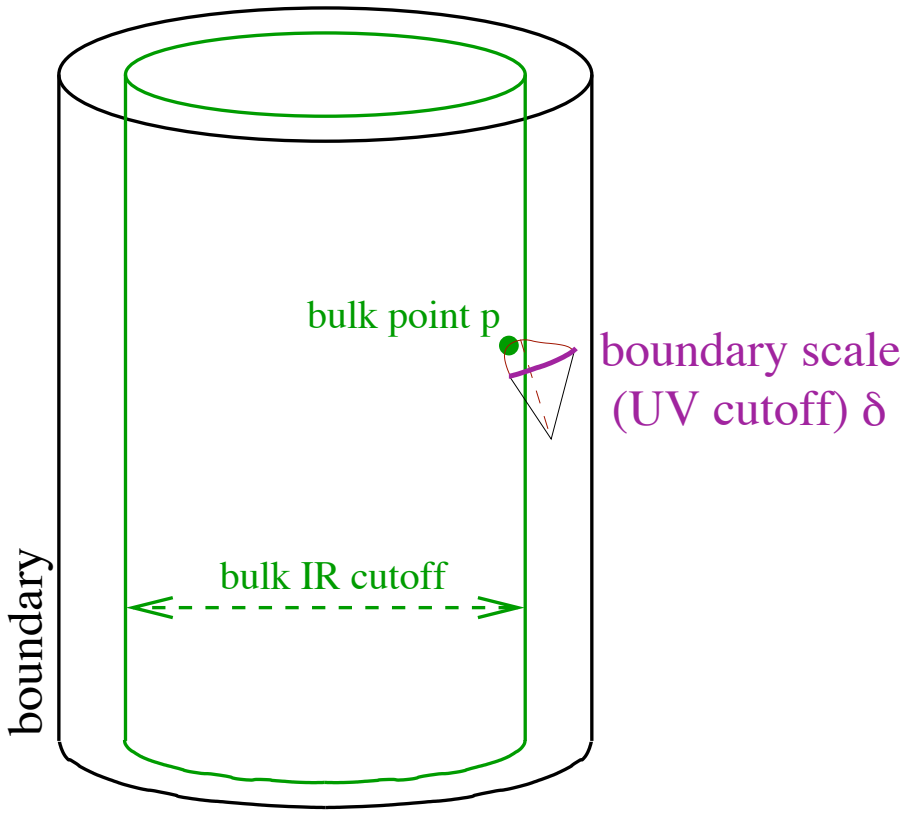




(b)

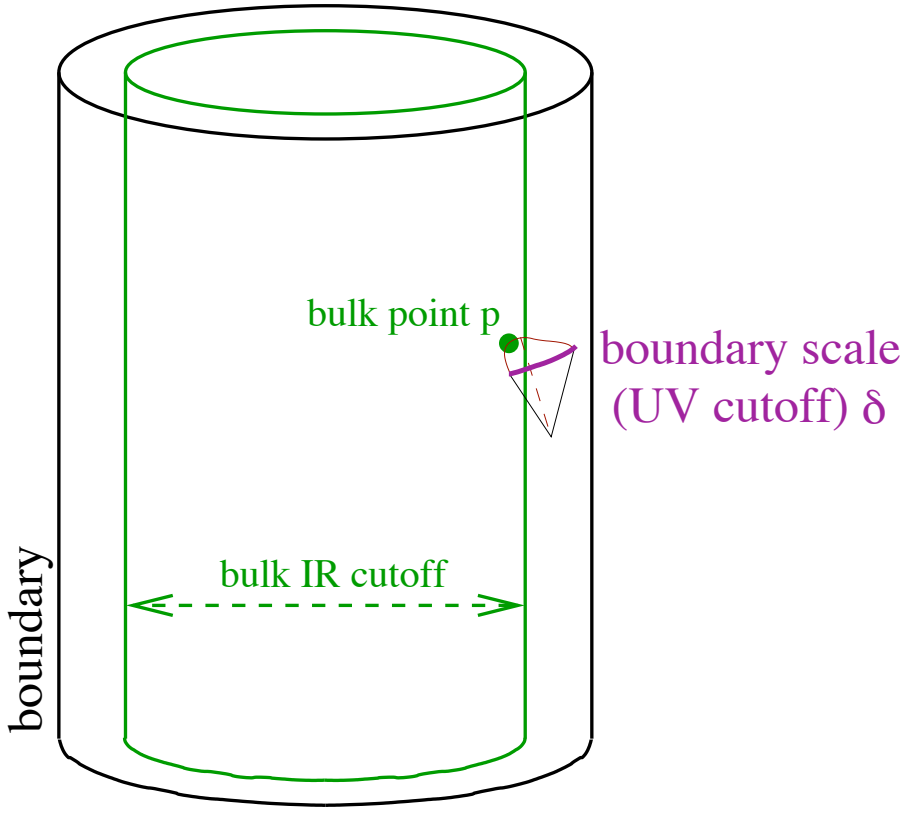
- The null direction orthogonal to B is not towards the center of the system
- The light-sheets miss most of the system, so $S \to 0$ as $A[B] \to 0$

AdS/CFT



- First complete, nonperturbative quantum theory of gravity
- An asymptotically AdS spacetime is described by a conformal field theory on the boundary

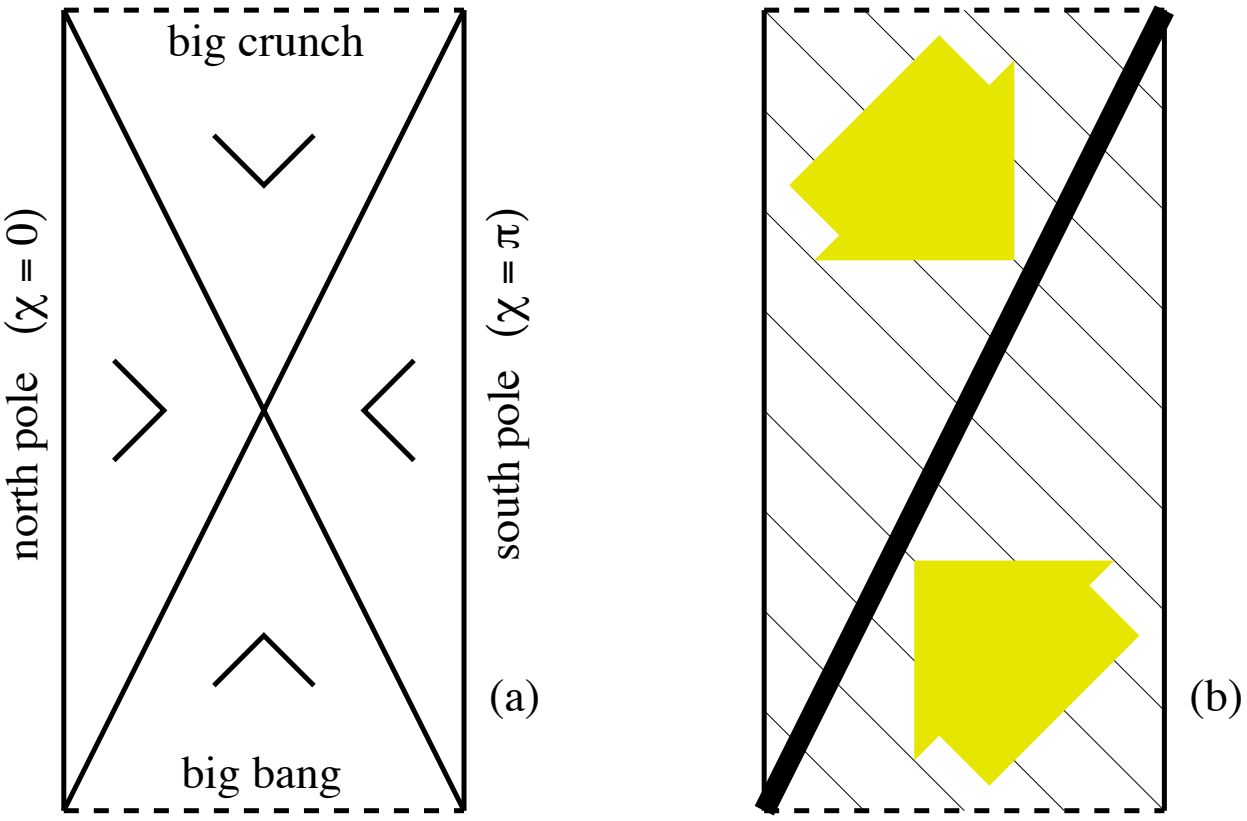
AdS/CFT



- There exists a cutoff version of this correspondence: A CFT with UV cutoff δ describes AdS out to a sphere of area A
- The relation $A(\delta)$ is such that the log of the dimension of the CFT Hilbert space is of order A
- The holographic principle is manifest!

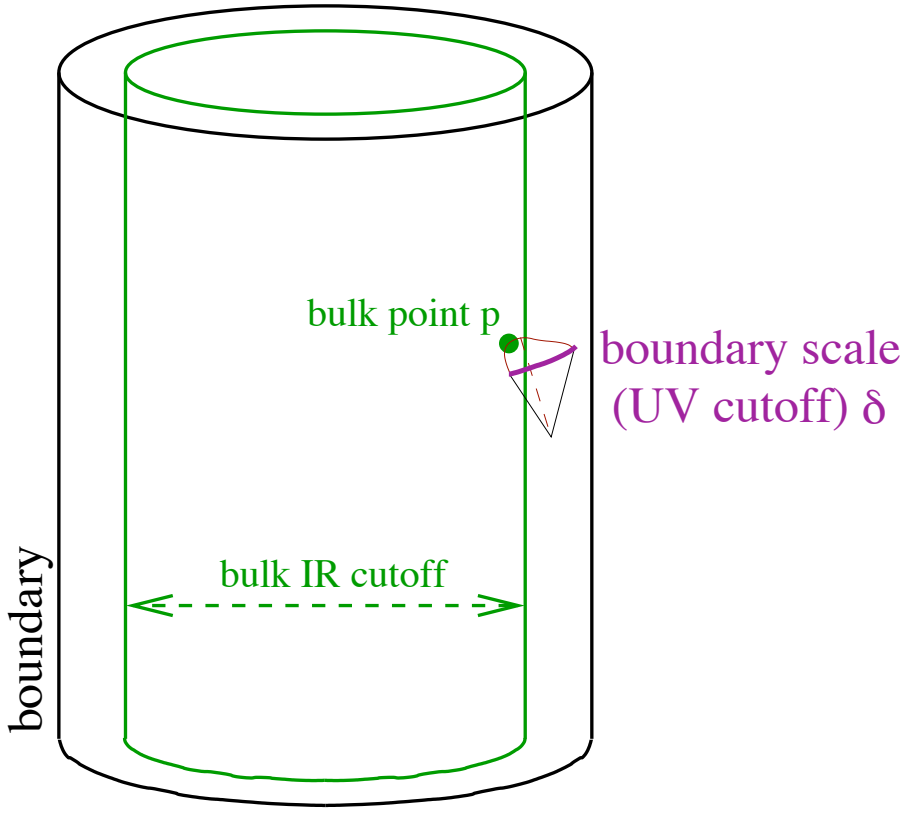
Holographic screens in general spacetimes

The world is a hologram. . .



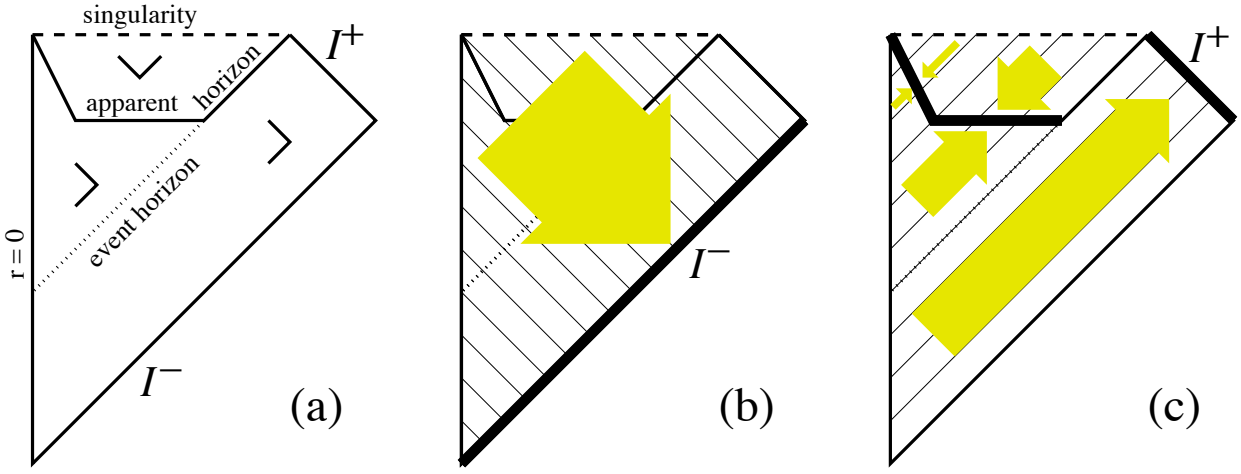
Holographic screens in general spacetimes

. . . but we don't know how the encoding works.



- AdS is very special; the dual theory is a unitary field theory sharing the same time variable
- This is related to a property of the holographic screen in AdS

Holographic screens in general spacetimes



- In general spacetimes, it would seem that the number of degrees of freedom has to change as a function of the time parameter along the screen
- The screen is not unique
- The screen can even be spacelike and it need not be connected

Ambiguity of the Vacuum in Curved Spacetime

In flat spacetime, the vacuum is the state from which no particles can be removed:

$$a_i|0\rangle = 0 \ ,$$

where a_i are the annihilation operators defined by decomposing the quantum field into positive and negative frequency modes:

$$\phi = \sum_i f_i a_i + f_i^* a_i^\dagger \ ,$$

and f_i are a complete set of modes of positive frequency with respect to the usual Minkowski time coordinate ($f_i \sim e^{ikx - i\omega t}$), obeying the equation of motion.

Ambiguity of the Vacuum in Curved Spacetime

In curved spacetime, there is no invariant notion of positive frequency. Decompositions into modes with positive and negative norm can be defined:

$$\frac{i}{2}\int_{\Sigma}(f_i\partial_a f_j^* - f_j^*\partial_a f_i)d\Sigma^a = \delta_{ij}\ ,$$

but the subspace of positive norm modes is not unique. If the spacetime contains asymptotic regions that are flat, then it is natural to choose a basis of modes which are positive or negative frequency with respect to the standard Minkowski timelike Killing vector field. If such regions are present at early (1) and late (3) times but spacetime is curved at intermediate times (2), then it is not guaranteed that the preferred vacuum state defined via (1) will agree with that defined from (3).

Particle Creation by Black Holes

Consider a black hole formed from collapse. At early times (long before the black hole forms), decompose the field with respect to Minkowski time

$$\phi = \sum_i f_i a_i + f_i^* a_i^\dagger \ ,$$

Note that the modes f_i are defined at all times through the equation of motion, $g^{ab}\nabla_a \nabla_b \phi = 0$.

Particle Creation by Black Holes

The black hole horizon and future null infinity also form a Cauchy surface. The distinction between them defines two sets of modes which together are complete:

$$\phi = \sum_i p_i b_i + p_i^* b_i^\dagger + q_i c_i + q_i^* c_i^\dagger \ ,$$

where the modes p_i are positive frequency at null infinity and purely outgoing, i.e., they vanish on the event horizon. Again, this decomposition is defined at all times through the equation of motion.

Bogoliubov Transformations

The modes and operators defined in these two ways are not independent. Annihilation operators of one type can be expressed as linear combinations of creation and annihilation operators of the other decomposition:

$$p_i = \sum_j \alpha_{ij} f_j + \beta_{ij} f_j^* \ ;$$

$$b_i = \sum_j \alpha_{ij}^* a_j + \beta_{ij}^* a_j^* \ .$$

Particle Creation by Black Holes

The initial state is defined as the infalling vacuum:

$$a_i|0\rangle = 0$$

for all i. Because of the mixing between creation and annihilation operators in the Bogoliubov transforms, however,

$$b_i|0\rangle \neq 0 \ !$$

The expected occupation number of the i-th mode at future null infinity will be nonzero:

$$\langle 0|b_i^\dagger b_i|0\rangle = \sum_j |\beta_{ij}|^2 \ .$$

Particle Creation by Black Holes

There is a technical complication: if backreaction is neglected, the black hole emits an infinite total amount of radiation in every mode:

$$\int_0^\infty d\omega' |\beta_{\omega\omega'}|^2 = \infty \ .$$

Handle this by defining wavepackets of finite temporal extent and characteristic frequency, p_{nj}.

Thermal spectrum

Let Γ_{nj} be the absorption probability for a wavepacket. One finds that

$$\int_0^\infty d\omega' |\beta_{jn\omega'}|^2 = \frac{\Gamma_{nj}}{e^{\omega/T} - 1} ,$$

with

$$T = \frac{\kappa}{2\pi} .$$

Thus, as far as the expected occupation number of each mode is concerned, the black hole emits as a thermal body.

One can show more generally that the state of the radiation is a thermal density matrix:

$$\rho = e^{-H/T} .$$

Information Loss

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.

Why not?

- Unitarity is a fundamental principle in quantum mechanics; black holes appear in the path integral that determines the S-matrix

Why not?

- Unitarity is a fundamental principle in quantum mechanics; black holes appear in the path integral that determines the S-matrix
- The apparent validity of nontrivial entropy bounds suggests that black holes do have a large enough Hilbert space to store the information—only to lose it when they evaporate?

Why not?

Black-hole physics seems to provide at least two ways in which the second law of thermodynamics may be transcended or violated:

a) Let an observer drop or lower a package of entropy into a black hole; the entropy of the exterior world decreases. Furthermore, from an exterior observer's point of view a black hole in equilibrium has only three degrees of freedom: mass, charge and angular momentum (1). Thus, once the black hole has settled down to equilibrium, there is no way for the observer to determine its interior entropy. Therefore, he cannot exclude the possibility that the total entropy of the universe may have decreased in the process. It is in this sense that the second law appears to be transcended (2).

Why not?

- Scattering in AdS is described by a manifestly unitary CFT

Where's the mistake?

Hawking's calculation assumes only the equivalence principle. It uses Quantum Field Theory on a curved spacetime background, in a regime of low curvature. What could go wrong?

Where's the mistake?

Hawking's calculation assumes only the equivalence principle. It uses Quantum Field Theory on a curved spacetime background, in a regime of low curvature. What could go wrong?

Planck scale remnants? But they would appear to lead to instabilities and they conflict with holographic entropy bounds.

Where's the mistake?

Hawking's calculation assumes only the equivalence principle. It uses Quantum Field Theory on a curved spacetime background, in a regime of low curvature. What could go wrong?

Planck scale remnants? But they would appear to lead to instabilities and they conflict with holographic entropy bounds.

The evaporation process takes a time $O(r_S^3)$, during which $O(r_S^2)$ quanta are emitted from a region of area $O(r_S^2)$. Hawking's calculation uses local QFT in a situation where the limitations imposed by the holographic principle are actually encountered.

Is Unitarity Consistent with the Equivalence Principle?

Just as unitarity lies at the core of QM, the equivalence principle lies at the heart of GR. On scales $\ll r_S$, for observers entering after the scrambling time, the horizon should be locally indistinguishable from Minkowski space. The horizon is not a special place.

Is Unitarity Consistent with the Equivalence Principle?

Just as unitarity lies at the core of QM, the equivalence principle lies at the heart of GR. On scales $\ll r_S$, for observers entering after the scrambling time, the horizon should be locally indistinguishable from Minkowski space. The horizon is not a special place.

Independently of the details of Hawking's calculation, it is not clear whether these two principles are mutually consistent.

Is Unitarity Consistent with the Equivalence Principle?

Just as unitarity lies at the core of QM, the equivalence principle lies at the heart of GR. On scales $\ll r_S$, for observers entering after the scrambling time, the horizon should be locally indistinguishable from Minkowski space. The horizon is not a special place.

Independently of the details of Hawking's calculation, it is not clear whether these two principles are mutually consistent.

Black Hole Complementarity is an attempt to reconcile the apparent conflict between the EP and Unitarity, by arguing that no contradictions can be operationally verified by any observer in the spacetime.

The Principles of Black Hole Complementarity

1. The Hawking radiation is in a pure state determined by a unitary S-matrix.
2. Effective Field Theory is valid outside the stretched horizon.
3. A black hole has a Hilbert space of dimension $\exp(A/4)$.
4. An infalling observer encounters nothing special on crossing the horizon.

Are they compatible?

I will first give an intuitive example, then a precise example of a contradiction between unitarity and the equivalence principle.

Membrane vs. Free Fall

According to the outside observer, the black hole is an object like any other. It behaves as a warm membrane located at the stretched horizon (a timelike hypersurface located l_P away from the mathematical event horizon). The membrane has specific electrical and mechanical properties that can be measured by lowering probes. (This is closely related to AdS/Hydrodynamics duality.) The membrane absorbs and thermalizes infalling objects, and later returns them as Hawking radiation. Objects do not cross the horizon. The infalling observer must see a violation of the equivalence principle.

Membrane vs. Free Fall

According to the infalling observer, matter and its information content can be carried across the horizon. By causality, this information cannot return to the outside observer. The outside observer should see information loss. This is consistent with Hawking's result that the out-state is a thermal density matrix.

Membrane vs. Free Fall

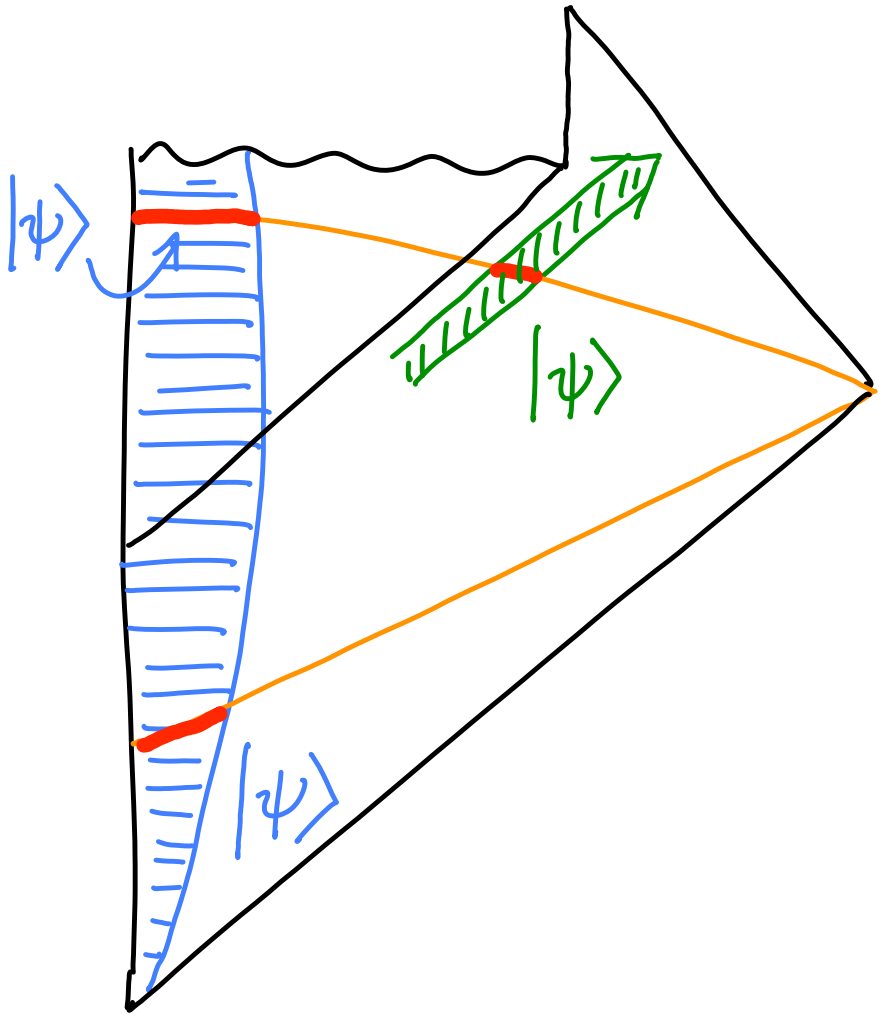
But by the time the infalling observer has been absorbed, $r_S \log r_S$, it is too late for the outside observer to send a signal to him. And by the time the infalling observer has crossed the horizon harmlessly, it is too late to report this fact to the outside observer!

Membrane vs. Free Fall

But by the time the infalling observer has been absorbed, $r_S \log r_S$, it is too late for the outside observer to send a signal to him. And by the time the infalling observer has crossed the horizon harmlessly, it is too late to report this fact to the outside observer!

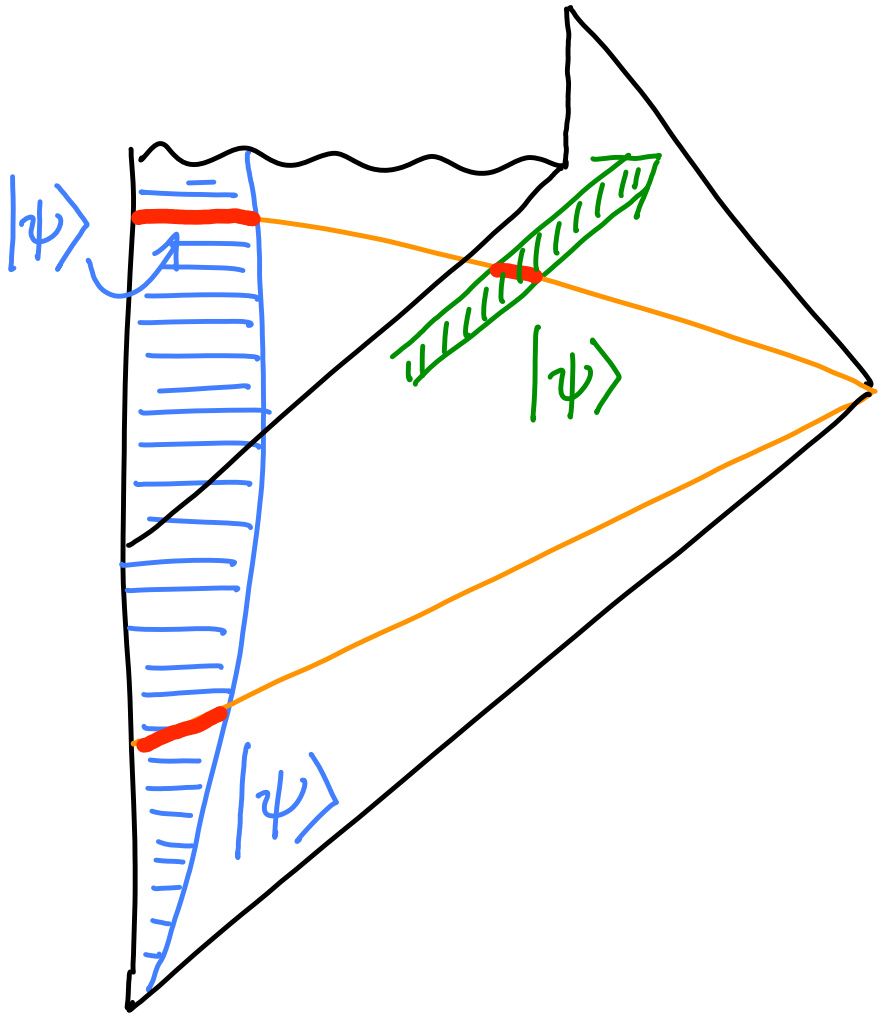
A sharper contradiction is obtained by considering evolution of the quantum state, while assuming that both principles hold.

Xeroxing paradox



A black hole forms by the collapse of an object in the pure state $|\psi\rangle$. Adopt the Heisenberg picture.

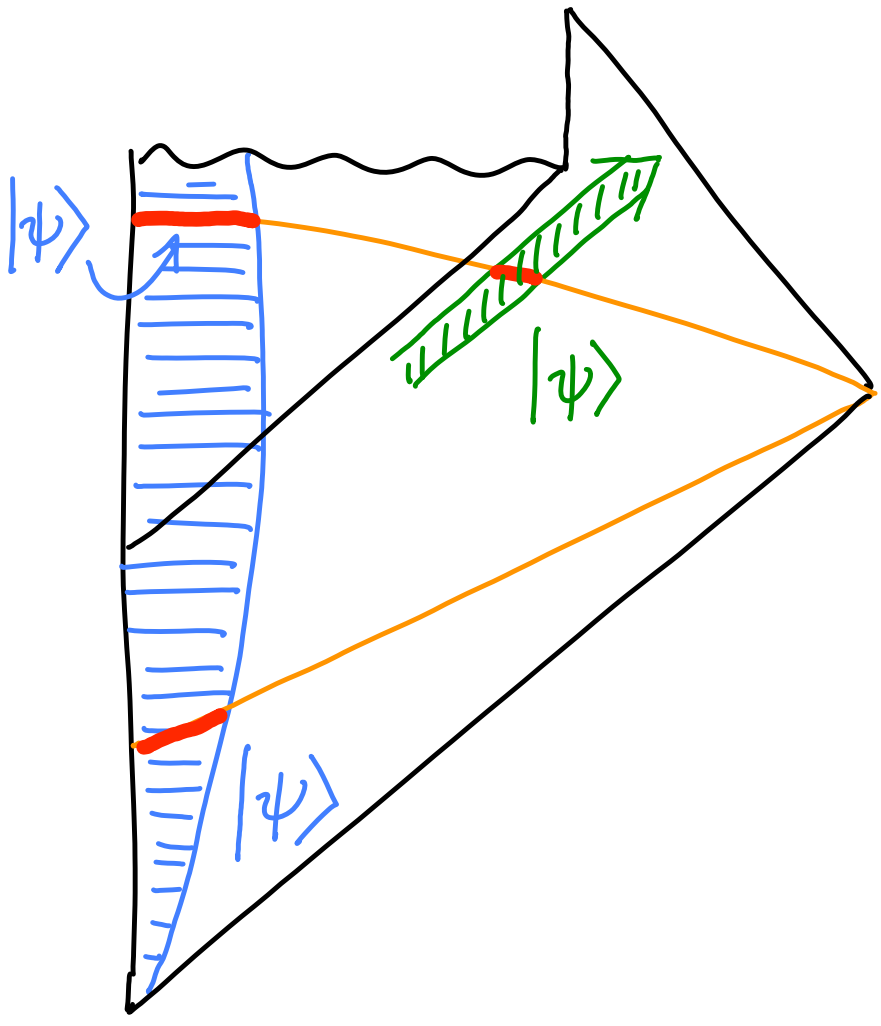
Xeroxing paradox



A black hole forms by the collapse of an object in the pure state $|\psi\rangle$. Adopt the Heisenberg picture.

After the black hole has evaporated, the Hawking cloud is in the quantum state $|\psi\rangle$, by unitarity.

Xeroxing paradox

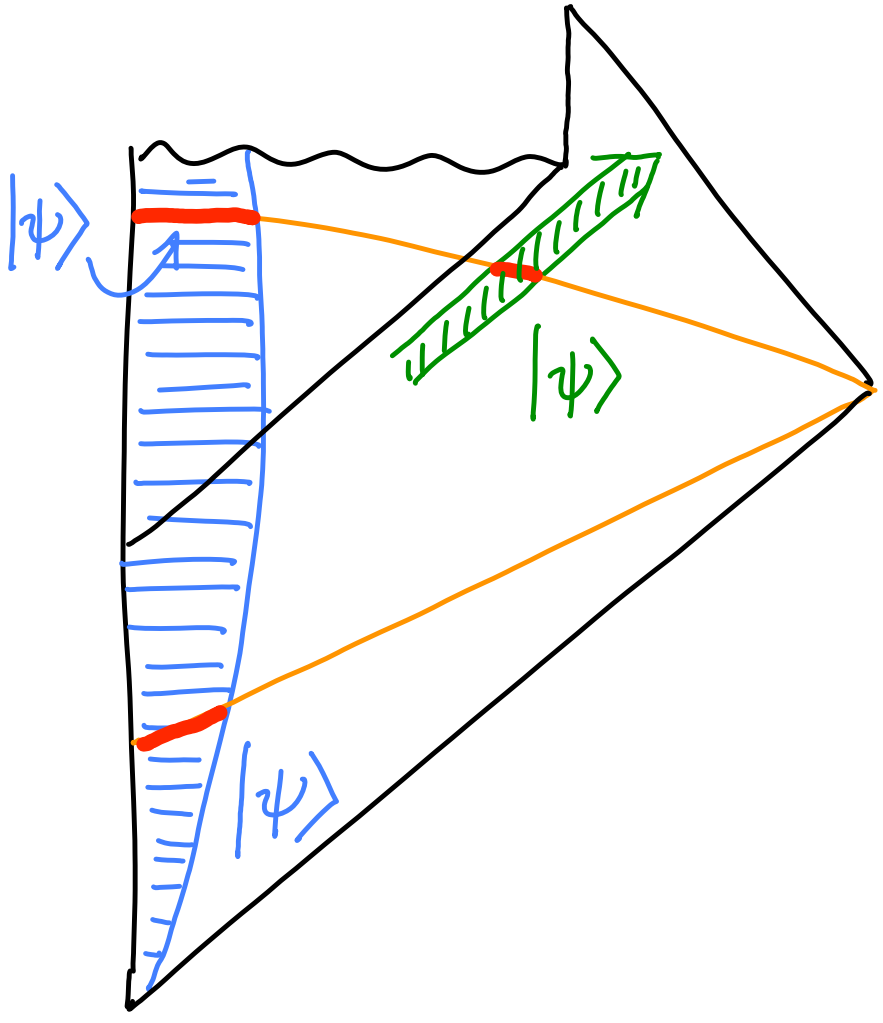


A black hole forms by the collapse of an object in the pure state $|\psi\rangle$. Adopt the Heisenberg picture.

After the black hole has evaporated, the Hawking cloud is in the quantum state $|\psi\rangle$, by unitarity.

At the same time inside the black hole, the object is still collapsing and remains in the quantum state $|\psi\rangle$, by the equivalence principle.

Xeroxing paradox

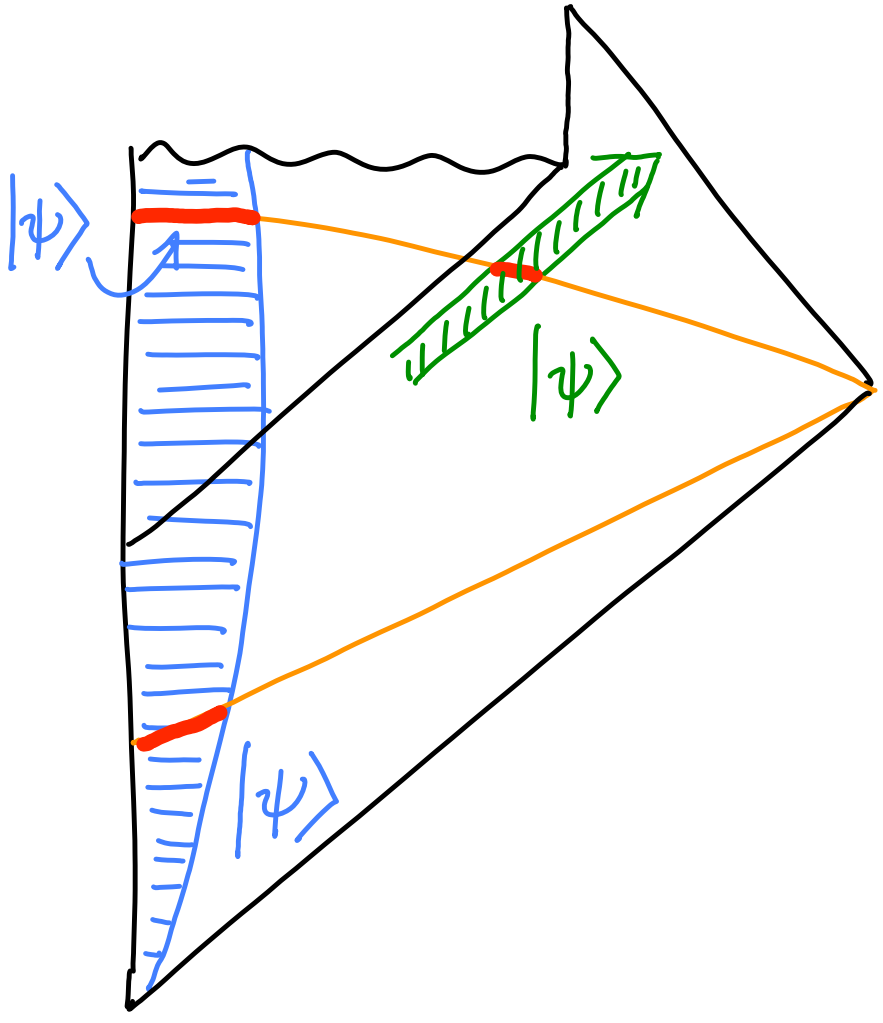


The black hole acts as a quantum xeroxing machine:

$$|\psi\rangle \rightarrow |\psi\rangle \otimes |\psi\rangle$$

This violates the linearity of quantum mechanics.

Xeroxing paradox



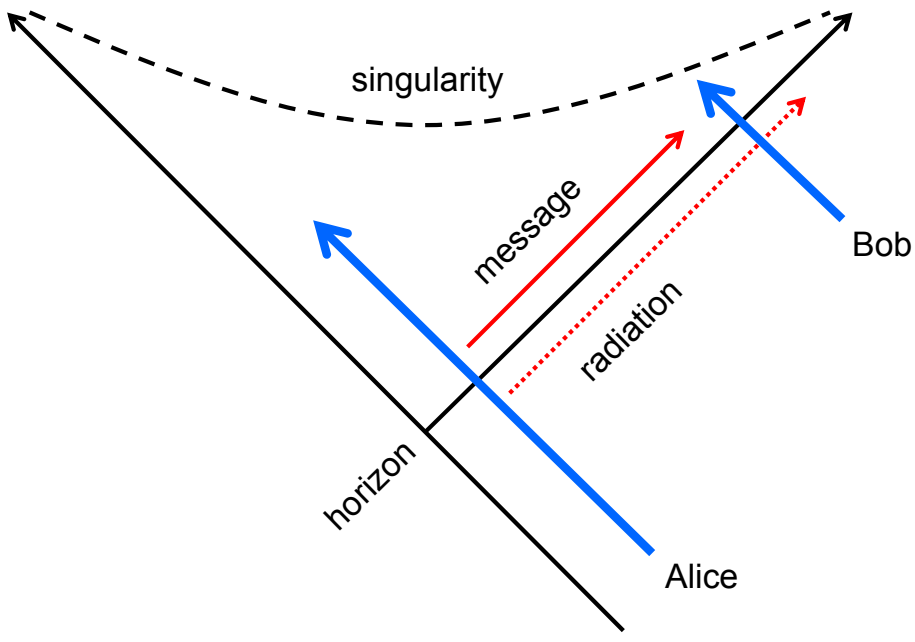
The black hole acts as a quantum xeroxing machine:

$$|\psi\rangle \rightarrow |\psi\rangle \otimes |\psi\rangle$$

This violates the linearity of quantum mechanics.

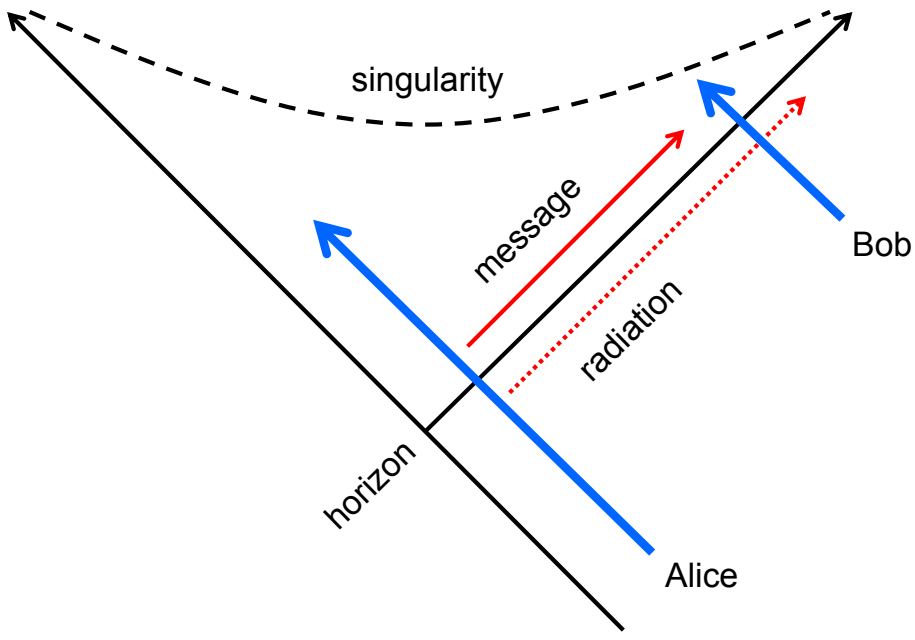
However, no observer can see both copies!

Strategy 1 (Susskind Thorlacius 1993)



Alice falls in with the star, sends a bit to Bob right after crossing the horizon. Bob waits until he can recover the bit in the Hawking radiation and then jumps in to receive Alice's message. But: Bob has to wait for $O(R^3)$ (Page 1993) until the first bit of information comes out. Alice would need energy $O(\exp R^2)$ to send a signal that reaches Bob before he hits the singularity. Bob fails to see both copies, by a lot.

Strategy 2 (Hayden Preskill 2007)



Bob already controls more than half of the Hawking radiation, then Alice throws in a bit. Since Bob's radiation is maximally entangled with the black hole, Alice's bit can be recovered as soon as the black hole has completely thermalized it. Speculative arguments suggest that the scrambling time can be as fast as $O(R \log R)$. In this case, Bob fails to see both copies, but only barely.

Xeroxing paradox

It appears that the following statements cannot all be true:

1. Unitarity: Hawking radiation contains the information about the state that formed the black hole
2. Equivalence Principle: Nothing special happens to an infalling observer while crossing the horizon

Xeroxing paradox

It appears that the following statements cannot all be true:

1. Unitarity: Hawking radiation contains the information about the state that formed the black hole
2. Equivalence Principle: Nothing special happens to an infalling observer while crossing the horizon
3. Omniscience: There must exist a consistent description of the entire spacetime, even if there are horizons, i.e., even if no-one can observe the entire spacetime.

Xeroxing paradox

It appears that the following statements cannot all be true:

1. Unitarity: Hawking radiation contains the information about the state that formed the black hole
2. Equivalence Principle: Nothing special happens to an infalling observer while crossing the horizon
3. Omniscience: There must exist a consistent description of the entire spacetime, even if there are horizons, i.e., even if no-one can observe the entire spacetime.

Let us sacrifice (3) in order to rescue (1) and (2). This leads to the principle of complementarity.

Complementarity

A fundamental description of Nature need only describe experiments that are consistent with causality. This principle can be applied to arbitrary spacetimes. The regions that can be probed are the *causal diamonds*: the intersection of the past and future of an arbitrary worldline.

Complementarity

Complementarity implies that there must be a theory for every causal diamond, but not necessarily for spacetime regions that are too large to be contained in any causal diamond. If we attempt to describe such regions, we may encounter contradictions, but such contradictions cannot be verified in any experiment.

Complementarity

As the thought-experiments show, complementarity resolves both of the paradoxes described above.

Complementarity

As the thought-experiments show, complementarity resolves both of the paradoxes described above.

But recently, a deep and subtle new gedankenexperiment was proposed by Almheiri, Marolf, Polchinski, Sully, arXiv:1207.3123

Maximal Entanglement

Let $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ and $N_A = \log |\mathcal{H}_A|$.

Let $\rho_A = \mathrm{Tr}_B\, \rho_{AB} = \sum_b \langle b|\rho_{AB}|b\rangle$.

Let $S_A = -\mathrm{Tr}_A\, \rho_A \log \rho_A$.

Assume the bipartite state is pure: $\rho_{AB} = |\psi\rangle\langle\psi|$.

Definition: *A* is maximally entangled with *B* iff $S_A = N_A$.

(This is equivalent to $\rho_A = e^{-N_A} I$, and it implies $N_A \leq N_B$.)

Maximal Entanglement

Example: 2 qubits.

The state $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ is maximally entangled.

Maximal Entanglement

Example: 2 qubits.

The state $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ is maximally entangled.

So are the other "Bell states"
$(|00\rangle - |11\rangle)/\sqrt{2}$,
$(|01\rangle + |10\rangle)/\sqrt{2}$,
$(|01\rangle - |10\rangle)/\sqrt{2}$.

Maximally entangled states form a complete basis if $N_A = N_B$.

Monogamy of Entanglement

Let $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$.

Let $\rho_{AB} = |\psi\rangle\langle\psi|$, where $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ is maximally entangled.

Can B and C be maximally entangled?

Monogamy of Entanglement

Let $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$.

Let $\rho_{AB} = |\psi\rangle\langle\psi|$, where $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ is maximally entangled.

Can B and C be maximally entangled?

No; in fact they cannot be entangled or even correlated at all. This follows from the strong subadditivity of the entanglement entropy:

$$S_{ABC} + S_B \leq S_{AB} + S_{BC} \ .$$

The Page Theorem

Note: this is not actually needed but it simplifies the argument for firewalls.

Consider a bipartite system, $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, with $N_A \leq N_B$, in a Haar-random pure state. The smaller subsystem will be nearly maximally entangled with the larger subsystem:

$$S_A \approx N_A - \frac{1}{2} e^{N_A - N_B} \ .$$

The Unruh State

The Minkowski vacuum can be written as a linear combination of product states with support in the left and right Rindler wedge, $|x| > |t|$:

$$|0\rangle_M = \prod_j \sum_{n_j=0}^{\infty} e^{-\pi\omega_j n_j} |n_j\rangle_R \otimes |n_j\rangle_L \ .$$

Notice that this state is highly entangled. Tracing over the left wedge yields a thermal state for the right wedge.

The Unruh State

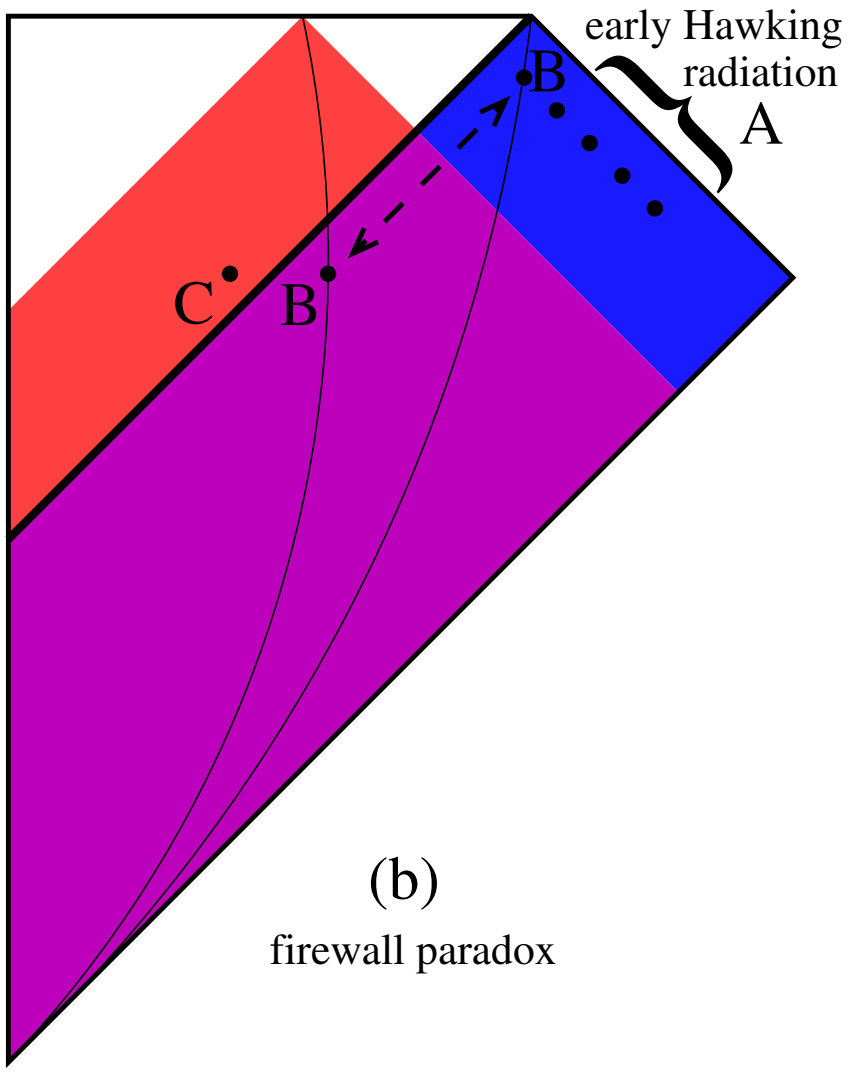
The Minkowski vacuum can be written as a linear combination of product states with support in the left and right Rindler wedge, $|x| > |t|$:

$$|0\rangle_M = \prod_j \sum_{n_j=0}^{\infty} e^{-\pi\omega_j n_j} |n_j\rangle_R \otimes |n_j\rangle_L \ .$$

Notice that this state is highly entangled. Tracing over the left wedge yields a thermal state for the right wedge.

In Hawking's calculation, this thermal state describes well the Hawking quanta when they are still in the "zone", and it explains the thermal nature of the Hawking radiation.

The AMPS argument: Unitarity

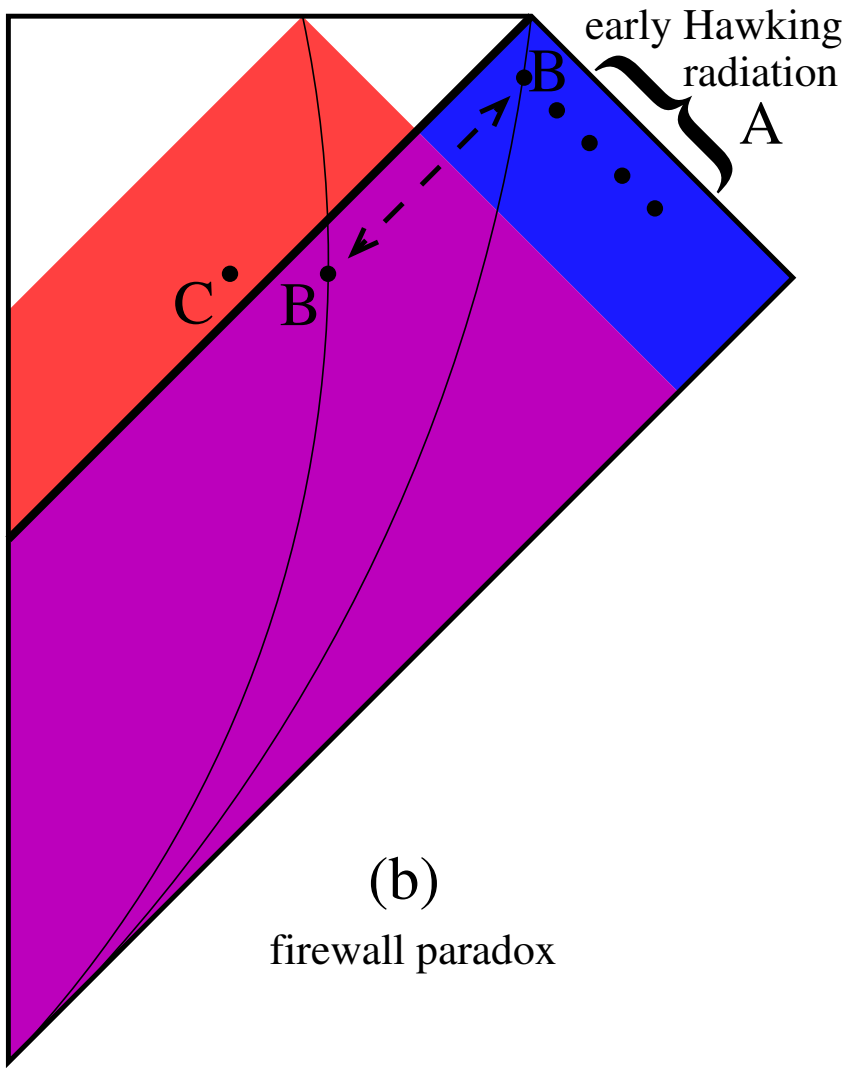


A black hole is allowed to evaporate more than half of its area.

The resulting "early Hawking radiation" *A* is maximally entangled with what remains of the black hole, by unitarity (Page 1993).

Equivalently, the early radiation is maximally entangled with the remaining "late" Hawking radiation.

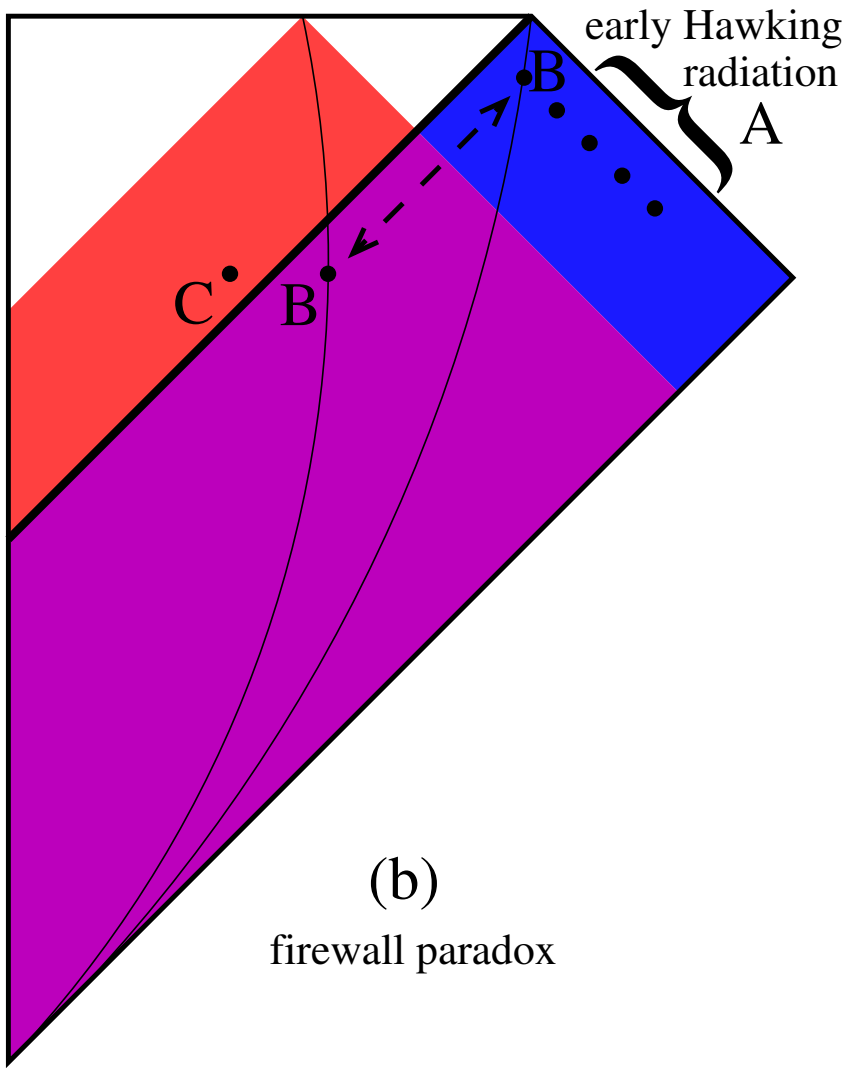
The AMPS argument: Unitarity



Consider one of the quanta of the late radiation, *B*, a wavepacket of size $O(R)$.

The early radiation *A* must in particular be maximally entangled with *B*.

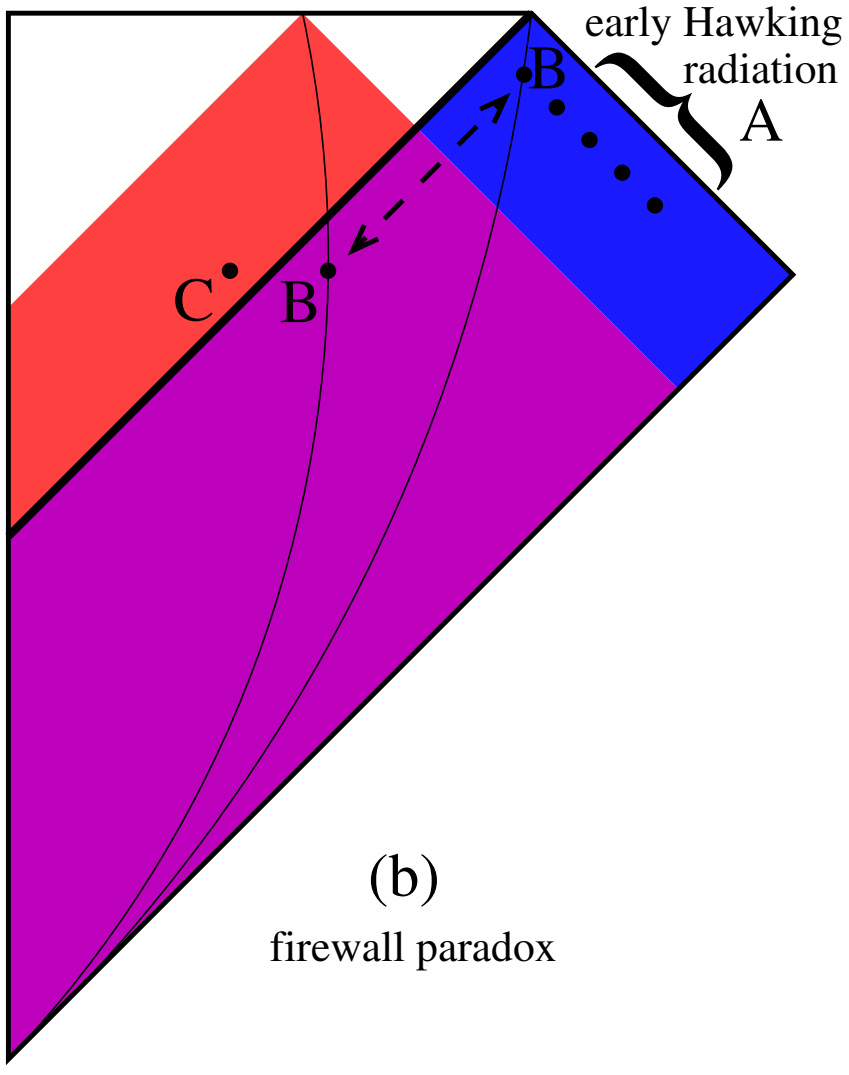
The AMPS argument: Effective Field Theory



Now consider an observer Alice who falls into the black hole.

She crosses the horizon when B is still close to the horizon and has size $1 \ll \lambda \ll R$.

The AMPS argument: Equivalence Principle

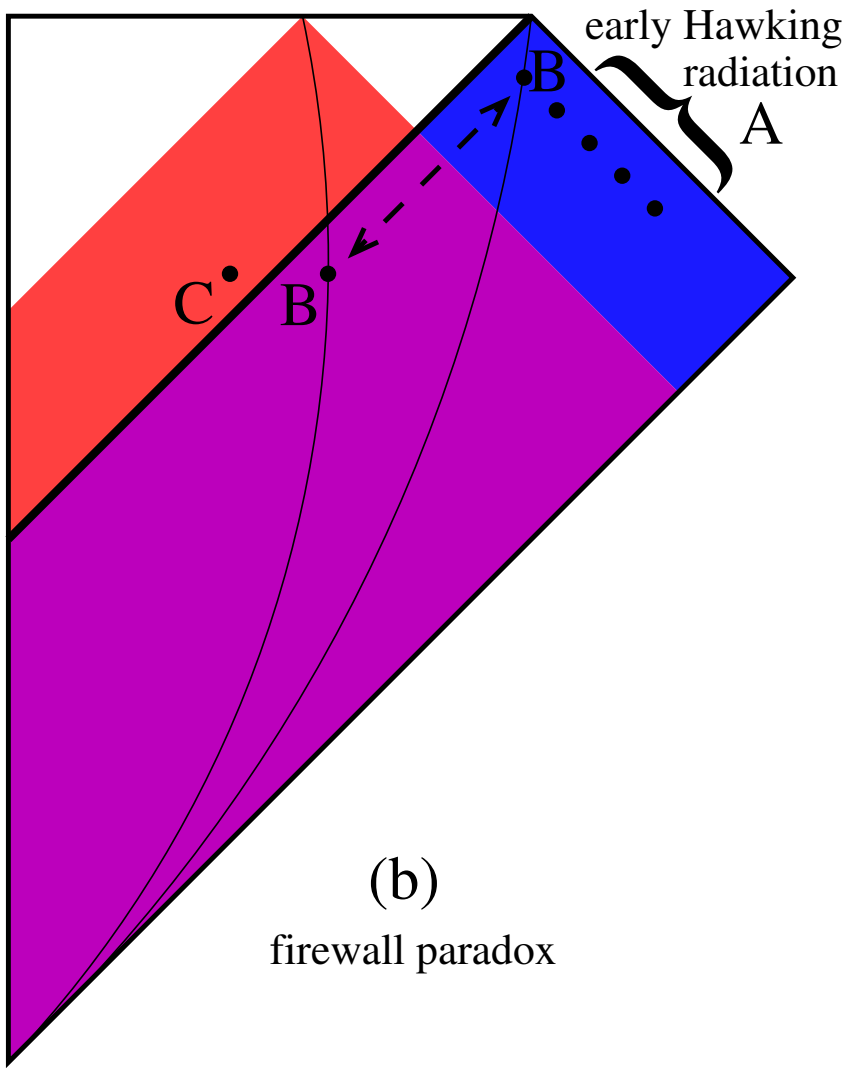


By the equivalence principle, Alice should see flat space behavior on such scales.

But in the Minkowski vacuum, modes with support on one side of a surface are maximally entangled with modes on the other side (Unruh state).

This implies that Alice must see entanglement between *B* and a mode *C* inside the horizon.

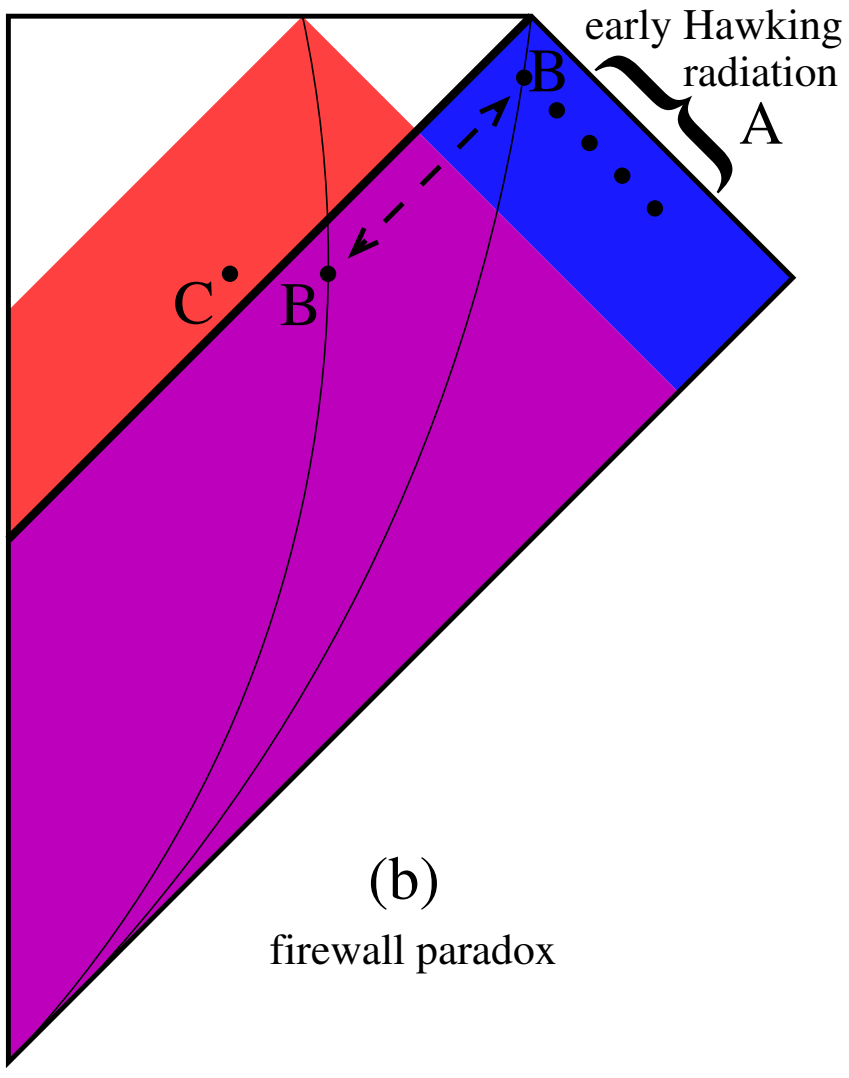
The AMPS argument: Contradiction



This is a contradiction.

By the strong subadditivity of entanglement entropy, *B* cannot be maximally entangled with *A*, and also entangled with *C*).

The AMPS argument: Contradiction

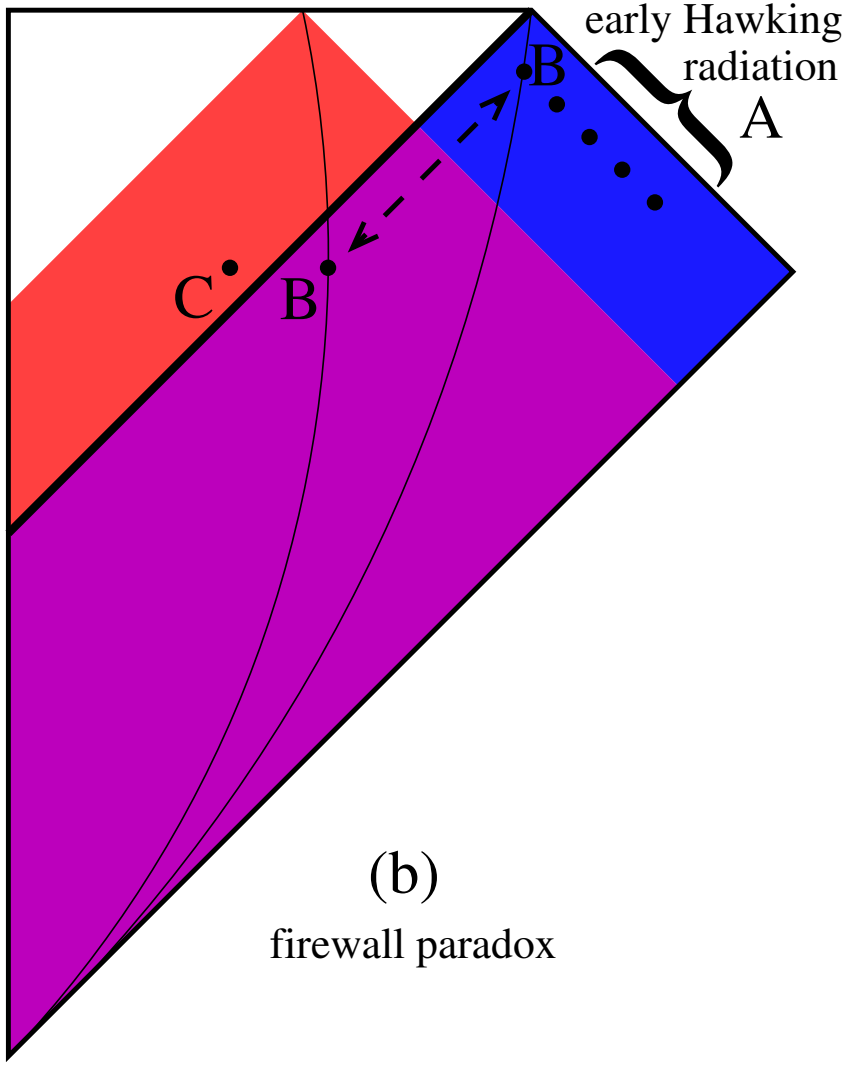


We must give up one of these:

1. Unitarity
   (entanglement of *B* with *A*)
2. QFT
   (evolution of the mode *B* from the near to the far region)
3. Equivalence Principle
   (entanglement of *B* with *C*)

AMPS argue that (3) is the most conservative choice.

The AMPS argument: Firewall



(b)

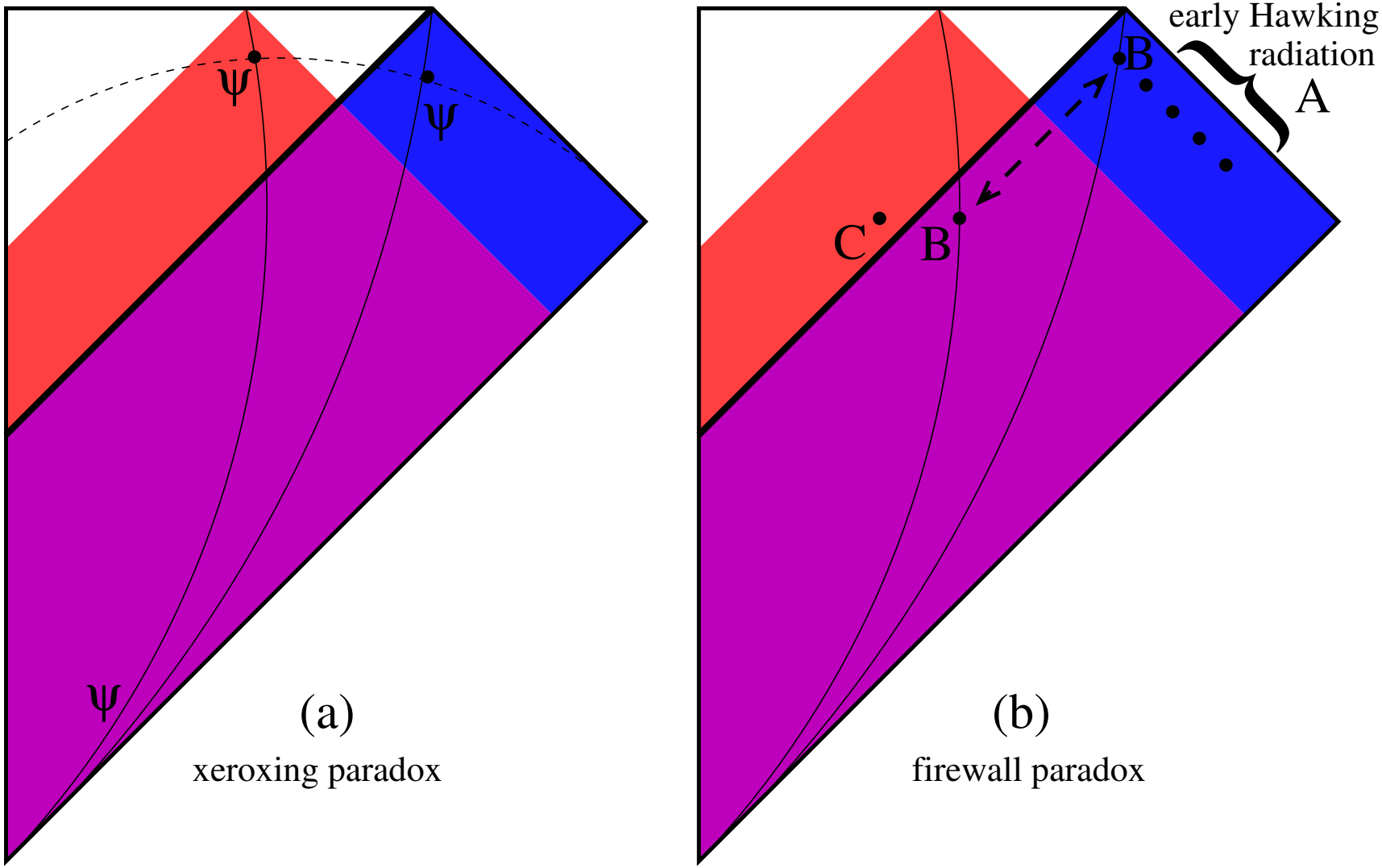
firewall paradox

By a mining argument, the AMPS argument applies not only to outgoing s-waves, but to higher-l near-horizon modes as well. The lack of Unruh entanglement of short distance modes across the horizon implies a divergent stress tensor at the stretched horizon—a "firewall" or singularity. This is similar to the Rindler vacuum.

Alice will hit the firewall and will not enter the black hole.

Why Complementarity is Not Enough

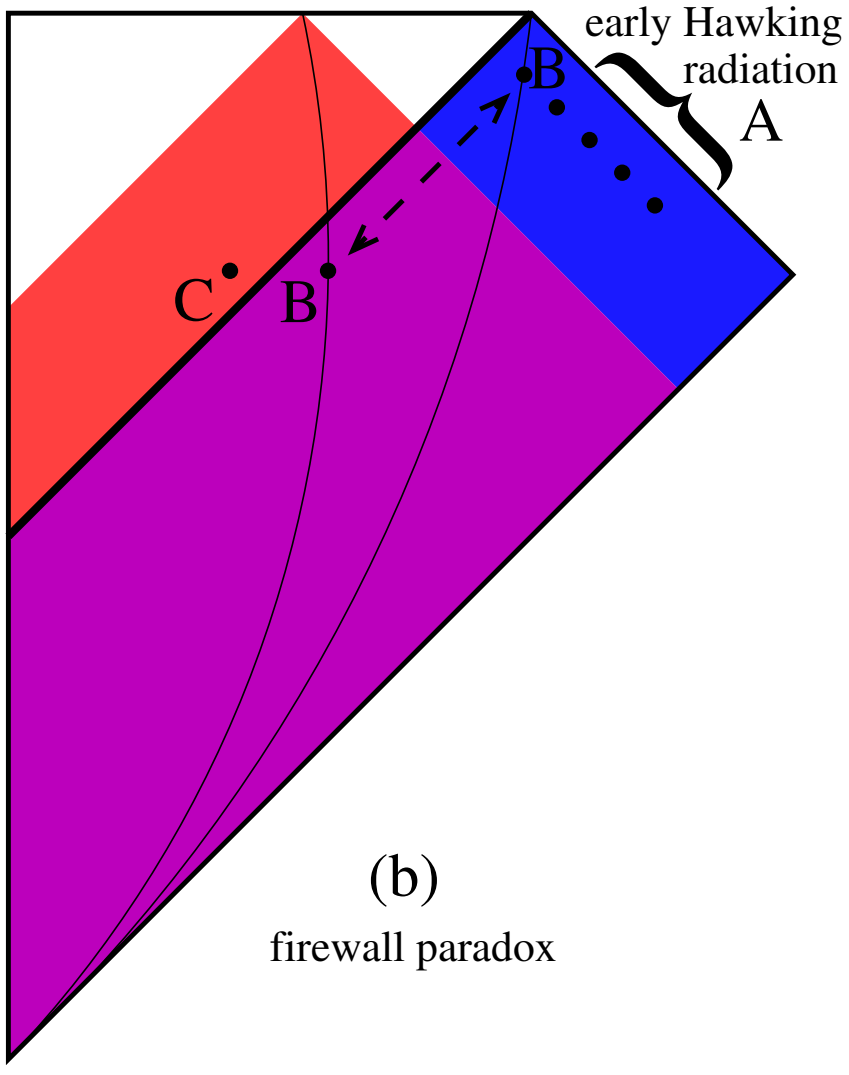
How Complementarity might help:



Xeroxing paradox: each observer sees only one copy

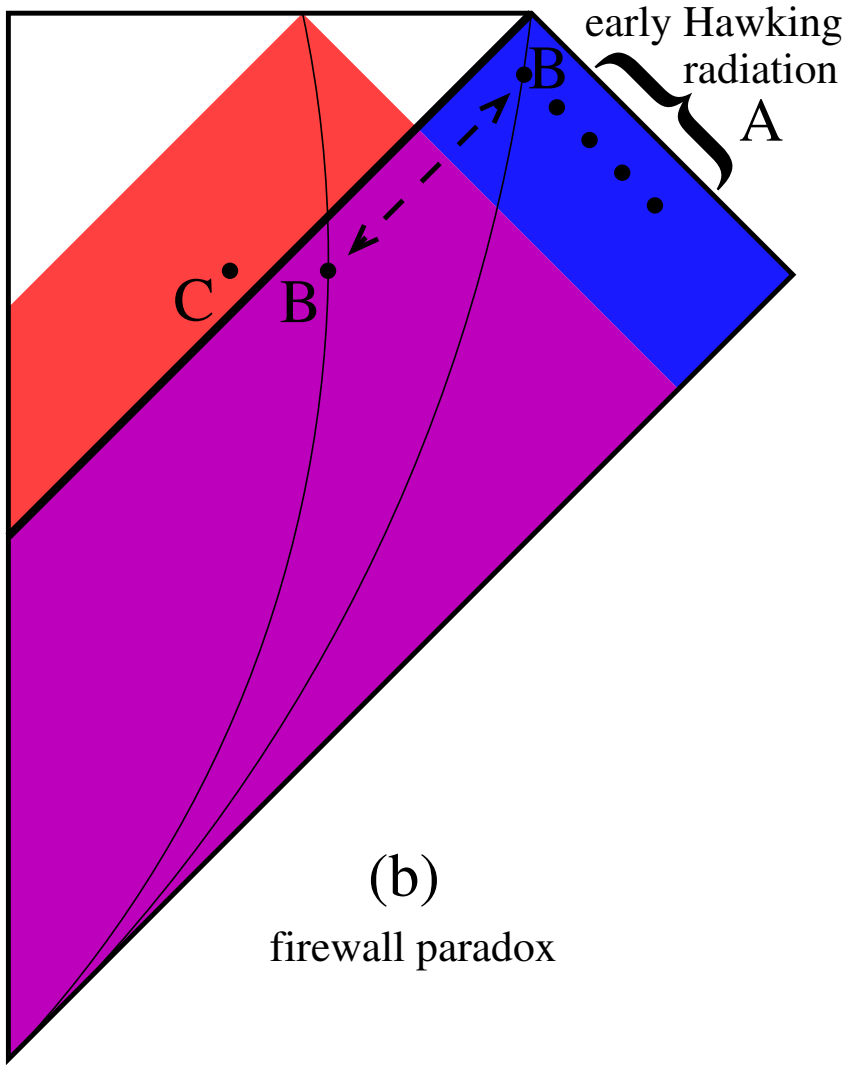
Firewall paradox: each observer sees only one entanglement

The infalling observer



Alice will find that the early Hawking radiation is consistent with unitarity. However, she cannot measure the full S-matrix and she cannot verify that the late radiation purifies the quantum state of the Hawking radiation. She can consistently assume that *B* is entangled only with *C*.

The infalling observer

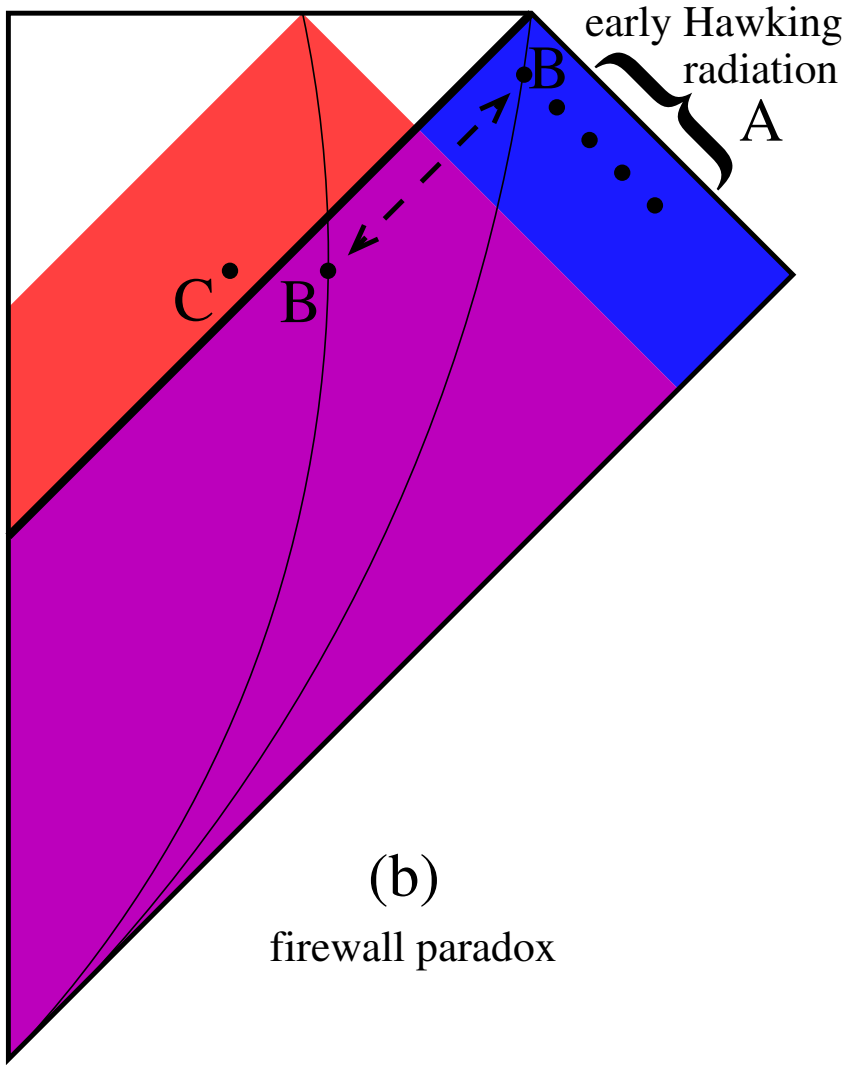


This implies that there is a separate theory for different infalling observers.

This is consistent with the version of complementarity advocated earlier.

It is not consistent with alternate statements of complementarity that refer to the black hole interior in its entirety.

The outside observer



Bob can measure the full S-matrix and finds it to be unitary. He cannot probe the horizon under free fall. To him, Alice seems to be absorbed by a membrane, and Alice can do nothing to prove him wrong. By the time she has verified that *B* is entangled with *C*, it is too late to send Bob a message, so Bob can consistently assume that *B* is entangled only with *A*.

Consistency check

For complementarity to resolve the xeroxing paradox, it is important that no observer can see both copies. This has been verified.

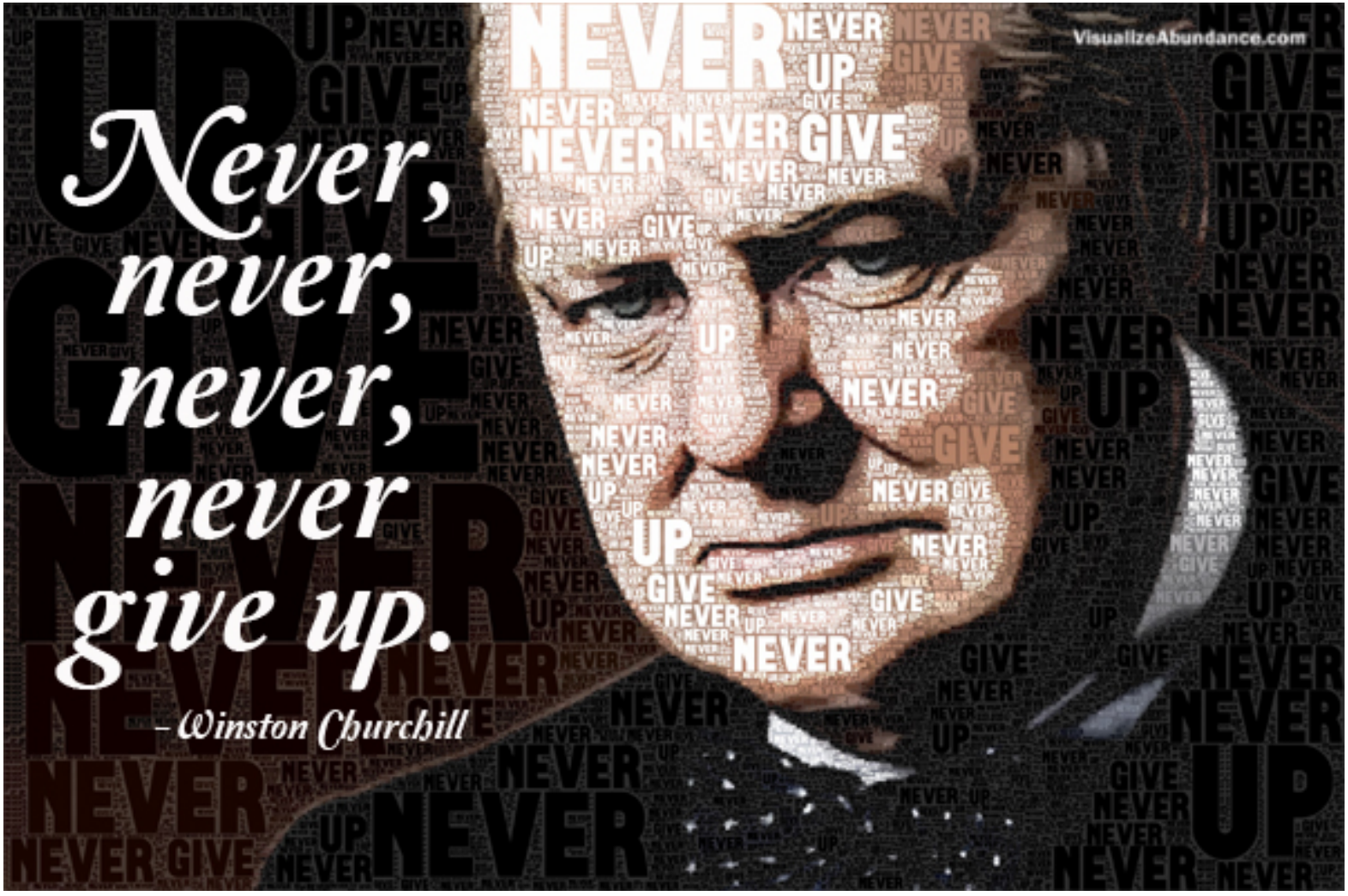
For complementarity to resolve the firewall paradox, it is important that no observer can see both entanglements. This consistency check appears to fail.

The problem

- Alice can measure B before she crosses the horizon (at least, I see no in-principle obstruction), so her theory at this point must contain a Hilbert space for B.
- Since she can still choose to return to infinity, her theory must contain a Hilbert space for R.
- By unitarity, B must be maximally entangled with some purification $R_B \subset R$ in the early Hawking radiation.
- Free fall without a firewall requires that B is maximally entangled with the interior modes A. This is impossible.

Now what?

Now what?


VisualizeAbundance.com
Never, never, never, never give up.
– Winston Churchill

Now what?

Except for giving up on unitarity, two possible stances:

(1) Attempt to preserve the equivalence principle.

Now what?

Except for giving up on unitarity, two possible stances:

(1) Attempt to preserve the equivalence principle.

- What will it take?

Now what?

Except for giving up on unitarity, two possible stances:

(1) Attempt to preserve the equivalence principle.

- What will it take?

(2) Make sense of firewalls:

- How do they form?
- When do they form?
- Do they grow?
- Why do they form?

Now what?

Sharp Paradox

=

Huge Opportunity