27 January 2021 Lecture 5: Hawking's original paradox

Last time, we derived a formula for the temperature of a black hole

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

Even before this derivation it was known that black holes obey

$$\underbrace{dM}_{\text{mass}} = \frac{\kappa\, dA}{8\pi} + \underbrace{\Omega}_{\text{angular velocity}}\, d\underbrace{J}_{\text{angular momentum}} + \underbrace{\Phi}_{\text{electric potential}}\, d\underbrace{Q}_{\text{charge}}.$$

$$dA \geq 0$$

This is a lot like the laws of thermodynamics

$$dU = T\,dS + \text{work-terms}$$

$$dS \geq 0$$

Before the derivation of the temperature one might have thought that this was a manifestation of "the same eqns have the same solution" (Feynman).

But the temperature derivation tells us we should take the thermodynamic interpretation seriously.

Also

$$S = \frac{A}{4}$$

So, black holes are thermodynamic objects with a lot of entropy.

For the sun

$$S \sim 10^{76} \,!$$

We will return to the significance of this issue repeatedly.

Now turn to our first paradox.

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 Info paradox was first outlined by Hawking in "Breakdown of predictability in gravitational collapse"

Several versions of the Info paradox have been developed, and it is common to use the term "Hawking paradox" for arguments that do not appear in Hawking's paper.

We will closely follow the argument in this paper.
This paradox is relatively easy to solve.
We will later encounter more formidable paradoxes.

From Hawking's paper

discovery by this author[12,13] that black holes create and emit particles at a steady rate with a thermal spectrum. Because this radiation carries away energy, the black holes must presumably lose mass and eventually disappear. If one tries to describe this process of black-hole evaporation by a classi-

This is simple to understand.

Since the black hole radiates at temperature T, its mass decreases at the rate

$$\frac{dM}{dt} = - C \, A \, T^{d+1}$$

Constant depends on greybody factors

Area of horizon

spacetime dim is T^{d+1}

Recall $M \propto r_h^{d-2}$; $T \propto \frac{1}{r_h}$; $A \propto r_h^{d-1}$

So

$$\frac{dr_h}{dt} \propto \frac{-1}{r_h^{d-1}}$$

So in

$$\text{time} \propto r_h^d \propto \frac{A}{T}, \quad r_h \to 0.$$

So after some time in an isolated Universe, a black hole would evaporate entirely.

The flux of radiation <u>increases</u> as the black hole becomes smaller.

to be very different from that of Minkowski space. For example, in the case of gravitational collapse which produces a black hole there is an event horizon which prevents observers at infinity from measuring the internal state of the black hole apart from its mass, angular momentum, and charge. This means that measurements at future infinity are insufficient to determine completely the state of the system at past infinity: One also needs data on the event horizon describing what fell into the black hole. One might think that one could have observers stationed just outside the event horizon

Next, Hawking points out that even classically the state at $\mathcal{I}^+$ is not enough to dermine the state on $\mathcal{I}^-$

We need data on $H \cup \mathcal{I}^+$

$\mathcal{I}^-$ ← to get data here

An operator Q which corresponds to an observable at future infinity will be composed only of the $\{b_j\}$ and the $\{b_j^\dagger\}$ and will operate only on the vectors $|A_I\rangle$. Thus the expectation value of this operator will be

$$\langle 0_-|Q|0_-\rangle = \sum\sum \rho_{AC} Q_{CA} \tag{3.4}$$

where $Q_{CA} = \langle C_I|Q|A_I\rangle$ in the matrix element of the operator Q on the Hilbert space of outgoing states and $\rho_{AC} = \sum \lambda_{AB}\bar{\lambda}_{CB}$ is the density matrix which completely describes all observations which are made only at future infinity and do not measure what went into the hole. The components of ρ_{AC} can be completely determined from the expectation values of polynomials in the operators $\{b_j\}$ and $\{b_j^\dagger\}$. Thus the density matrix is independent of the ambiguity in the choice of the $\{q_j\}$ which describes particles going into the hole.

As an example of such a polynomial consider

previously with the field expanded as

$$\phi = \sum_\ell \int d\omega \left[A_{\omega,\ell} f^{out}(\omega,\ell,r_*) + B_{\omega,\ell} f^{in}(\omega,\ell,r_*)\right] e^{-i\omega t} Y_\ell(\Omega) + h.c.$$

We defined $a_{\omega,\ell}$ as smeared versions of $A_{\omega,\ell}$ and found that

$$\langle a_{\omega,l}\, a^{+}_{\omega,l} \rangle = \frac{1}{1-e^{-\beta\omega}}$$

This leads to a flux at future null infinity.

$$F^{out}(\omega, l, r_*)\, e^{-i\omega t} \sim \frac{1}{r^{(d-1)/2}}\, e^{i\omega u}\, t_{\omega,l}.$$

So $T_{uu} \sim \partial_u\phi\, \partial_u\phi$ is non-zero at I^+

But (a) This flux does not care about the state of the $B_{\omega,l}$ [as we discussed during the lectures] therefore

$$\langle n_j \rangle = |t_\omega|^2 (e^{2\pi\omega\kappa^{-1}} - 1)^{-1} . \tag{3.10}$$

This is precisely the expectation value for a body emitting thermal radiation with a temperature $T = \kappa/2\pi$. To show that the probabilities of emitting different numbers of particles in the jth mode and not just the average number are in agreement with thermal radiation, one can calculate the expectation values of n_j^2, n_j^3, and so on. For example,

$$\langle n_j^2 \rangle = \langle 0_- | b_j^\dagger b_j b_j^\dagger b_j | 0_- \rangle$$
$$= \langle n_j \rangle + \langle 0_- | (b_j^\dagger)^2 (b_j)^2 | 0_- \rangle . \tag{3.11}$$

One can evaluate the second term on the right-hand

(1) We can compute this flux. Hawking computes the term with the greybody factor, which is directly related to the flux at infinity

We compute for simplicity polynomials of

$$N_\omega = a_\omega^\dagger a_\omega$$ ← Note ordering

$$N_\omega = \frac{1}{e^{\beta\omega}-1} \; ; \quad N_\omega^2 = \frac{1+e^{\beta\omega}}{(e^{\beta\omega}-1)^2} \; \ldots$$

One way to see this is to focus on the equation

$$\left(\tilde{a}_{\omega,l} - a^{+}_{\omega,l} e^{-\beta\omega/2}\right)|\psi\rangle = 0$$

If we focus <u>only</u> on the state of these two harmonic oscillators we can "solve"

$$|\psi\rangle = e^{e^{-\beta\omega/2} a^{+}_{\omega,l} \tilde{a}^{+}_{\omega,l}} \, |N_\omega = 0, \tilde{N}_\omega = 0\rangle$$

$$\otimes \, (\text{other dof}).$$

For such a state

$$\langle N^{a}_{\omega,l} \rangle = \mathrm{tr}\left(P_{\omega,l} N^{a}_{\omega,l}\right)$$

where

$$p_{\omega,l} = \frac{1}{1-e^{-\beta\omega}} e^{-\beta\omega N_{\omega,l}}$$

is the thermal density matrix.

If one specifies the expectation value of all polynomials of $N_{\omega,l}$ for arbitrary ω, l this seems to specify the state at $\mathcal{I}^+$.

Note this density matrix does not correspond to a pure state.

So even if we start with a pure state on $\mathcal{I}^-$, we seem to always end up with a thermal state on $\mathcal{I}^+$.

This is obviously a paradox. Under unitary evolution

$$|\psi\rangle \to U|\psi\rangle$$

and so the initial density matrix

$$|\psi\rangle\langle\psi| \to U|\psi\rangle\langle\psi|U^{\dagger}$$

The black hole seems to produce a mixed state at late times

$$|\psi\rangle\langle\psi| \longrightarrow \sum |c_i|^2 \; |\psi_i\rangle\langle\psi_i|$$

which is inconsistent with unitary evolution.

A pure state contains all information about the system.

A mixed state is a probabilistic mixture of pure states.

A mixed state can always be thought of as arising because we have a large system and throw away information about one part.

So the evolution from a pure to a mixed state suggests loss of information.

time regions about which one has no knowledge. One therefore has to introduce a hidden surface around each of these holes and apply the principle of ignorance to say that all field configurations on these hidden surfaces are equally probable provided they are compatible with the conservation of mass, angular momentum, etc. which can be measured by surface integrals at a distance from the hole.

Let H_1 be the Hilbert space of all possible data on the initial surface, H_2 be the Hilbert space of all possible data on the hidden surface, and H_3 be the Hilbert space of all possible data on the final surface. The basic assumption of quantum theory is that there is some tensor S_{ABC} whose three indices refer to H_3, H_2, and H_1, respectively, such that if

$$\xi_C \in H_1, \quad \zeta_B \in H_2, \quad \chi_A \in H_3,$$

then

$$\sum \sum \sum S_{ABC} \chi_A \zeta_B \xi_C$$

is the amplitude to have the initial state ξ_C, the final state χ_A, and the state ζ_B on the hidden surface. Given only the initial state ξ_C one cannot determine the final state but only the element $\sum S_{ABC}\xi_C$ of the tensor product $H_2 \otimes H_3$. Because one is ignorant of the state on the hidden surface one cannot find the amplitude for measurements on the final surface to give the answer χ_A but one can calculate the probability for this outcome to be $\sum\sum\sum \rho_{CD} \bar{\chi}_C \chi_D$, where

$$\rho_{CD} = \sum \sum \sum \bar{S}_{CBE} S_{DBF} \bar{\xi}_E \xi_F$$

is the density matrix which completely describes observations made only on the future surface and not on the hidden surface. Note that one gets this density matrix from $\sum S_{ABC}\xi_C$ by summing with equal weight over all the unobserved states on the "hidden" surface.

One can see from the above that there will not

Hawking explained his intuition for why this was happening.

Since the observer at $\mathcal{I}^+$ does not know about the "Bw" modes, this observer must adopt a principle of ignorance about the interior.

Say the state on $\mathcal{I}^-$ can be represented in a Hilbert space H_1

$$\psi^-_A \in H_1$$

↑ to emphasize state is a vector

$\psi^H_B \in H_2$ ← state of horizon or interior "hidden surface"

$\psi^+ \in H_3$ ← state on $\mathcal{I}^+$

$\boxed{\text{Final state part of } H_2 \otimes H_3}$ ← Note assumption. follows from assuming locality.

Then unitary evolution tells us we can compute an amplitude using a "S-matrix"

$$S_{ABC}\, \psi^-_A\, \psi^H_B\, \psi^+_C$$

If we later "throw away" information about ψ^H_B, we can only compute the provability to obtain ψ^+ as the final state not the amplitude.

The provability is

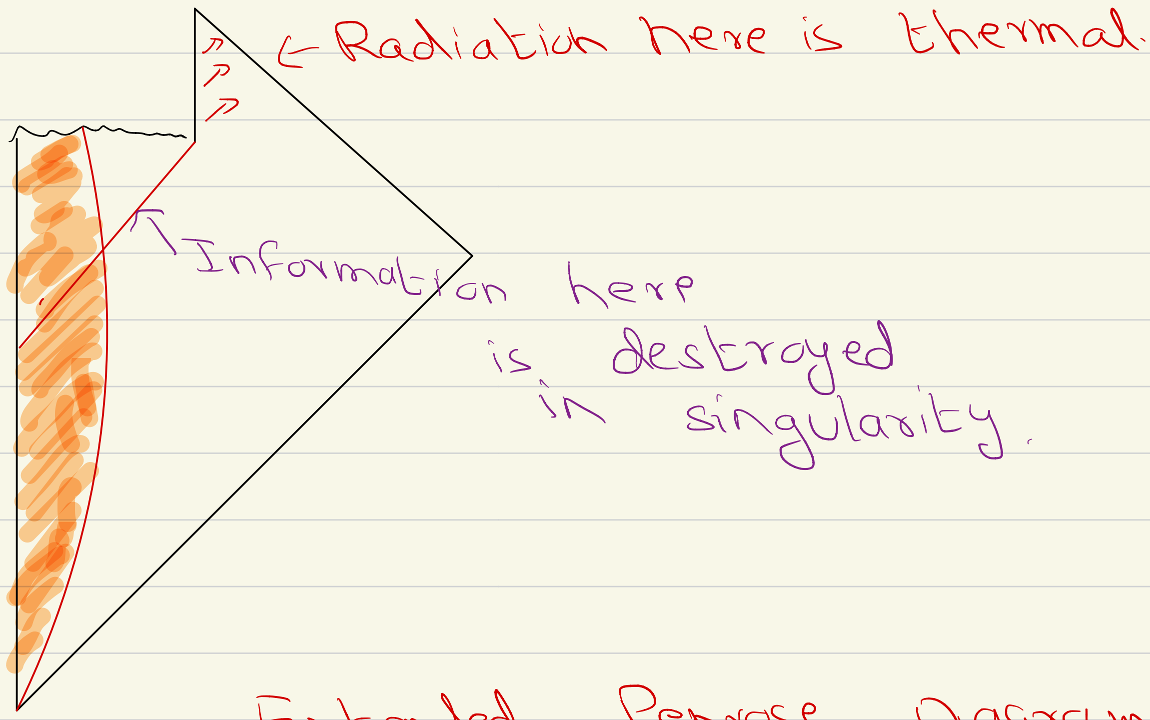
$$\sum_{\psi^H} S_{ABC}\, \psi^-_A\, \psi^H_B\, \psi^+_C\, S^*_{\tilde{A}\tilde{B}\tilde{C}}\, \psi^{-*}_{\tilde{A}}\, \psi^{H*}_{\tilde{B}}\, \psi^{+*}_{\tilde{C}}.$$

$$= \underbrace{\sum S_{ABC}\, S^*_{\tilde{A}B\tilde{C}}\, \psi^-_A\, \psi^{-*}_{\tilde{A}}}_{P_{C\tilde{C}}}\, \psi^+_C\, \psi^{+*}_{\tilde{C}}.$$

By "tracing out" part of the final state using the "principle of ignorance", we get a density matrix.

Hawking referred to a "superscattering operator" that would take us between pure states and mixed states.

There is a way to summarize this argument diagrammatically



Extended Penrose Diagram

This is not a rigorous Penrose diagram.

It is meant to schematically show how information is lost, and the causal structure of a B.H. spacetime is different from Minkowski space.

formation in the final stages of the evaporation. However, information like baryon number requires energy and there is simply not enough energy available in the final stages of the evaporation. To carry the large amount of information needed would require the emission in the final stages of about the same number of particles as had already been emitted in the quasistationary phase.

Because one ends up with a density operator rather than pure quantum space, the process of

There is a final important point Hawking made.

In the final stages of evaporation, the B.H. becomes very small and Q.G. is important.

could all the information emerge in the end?

No! Information is stored in microstates

If the number of microstates below a given energy is bounded, information about an arbitrarily large black hole cannot be stored with small energy.

This is why "remnants" are not viable

We would now like to examine if Hawking's argument for information loss is valid.

The main calculation is the thermal spectrum for the $a_{\omega,\ell}$ modes.

This is supported by intuition from the causal structure of the geometry.

We would like to ask two questions

1) Does Hawking's calculation really imply that the final state is mixed?

2) Is the intuition from the causal structure valid at the accuracy required.

First we will take a detour into quantum statistical mechanics.