Black Hole Information Paradox

Delon Shen

Notes for Suvrat Raju's Black Hole Information Paradox course at ICTP(well really online) during Spring 2021. Course website can be found here which contains notes, assignments, and links to lecture videos. This course closely follow a review posted here. If you have any comments let me know at hi@delonshen.com.

Last updated on February 4, 2021

Lecture 1: Introduction and Two-Point QFT Correlators

January 13,2021

The main organization of this course

(a) Hawking's Original Paradox → Thermalization and exponentially small corrections.

(b) Paradoxes about interior of evaporating Black Holes → holography of information, islands and page curve.

(c) Paradoxes about large Black Holes in AdS/CFT→ Mirror operators, state-dependence, and firewalls/furrballs

Lets start by talking about **Hawking Radiation**, it's the effect that underlies the information paradox. Take a black hole in asymptotically flat space. This black hole radiates with a temperature $\propto$ surface gravity. We should also recall that hawking radiation relies on short distance QFT physics and on global late-time properties of the black hole geometry. The interesting thing is that the derivation for Hawking's radiation also implies the existence of the entangled modes across horizons. So what are the common derivations of hawking radiation(TODO (a) is in appendix of review paper and (b) might be in wald)?

(a) Hawking's original derivation

(b) Rindler $\leftrightarrow$ Minkowski Bogolivlov transformation

In this course we'll consider a different derivation from both of these

Lets take a second to step back from black hole and look at Quantum Fields near a null surface. We'll apply what we learn here to black holes later. What we want to show is that across any null surface in a smooth state (TODO smooth state who?) we can isolate a "local" QFT (which we'll define in a bit) with universal entanglement. This is useful because we'll find that in a black hole spacetime local degrees of freedom near the horizon gives global modes in blackhole geometry.

First lets define what we mean by a smooth metric around some point. Consider a point in some $D = d + 1$ space and let this point be the origin. We have U, V, two null coordinates, and $d - 1$ transverse coordinates. A metric is smooth around some point if around some point we can locally choose some coordinates so the metric takes the following form. (think light cone variant Kruskal coordinates in arbitrary dimensions?)

$$ds^2 = -dUdV + \delta_{\alpha\beta} dy^\alpha dy^\beta + \dots$$

Where $dUdV$ are two null coordinates and α, β is over $d - 1$ indices and where the $\dots$ terms vanish near origin.

$$figure$$

We also want to make an additional demand. Consider a scalar field ϕ and points near $U = 0$. If we're still thinking in terms of Kruskal coordinates this means we're thinking of things close to eachother on each side of the horizon? In the limit where x_1 approaches x_2 for any nonsingular state the two point correlation function (Wightman function?) becomes.

$$\langle \phi(x_1)\phi(x_2)\rangle = \frac{N}{|x_1 - x_2|^{d-1}} + \dots$$

We also impose the following scales

(a) $|x_1 - x_2| \ll \ell_{\text{curvature}}$

(b) $|x_1 - x_2| \ll \frac{1}{m}$

(c) $|x_1 - x_2| \gg \ell_{\text{Pl}}$ or any UV scale where EFT breaks down.

These length scales give us the normalization if we consider a free field (e.g. $\mathcal{L} = 1/2(\partial_\mu \phi)^2$).

$$N = \frac{\Gamma(d-1)}{2^d \pi^{d/2} \Gamma(d/2)} \Rightarrow \langle \phi(x_1)\phi(x_2)\rangle = \frac{\Gamma(d-1)}{2^d \pi^{d/2}\Gamma(d/2)} \frac{1}{|x_1 - x_2|^{d-1}} + \dots$$

Because of the length scales we assume we can say that the structure of the two point function is universal (TODO what in the world.) Before we continue lets look at a few things that will be useful

$$|x_1 - x_2|^2 = -\delta U \delta V + \delta_{\alpha\beta}\delta y^\alpha \delta y^\beta \qquad \delta O = O_1 - O_2$$

Also if we grind through some calculations we'll find that

$$\langle \partial_{U_1}\phi(x_1)\partial_{U_2}\phi(x_2)\rangle = -\frac{d^2-1}{4}\frac{N(\delta V)^2}{|x_1 - x_2|^{d+3}} + \dots$$

Taking $\delta V \to 0$, e.g. we take δV to be the smallest separation, then we find that

$$\lim_{\delta V \to 0} \frac{(\delta V)^2}{(-\delta U \delta V + \delta y^\alpha \delta y^\beta \delta_{\alpha\beta})^{(d+3)/2}} \neq 0$$

It's not zero since it does receive a contribution when $y^\alpha = 0$. To see this we can do an integral over all the transverse separations.

$$\int \frac{(\delta V)^2}{(\delta U \delta V + \delta y^\alpha \delta y^\beta \delta_{\alpha\beta})^{(d+3)/2}} d^{d-1}\delta y^\alpha$$

We'll also take the $|x_1 - x_2|$ is positive. Now with the substitution $\delta \tilde{y}^\alpha = \delta y^\alpha / \sqrt{-\delta U \delta V}$ we get

$$\frac{1}{(\delta V)^2}\int \frac{d\delta \tilde{y}^\alpha}{[1 + \delta\tilde{y}^\alpha \delta \tilde{y}^\alpha]^{(d-3)/2}}$$

What happens in the end is that all factors of δV cancel. In the notes Suvrat says that for $\delta y^\alpha \neq 0$ the integral vanishes. I think this might be due to symmetry, there's a ring of δy^α with appropriate sign that cancels out when we do the integral. But since $\delta y^\alpha = 0$ doesn't have this cancellation it remains finite. In the end we get

$$\boxed{\lim_{\delta V \to 0} \langle \partial_{U_1}\phi(x_1)\partial_{U_2}\phi(x_2)\rangle = -\frac{1}{4\pi}\frac{\delta^{d-1}(\delta y^\alpha)}{(U_1 - U_2 - i\epsilon)^2}}$$

Something to note: the reason we did an integral was to pick up the coefficient of the delta function. How do we sniff out the presence of a delta function. We use the property $\int \delta(x)dx = 1$ with the fact that the integral vanishes for $x \neq 0$. Thus if the integral we considered above gave a finite answer then we know the two point correlation function of the derivatives of the states would be proportional to a delta function.

The next step which will happen in the next lecture would to define modes as approximately

$$\int \partial_\mu \phi (-U)^{i\omega}$$

What is this doing? It's picking up the right moving modes with constant V. (TODO huh?)

Two-point function normalization

First we'll consider a free field governed by the lagrangian in $D = 3+1$ flat space time.

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 + \frac{m^2}{2}\phi^2$$

Lecture 2: Entangled Modes accross Null Surfaces

January 14,2021

Last time we mentioned that we can extract right moving modes accross the null surface $U = 0$ with an integral (this is an approximate expression, we'll get a more precise integral in a bit)

$$\int \partial_U \phi (-U)^{i\omega} dU \qquad \int \partial_U \phi U^{i\omega} dU$$

Note that we can't integrate over a large region in U because this would violate our limit of x_1 approaching x_2 which we derived the form of the two point function for in the last lecture. So then we know that we should integrate over a small region of U instead(with respect to the length scales defined in the last lecture).

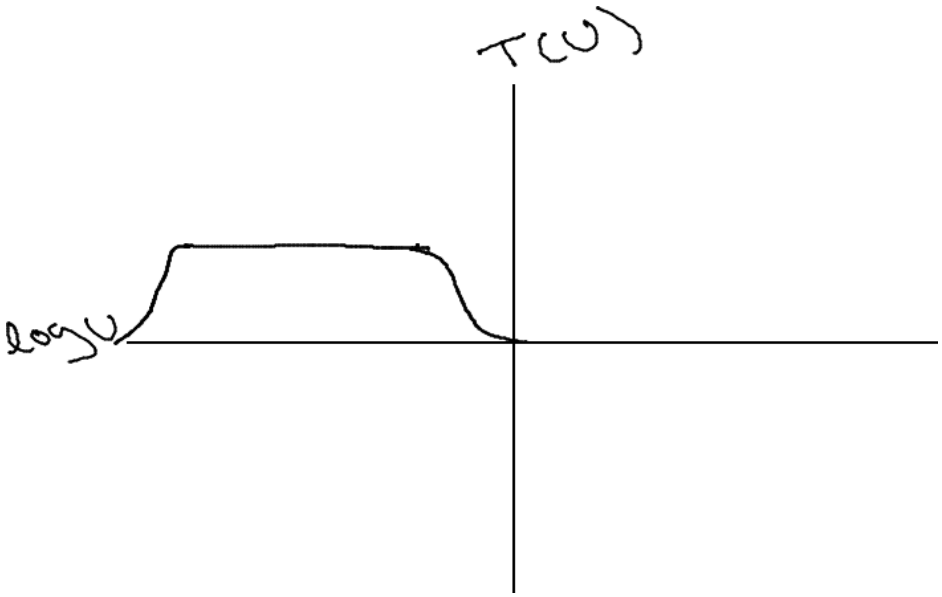
We'll start by introducing a smearing function (TODO what?) $T(U)$ with the following properties

(a) $T(U)$ dies off smoothly near $U \to 0$

(b) Support in interval $[U_l, U_r]$ where $\ell_{\text{UV}} \ll U_r, U_l \ll \ell_{\text{curv}}$ and $\frac{U_r}{U_\ell} \gg 1$ and $U_l, U_r > 0$

(c) We normalize the semaring function $\int T(U)^2 dU/U = 2\pi$. Note that $dU/U = d\log U$.

(d) $T(U)$ is flat for a large range of $\log U$.



The next thing we need to do is define what's happening in the transervse direction by integrating over a volume Vol in the transverse direction which is smaller than a cube of the curvature scale. From this we can write a more precise expression for the mode.

$$a_{\omega_0} = \int (\partial_U \phi(U, V = 0, y^\alpha))(-U)^{-i\omega_0} T(-U) dU \frac{d^{d-1}y^\alpha}{\sqrt{\pi\omega_0 \text{Vol}}}$$

We can similarly define a similar integral for the other side of the null surface. Note we let $V = -\epsilon$ to ensure that we're considering points that are spacelike separated.

$$\tilde{a}_{\omega_0} = \int (\partial_U \phi(U, V = -\epsilon, y^\alpha))(U)^{i\omega_0} T(U) dU \frac{d^{d-1}y^\alpha}{\sqrt{\pi\omega_0 \text{Vol}}}$$

The $T(U)$ insures we only integrate over a small region in U and we have some normalization factors put in that aren't motivated from what I can tell.

Let's now compute the two point function of a and $\tilde{a}$ and we will find that this will only depend on the short distance field correlator that we found last lecture. First spelling things out

$$\begin{aligned}\langle a\tilde{a}\rangle &= \frac{1}{\pi \text{Vol}\omega_0} \int dU_1 dU_2 \langle \partial_{U_1}\phi(U_1, V = 0, y_1)\partial_{U_2}\phi(U_2, V = -\epsilon, y_2)\rangle \times \\ &\times (-U_1)^{-i\omega_0} U_2^{i\omega_0} T(-U_1)T(U_2) d^{d-1}y_1 d^{d-1}y_2 \\ &\text{Remember that the correlator gives a delta function} \\ &= -\frac{1}{4\pi^2\omega_0} \int \frac{1}{(U_1 - U_2)^2} \left(\frac{U_2}{-U_1}\right)^{i\omega_0} T(-U_1)T(U_2) dU_1 dU_2\end{aligned}$$

To do this integral we need the identity

$$\frac{1}{U_1 - U_2}^2 = \frac{1}{(-U_1)U_2} \int_{-\infty}^{\infty} \frac{\omega e^{-\pi\omega}}{1 - e^{-2\pi\omega}} (U_2/(-U_1))^{-i\omega} d\omega$$

When $U_1 < 0$ and $U_2 > 0$. Assume that $|U_1| > |U_2|$. This lets do a contour integral where there are poles at $\omega = in$ where $n \in \mathbb{N}$. From here we can sum the residuals. What are the residuals at the poles? Well mathematica can tell us and so can Suvrat.

$$\frac{1}{(U_1 - U_2)^2} = \frac{1}{|U_1|U_2} \sum_{n=1} -n(-1)^n (U_2/|U_1|)^n$$

We can also plug in the identity (before computing the residuals) into $\langle a\tilde{a}\rangle$ to get

$$\begin{aligned}\langle a\tilde{a}\rangle &= \frac{1}{4\pi^2\omega_0} \int \frac{dU_1}{U_1}\frac{dU_2}{U_2} (U_2/(-U_1))^{-i(\omega-\omega_0)} \frac{\omega e^{-\pi\omega}}{1 - e^{-2\pi\omega}} T(-U_1)T(U_2) d\omega \\ &= \int T(-U_1)(1/(-U_1))^{i(\omega_0-\omega)} dU_1/U_1 \times \int T(U_2) U_2^{i(\omega_0-\omega)} dU_2/U_2 \times \int \frac{\omega e^{-\pi\omega}}{1 - e^{-2\pi\omega}} d\omega\end{aligned}$$

Now note that we can rewrite this in terms of $\log(U)$ since $dU/U = d\log U$.

$$= \int T(-U_1) e^{-i(\log[-U_1])(\omega_0-\omega)} d\log U_1 \times \int T(U_2) e^{i(\log U_2)(\omega_0-\omega)} d\log U_2 \times \int \frac{\omega e^{-\pi\omega}}{1 - e^{-2\pi\omega}} d\omega$$

The first two integrals are fourier transforms of T. Namely $S(\gamma) = \frac{1}{2\pi}\int_0^\infty T(U)U^{-i\gamma} dU/U$

$$= \frac{1}{\omega_0} \int \frac{\omega e^{-\pi\omega}}{1 - e^{-2\pi\omega}} |S(\omega - \omega_0)|^2 d\omega$$

Now note that since we said T is very flat for a large range of U then we know that the fourier trasnform of T has to be very big at $T = 0$ and thus the fourier transform of T becomes basically a delta function. This gives us finally

$$\boxed{\langle a\tilde{a}\rangle = \frac{e^{-\pi\omega_0}}{1-e^{-2\pi\omega_0}} + \dots} \tag{1}$$

There are similar calculations we can do to find

$$\langle aa^\dagger\rangle = \frac{1}{1-e^{-2\pi\omega_0}} \quad \langle \tilde{a}\tilde{a}^\dagger\rangle = \frac{1}{1-e^{-2\pi\omega_0}} \quad [a,\tilde{a}] = 0 \quad [a,a^\dagger] = [\tilde{a},\tilde{a}^\dagger] = 1 \quad \langle a^\dagger\tilde{a}\rangle = \langle a^\dagger a^\dagger\rangle = 0 \tag{2}$$

Something to note is that in some quantum field theories $\langle i_\omega \tilde{i}_{\omega'}\rangle = e^{-\pi\omega}1 - e^{-2\pi\omega}\delta(\omega-\omega')$. However in this case that is not true. Here we have (TODO how?) $\langle \tilde{a}_{\omega_0} c_{\omega_0'}\rangle \approx 0$ where $\omega_0 \neq \omega_0'$.

In the special case where we have a spacetime with spherical symmetry

$$ds^2 = -dUdV + r_0^2 d\Omega_{d-1}^2 + \dots$$

In this we can derive a analgous form of the modes where

$$a = \frac{r_0^{d-1}}{\sqrt{\pi\omega_0}} \int \partial_U \phi(U, V=0, \Omega)(-U)^{-i\omega_0} T(-U) dU Y_l^*(\Omega) d\Omega \qquad \tilde{a} = \dots$$

Where Y_l are our spherical harmonic functions. These satisfy all the same correlator and commutation relation as we found before.

Now moving to a different topic. We've been writing $\langle\dots\rangle$ for correlators but we need to describe what state we're calculating these correlators for. Say we are in a state $|\psi\rangle$. What we want to show is that $\tilde{a}|\psi\rangle \propto a^\dagger|\psi\rangle$. Thinking about this geometrically this means that $\tilde{a}$ and $a^\dagger$ are parallel to eachother. To prove this consider the decomposition of $\tilde{a}|\psi\rangle$ (TODO: is that actually a complete set?)

$$\tilde{a}|\psi\rangle = c_1 a|\psi\rangle + c_2 a^\dagger|\psi\rangle + |\chi\rangle$$

Where $|\chi\rangle$ is orthogonal to $a|\psi\rangle$ and $a^\dagger|\psi\rangle$. From here we can use the correlators we have in (2) to get $c_1 = 0$. Similarly we can use (1) to get

$$\frac{e^{-\pi\omega_0}}{1-e^{-2\pi\omega_0}} = \langle\psi|a\tilde{a}|\psi\rangle = c_2\langle\psi|aa^\dagger|\psi\rangle + \langle\psi|a|\chi\rangle = \frac{c_2}{1-e^{-2\pi\omega_0}} \Rightarrow c_2 = e^{-\pi\omega_0}$$

Finally we can find $|\chi\rangle$ through the following. From (2) we have $[\tilde{a},\tilde{a}^\dagger] = 1$. This means

$$\langle\tilde{a}\tilde{a}^\dagger\rangle - 1 = \langle\tilde{a}^\dagger\tilde{a}\rangle = \frac{e^{-2\pi\omega_0}}{1-e^{-2\pi\omega_0}}$$

Now with this we have (after $|\dots|^2$ both sides)

$$\frac{e^{-2\pi\omega_0}}{1-e^{-2\pi\omega_0}} = |c_2|^2\langle\psi|aa^\dagger|\psi\rangle + \langle\chi|\chi\rangle + 0 \Rightarrow \langle\chi|\chi\rangle = 0 \Rightarrow |\chi\rangle = 0$$

Where the last term vanishes because we assume χ is orthogonal to $a^\dagger|\psi\rangle$ and $a|\psi\rangle$. After all this we get

$$\boxed{\tilde{a}|\psi\rangle = e^{-\pi\omega_0}a^\dagger|\psi\rangle} \tag{3}$$

Similarly we can also show that

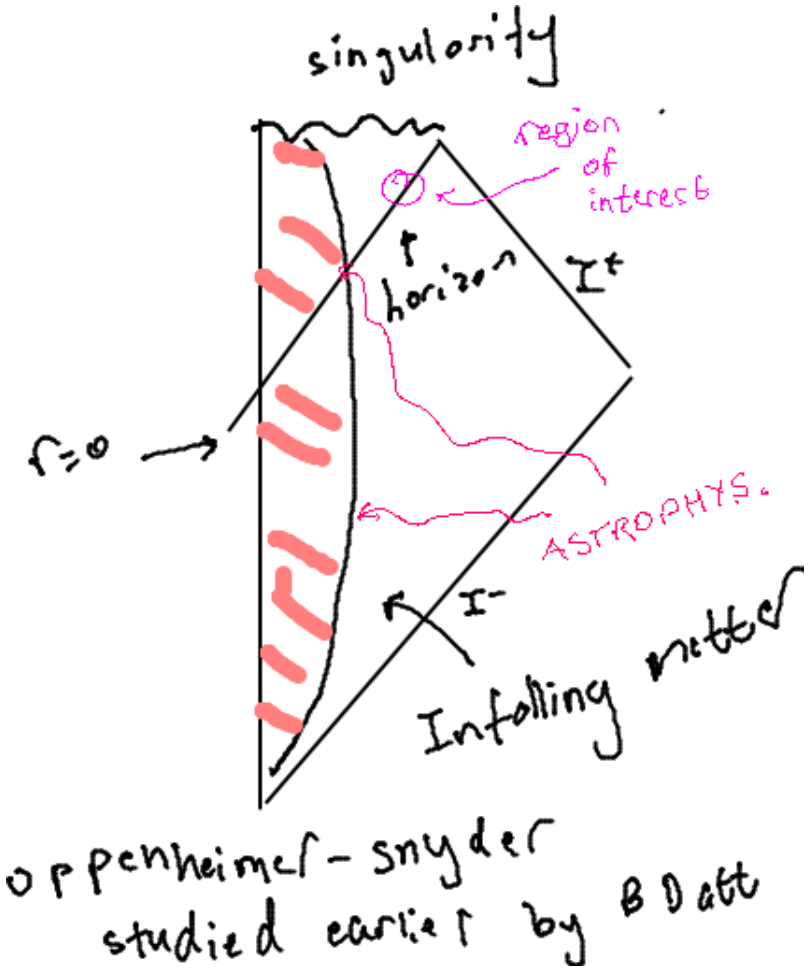
$$\boxed{\tilde{a}^\dagger|\psi\rangle = e^{\pi\omega_0}a|\psi\rangle} \tag{4}$$

Next lecture we'll apply these results to black holes

Lecture 3: Quantum Fields in a Black Hole Background

January 20,2021

Lets start by review Black Holes in flat space.



In the late time limit the metric becomes very simple

$$ds^2 \to -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega_{d-1}^2 \qquad f(r) = 1 - \frac{\mu}{r^{d-2}}$$

Where μ is the mass parameter is related to the mass by

$$\mu = 8\pi^{(2-d)/2}\Gamma(d/2)GM_{\text{real}}/(d-1)$$

The horiizon is when $f(r_h) = 0$ and thus $r_h = \mu^{d-2}$. Lets talk about what we mean by as $t \to \infty$. We mean that $t \gg r_h$ after collapse but $t \ll t_{\text{evap}}$. This is because the collapsing black hole is not valid for when the black hole is evaporating. Also note that the region of interest hast the property that there is a lot of time in that region. It's also useful to go to Tortoise coordinates in order to examine propagating fields

$$dr_* = \frac{dr}{f(r)} \qquad \text{Near } r \to \infty,\ f(r) \to 1 \Rightarrow r_* \to \infty \qquad r \to r_h,\ f(r) \to 2k(r-r_h) \Rightarrow r_* \to \frac{1}{2k}\log[(r-rh)\underline{2k}]$$

Where $k = f'(r_h)/2 =$ surface gravity. The underlined term is a choice of constant. So now we have

$$ds^2 = f(r)[-dt^2 + dr_*^2] + r(r_*)^2 d\Omega^2$$

Something else we should note is that the horizon is not as special as we think. Lets go to Kruskal coordinates

$$U = -\frac{1}{k}e^{k(r_*-t)} \Rightarrow dU = (dt - dr_*)e^{k(r_*-t)} \qquad V = \frac{1}{k}e^{k(r_*+t)} \Rightarrow dV = (dr_* + dt)e^{k(r_*+t)}$$

$U < 0$ outside the horizon. This means we have $dUdV = (dr_*^2 - dt^2)e^{2kr_*}$. However near the horizon we found that the exponential becomes $2k(r - r_h)$. The metric in Kruskal coordinates becomes

$$ds^2 \to -dUdV + r^2 d\Omega_{d-1}^2 \text{ near } r \to r_h$$

Horizon is at $U = 0$ while V remains finite so $t \propto \log(V/U) \to \infty$. The coordinates are basically flat near the horizon. Behind the horizon U becomes positive. For $r < r_h$ we find that $f(r)$ changes negative so t is a spacelike coordinate and r_* is a time coordinate.

We're done reviewing classical black holes so now lets consider the propagation of fields. Consider the field that is minimally coupled

$$\left(\frac{1}{\sqrt{-g}}\partial_\mu g^{\mu\nu}\sqrt{-g}\partial_\nu - m^2\right)\phi = 0$$

In tortoise coordinates $\sqrt{-g} = f(r)r^{d-1}$(sphericalcontribution) and $g^{**} = -g^{tt} = 1/f(r)$. The wave equation becomes

$$\frac{1}{f(r)r^{d-1}}\partial_* r^{d-1}\partial_*\phi = \frac{1}{f(r)}\partial_t^2\phi + \frac{1}{r^2}\Box_\Omega\phi - m^2\phi$$

We can solve the above by noting near the horizion $f(r) \to 0$ so the equation becomes

$$\frac{1}{f(r)}(\partial_*^2\phi - \partial_t^2\phi) = 0$$

This is independent of the angular part, mass, and additional interactions. We can then write

$$\phi \to \int d\omega e^{-i\omega t}\left[A_\omega(\Omega)e^{-i\omega r_*} + B\omega(\Omega)e^{i\omega r_*}\right] + \text{hermitian conjugate}$$

This however isn't the most convenient thing we could do. First let $Y_\ell(\Omega)$ as a spherical harmonics where ℓ is a collective symbol for all the angular quantum numbers. We can choose spherical harmonics as our basis of solutions and have (where we choose f_{in} and f_{out}. These are solutions which we choose as our basis)

(a) $f_{\text{in}}(\omega, \ell, r_*)e^{-i\omega t}Y_l(\Omega)$ where as $r \to r_h$ $f_{\text{in}} \to h_{\omega,\ell}e^{-i\omega r_*}$

(b) $f_{\text{out}}(\omega, \ell, r_*)e^{-i\omega t}Y_l(\Omega)$ where as $r \to r_h$ $f_{\text{in}} \to e^{i\omega r_*} + g_{\omega,\ell}e^{-i\omega r_*}$.

f_{in} has the property that as t increase r_* must decrease. On the other hand f_{out} has the property that as t increase r_* must increase. Both of these happen to keep phase constant. These solutions above are chosen so that the they're orthogonal in the Klein-Gordon norm. This isn't enough to fix those however. We also choose $g_{\omega,\ell}$ so that as $r_* \to \infty$ we have $f_{\text{out}} \to b_{\omega,\ell}r^{(1-d)/2}e^{i\omega r_*}$.

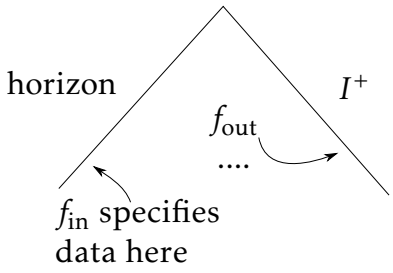


Figure 1: Description of how information in specified in Penrose diagram

f_{in} as $r \to r_h$ looks like $e^{-i\omega(r_*+t)}$. Also remember that $r_* \to -\infty$ at horizon but $t \to \infty$ so $(r_* + t)$ remains finite. The point of this whole song and dance is for

$$\phi = \sum_\ell \int d\omega \Big[A_{\omega,\ell} f^{\text{out}}(\omega,\ell,r_*) + B_{\omega,\ell} f^{\text{in}}(\omega,\ell,r_*)\Big] e^{-i\omega\bar{t}} Y_\ell(\Omega) + \text{hermitian conjugate}$$

A, B are not normalized but are anhilation operators and hermitian conjugate are the creation operator. What happens when we cross the horizion. Behind the horizion we can write a similar expansion

$$\phi = \sum_\ell \int d\omega \Big[\tilde{A}_{\omega,\ell} e^{i\omega t} Y^*_\ell(\Omega) + C_{\omega,\ell} e^{-i\omega t} Y_\ell(\Omega)\Big] \tilde{f}^{\text{out}}_{\omega,\ell}(r_*) + \text{hermitian conjugate}$$

Where as $r \to r_h$ from inside we have $\tilde{f}^{\text{oout}}_{\omega,\ell}(r_*) \to e^{-i\omega r_*}$. Note that we do not go deep inside the horizon. This is because our expansion starts breaking down if we go too far in. By continuity the $C_{\omega,\ell}$ modes have an expansion

$$C_{\omega,\ell} = A_{\omega,\ell} h_{\omega,\ell} + B_{\omega,\ell} g_{\omega,\ell}$$

Also we know that $\tilde{A}_{\omega,\ell}$ are new modes that aren't related to modes outside of the horizon

Lecture 4: Hawking Radiation

January 21,2021

Lets first recap. We took a field in KG equation in BH ST and said this field has some form that look like

$$\phi = \sum_\ell \int d\omega \Big[A_{\omega,\ell} f^{\omega t}(\omega,\ell,r_*) Y_\ell(\Omega) + B_{\omega,\ell} f^{\text{in}}(\omega,\ell,r_*) Y_\ell(\Omega)\Big] e^{-i\omega t} + \text{hermitian conjugate}$$

We could in principle solve for $f^{\text{in,out}}$ but we could just look at the near horizion behavior

$$f^{\text{out}} \to e^{i\omega r_*} + g_{\omega,\ell} e^{-i\omega r_*} \qquad f^{\text{in}} \to h_{\omega,\ell} e^{-i\omega r_*}$$

We can also look at the field inside the black hole and find

$$\phi = \sum d\omega \ldots$$

And again looking at the near horizion behavior

$$\tilde{f}_{\text{out}}(\omega, \ell, r_*) \to e^{-i\omega r_*}$$

And asserting continuity we get

$$C_{\omega,\ell} = A_{\omega,\ell} h_{\omega,\ell} + B_{\omega,\ell} g_{\omega,\ell}$$

Near the horizion the metric looks like

$$ds^2 \to -dU dV + r^2 d\Omega_{d-1}^2$$

This should remind us of near horizion modes that we saw in the first two lectures. In our older notes we have

$$a_{nh} = \frac{r_0^{d-1}}{\sqrt{\pi\omega_0}} \int \partial_U \phi(U, V = 0, \Omega)(-U)^{-i\omega_0} T(-U) Y_l^*(\Omega) dU d\Omega$$

$$\tilde{a}_{nh} = \frac{r_0^{d-1}}{\sqrt{\pi\omega_0}} \int \partial_U \phi(U, V = -\epsilon, \Omega) U^{i\omega_0} T(U) Y_\ell(\Omega) d\Omega$$

Since we have the field expansion we could in principle plug in ϕ and evaluate the integral very carefully. But in fact we don't have to do the integral becuase near the horizion what does the field look like?

$$\phi \approx \sum_{\omega,\ell} e^{-i\omega t} A_{\omega,\ell} (e^{i\omega r_*} + g_{\omega,\ell} e^{-i\omega r_*}) + (\text{terms with } B_{\omega,\ell}) \approx \sum_{\omega,\ell} A_{\omega,\ell} (U^{i\omega/k} + g_{\omega,\ell} V^{-i\omega/k}) Y_\ell(\Omega)$$

Where the above is up to constants. This is because $U \propto e^{k(r_*-t)}$. The $B_{\omega,\ell}$ multiply a different set of terms $e^{-i\omega t} e^{-i\omega r_*} Y_\ell(\Omega)$ which is approximately $B_{\omega,\ell} V^{-i\omega/k}$.

Now looking at a_{nh} and $\tilde{a}_{nh}$ which are basically fourier transforms in $\log U$ we see that the integral picks up the modes with $\omega = \omega_0$ and $-\omega_0$ respectivley. So we can say that

$$a_{\omega,\ell} \approx \int A_{\omega',\ell} q(\omega', \omega) d\omega' +$$

In fact we will find that

$$a_{nh} = a_{\omega_0 k, \ell}$$

Since $q(\omega', \omega)$ picks up modes from near $A_{\omega_0 k}$. We only need that

(a) $q(\omega', \omega)$ is sharply peaked around $\omega' = \omega$.

(b) after smearing $[a_{\omega,\ell}, a_{\omega,\ell}^\dagger] = 1$.

The summary is that the near horizion modes become slightly smeared global modes centered around frequency $\omega_0 k$. We can now immediatley compute the two point function of these modes. Using $\omega_0 = \omega/k$ lets define $\beta = 2\pi/k$.

$$\boxed{\langle a_{\omega,\ell} a_{\omega,\ell}^\dagger \rangle = \frac{1}{1 - e^{-2\pi\omega_0}} = \frac{1}{1 - e^{-\beta\omega}}}$$

And look at that! We have *Hawking radiation*. Lets look at bit more at the $\tilde{a}$ modes

$$\tilde{a}_{nh} \approx \tilde{a}_{\omega_0 k,\ell} = \text{ smeared version of } \tilde{A}_{\omega,\ell}$$

We're actually going to put this aside for now if we're only talking about the exterior of the BH. To get thermal occupancy for $a_{\omega,\ell}$ modes we assumed

(a) Horizion was smooth. this is what fixed the occupancy of the near horizion modes

(b) There was a late time emergent $t \to t + \delta t$ isometry. this is what allows us to relate the near horizion (nh) modes to global modes.

The short distance properties correspond to the smoothness of horizion and the long distance property corresponds to the late time emergent isometry. Something to note is that $a_{\omega,\ell}$ is proporitional to $A_{\omega,\ell}$ which is outgoing to I^+. The stress tensor far away also depends on the properties of f_{out}. Also our derivation does not constrain $B_{\omega,ell}$ since it doesn't appear in the near horiizon mode. So we can choose some things for $B_{\omega,\ell}$

(a) We could put $B_{\omega,\ell}$ in a vacuum $\to$ Unruh state

(b) populate $B_{\omega,\ell}$ thermally $\to$ Kruskal state

(c) ...

The advantage of this derivation is that it is clear how we can correct this theory by finding the higher order terms. This is an advantage to the rindler anlogy or the ray tracing arguemnt where finding higher order terms is less well defined.

We can also consider AdS Black Holes. In asympotitically global AdS

$$ds^2 \overset{r\to\infty}{\longrightarrow} -(1+r^2)dt^2 + \frac{dr^2}{1+r^2} + r^2 d\Omega_{d-1}^2$$

The AdS radius is 1. So now we want to consider black holes that form from the collapse of matter.

$$ds^2 \overset{t\gg 1}{\longrightarrow} -f(r)dt^2 + \frac{dr^2}{f(r)^2} + r^2 d\Omega_{d-1}^2$$

Which is identitical with what we had above except $f(r) = 1 + r^2 - \mu/r^{d-1}$ and μ has the same relationship with M. The horizion once again is at $f(r_h) = 0$ meaning that we are interested in a late infalling observer "take enough" that effect of infalling matter have died out.

Behind the horizion U becomes positive for $r < r_h$ $f(r)$ changes sign so for $r < r_h$ t is a psace-likecoordinate r_* is a time coordinate

$$T = V + U \qquad x = V - U$$

We are considering fields which are normalised

$$\phi \to 0 \qquad \text{as} \qquad r \to \infty$$

System undisturbed by external sources evolves autonomously. For fields of mass m

$$\phi \to \frac{1}{r^D}$$

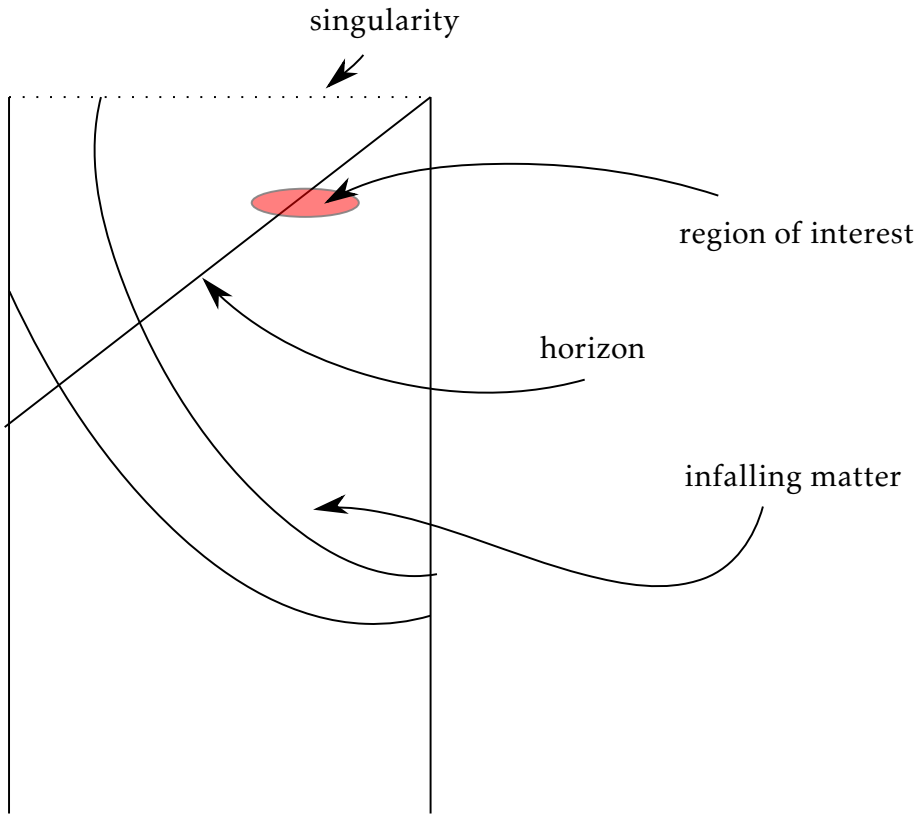


Figure 2: Penrose diagram for AdS black hole

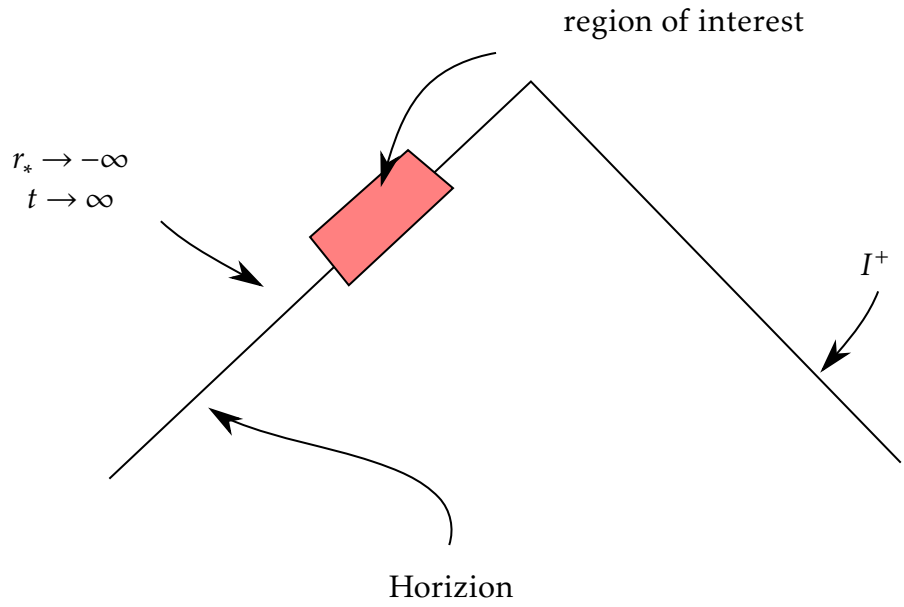


Figure 3: Zooming in on penrose diagram for late time ignore infalling matter

Where $D = d/2 + \sqrt{d^2/U + m^2}$ is also dimension of the dual operator to ϕ. Like before we'll consider tortoise coordinates

$$dr_* = \frac{dr}{f(r)}$$

Near the horizion once again $r_* \to -\infty$ and near infinity $r_* \to$ const. We can once again write down a expansion

$$\phi = \sum_\ell \int d\omega A_{\omega,\ell} f^{\text{st}}(\omega, \ell, r_*) e^{-i\omega t} Y_\ell(\Omega) + \text{hermitian conjugate}$$

Where f^{st} is a standing wave solution. We will also demand near $r_* \to -\infty$

$$f_{\text{st}} \to e^{i\omega r_*} + \underline{e^{-i\delta_{\omega,\ell}}} e^{-i\omega r_*}$$

Where the underlined term is a phase. We need to satisfy

$$r^{d-1}[\phi, \dot{\phi}] = i\delta(r_* - r'_*)\hat{\delta}(\Omega, \Omega')$$

Near the horizion

$$[\phi, \dot{\phi}] \approx \int d\omega \, [A_{\omega,\ell}, A^\dagger_{\omega',\ell}] \left(e^{i\omega r_* - \omega' r'_*} + g_{\omega,\ell} g^*_{\omega',\ell'} e^{-i\omega r_* + \omega' r'_*} \right) + (\text{linear in } g_{\omega,\ell}) e^{i\omega r_* + \omega' r'_*}$$

(TODO what?). He says if you didn't follow you can work it out yourself. In the same way behind the horizion we can write down a expansion again

$$\phi = \int d\omega [A_{\omega,\ell} e^{-i\delta\omega\ell} e^{-\omega t} Y_\ell(\Omega) + \tilde{A}_{\omega,\ell} e^{i\omega t} Y^*_\ell(\Omega)] \tilde{f}_{\text{st}}(\omega, \ell, r_*) + \text{hermition conjugate}$$

Where $\tilde{f}_{\text{st}}(\omega, \ell, r_*) \overset{r \to r_h^-}{\longrightarrow} e^{-i\omega r_*}$. As in flat space we introduce near horizion modes and once again

$$a_{nh} = a_{\omega_0 k,\ell} \quad \tilde{a}_{nh} = \tilde{a}_{\omega_0 k,\ell} \quad k = f'(r_h)/2 \quad a_{\omega,\ell} = \int A_{\omega,\ell} a_\ell(\omega, \omega') d\omega' \quad \tilde{a}_{\omega,\ell} = \int \tilde{A}_{\omega,\ell} \tilde{q}_\ell(\omega, \omega') d\omega'$$

And this leads us

$$\langle a_{\omega,\ell} a^\dagger_{\omega,\ell} \rangle = \frac{1}{1 - e^{-\beta\omega}}$$

This is the same as before but here we don't have any flux at infinity. Even ghough $A_{\omega,\ell}$ smeared is thermally occupied there is no flux at inifnity. We obtain a black holes with a thermal atmosphere around it.

$$f^{\text{st}} \overset{r \to \infty}{\longrightarrow} \frac{\sqrt{G_{\omega,\ell}}}{r^D}$$

And the dual operator $O_{\omega,\ell} = \sqrt{G_{\omega,\ell}} A_{\omega,\ell}$.

Lecture 5: Hawking's Original Paradox

January 27, 2021

Today we're going to formulate the paradox as Hawking stated it. Last time we derived a formula that gives the Hawking temperature

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

Before this was derived people knew that

$$dM = \frac{kdA}{8\pi} + \Omega dJ + \phi dQ \qquad dA \geq 0$$

Where A is the area, Ω is the angular velocity of the horizion, J is angular momentum, ϕ is potential of the horizion and Q is chrage. This comes from analyzing classical processes around the particle. E.g. you throw some particle of charge and mass M into a Black Holes and then see what happens. We also know that the area of a black hole is alwyas increasing. This in fact looks a lot like thermodynamics

$$dU = TdS + \text{work terms} \qquad dS \geq 0$$

Feynman likes to say that the same equations have the same solutions. For example LC circuits and springs. They're different physical systems but they can solved in the same way. The finding of the black hole temperature elavated the connection between thermodynamics and blackhole from a formal one to a physcial one. The entropy of the black hole is

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

Something we should note is that black holes have a lot of entropy compared to other objects in the universe.

Lets look at Hawking original paper. In it he argued that if you start with some inital state, the final state is a mixed state because you end up losing some information. The first point that Hawking made is that Hawking had argued in a previous paper that black holes steadily create and emit particle with a thermal spectrum. This radiation should take away energy so the black hole must lose mass and eventually evaporate. The details can seen here. We worked out

$$\frac{dM}{dt} = -cAT^{d+1}$$

To work out the constant c you need to work out the greybody factors. This almost follows from dimensional analysis. In flat space we found that

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

Putting these proportionalies we get

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

The interesting part of this is that in a time $\propto r_h^d \propto \frac{A}{T}$ we have that $r_h \to 0$. What this tells us is if there's a black hole in a universe where nothing is trying to fight against hawking radiation

the black hole will evaporate in a time propotional to the power of the inital radius.

Hawking's next point is that if you take a black hole you need data on the horizion and I^+ to determine I^-. This is different from minkowski where you only need information of I^+ to know the inital state of I^-. What this is saying is that you lose information. From here hawking notes that this is also true quantum mechanically. He does this by computing the occupancy of something. In the first we lectures we had fields as

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

We then define $a_{\omega,\ell}$ which are smeared versions of $A_{\omega,\ell}$ and we got from these

$$\left\langle a_{\omega,\ell} a^\dagger_{\omega,\ell} \right\rangle = \frac{1}{1-e^{\beta\omega}}$$

We then found that there was some flux at infinity. Hawking's point is that this flux you compute does not depend on the state of the $B_{\omega,\ell}$ modes. Not only does it not depend on the B modes but you can compute the state of the $A_{\omega,\ell}$ modes. First we have

$$N_\omega = a^\dagger_\omega a_\omega \qquad \langle N_\omega \rangle = \frac{1}{e^{\beta\omega}-1} \qquad \left\langle N^2_\omega \right\rangle = \frac{1+e^{\beta\omega}}{(e^{\beta\omega}-1)^2}$$

The derive this you can of the following. First we know that

$$(\tilde{a}_{\omega,\ell} - a^\dagger_{\omega,\ell} e^{-\beta\frac{\omega}{2}}) |\psi\rangle = 0$$

We find that the kind of state that satisfies this equation is

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

Then we can find correlators of any polynomial can be worked by writing out a density matrix

$$\left\langle N^q_{\omega,\ell} \right\rangle = \text{tr}(\rho_{\omega,\ell} N^q_{\omega,\ell}) \qquad \rho_{\omega,\ell} = \frac{1}{1-e^{-\beta\omega}} e^{-\beta\omega N_{\omega,\ell}}$$

(something about therma distribution here.) The important part about of all of this is the fact that this is not a pure state. And so we find is that whatever state we start with we have on I^- we end up with a thermal state on I^+ (characterized by the thermal distribtuin.) This is in paradox with the unitarity of quantum mechanics.

Quantum mechniacal evolution takes a state and evolves ith with a unitary operator. So lets say we have some density matrix. AFter time evolution we have

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

So the fact that you can start in a pure state and go to a mixed state is a sign of the fact that we've lost information along the way. This is because we can think of a mixed state as starting with a pure state and throwing away some information. This is our first paradox! Hawking then gives some intuiion for why this might happen. There is some hidden surface behind the

horizion so we should adopt a principle of ignorance. The point of this is say we have some state on I^- as a vector in some hilbert space H_1

$$\psi_A \in H_1$$

And then we have $\psi_B^H \in H_2$ which is a state at the horizion or interor hiddens urface. And finally we have $\psi^+ \in H_3$ which is a state on I^+. Now we assume locality. So we have some S matrix

$$S_{ABC}\psi_A^-\psi_B^H\psi_C^+$$

Is the amplitude to go from $\psi^- \to \psi^H \otimes \psi^+$. So if you sum over all possible states of the hidden surface

$$\sum_{\psi^H} S_{ABC}\psi_A^-\psi_B^H\psi_c^+ S^*_{\tilde{A}\tilde{B}\tilde{C}}\psi_{\tilde{A}}^{-*}\psi_{\tilde{B}}^{H*}\psi_{\tilde{C}}^{+*} = \underbrace{\sum_B S_{ABC}S^*_{\tilde{A}B\tilde{C}}\psi_A^-\psi_{\tilde{A}}^{-*}}_{\rho_{c\tilde{c}}}\psi_C^+\psi_{\tilde{C}}^{+*}$$

So we get a superscattering operator (some generalzation of the unitary evolution operator) that takes pure states to mixed states. There is a nice diagrammitic way to summarize the argument which is in Figure 4. Information requires energy to be stored. Some system has a lot

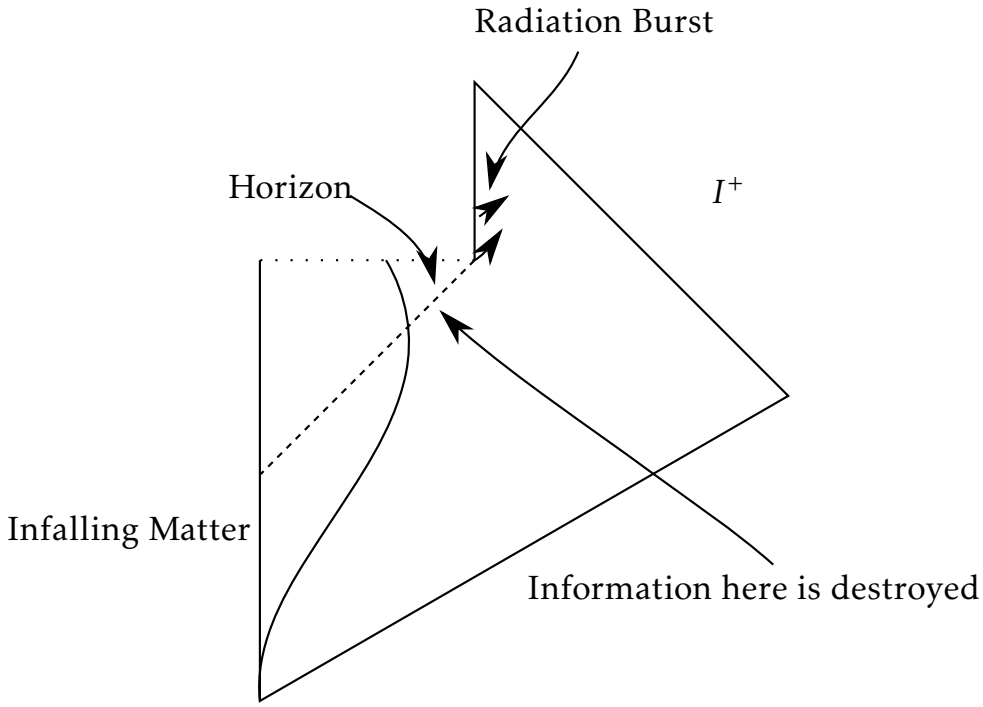


Figure 4: Diagrammitic summary of Hawking's argument of information paradox

of microstates and the microstate gives the information. This is how hard-disks store information. This extends to quantum information. And so the fact taht you have a large number of microstates means you need a lot of energy. The final radiation burst when the black hole evaporates doesn't have enough energy to encode the information so that radiation burst shouldn't have the lost information. So remnants are not viable candiate to resolve information paradox.

Elaboration on energy needing energy: you can store information in microstates. The number of microstates is e^S. The fact that information requires energy is the fact that S at some energy that is bounded.

So there are two elements of Hawking's argument that are crucial

(a) Computation of the thermal density matrix. This is the key computation

(b) Intution from the casual structure of space time.

What we'll ask tomorrow is it really true that computation shows mixed state? And also we'll ask is that is the intuiton about the casual strucutre valid at the accuracy required.

Lecture 6: Mixed and Pure States

January 28, 2021

We'll start by asking a question that seems unrelated to the information paradox: how close are pure states to mixed states? Lets start by discussing the difference between pure states and mixed states. In quantum mechanics our observations are inherantly probailistic. Consider schrodinger's cat. The state of a cat is

$$|\psi\rangle = a_1 |\text{dead}\rangle + a_2 |\text{alive}\rangle$$

And the probability of finding the cat dead is $|a_1|^2$ and alive $|a_2|^2$. This is a pure state. We could similarly have a classical mixture where we have the same probability of having a dead or alive cat. These two states (quatnum state and clasical mixture) can be distinguished. Well the EV of an operator A can be found with

$$\langle\psi|A|\psi\rangle = |a_1|^2 \langle 1|A|1\rangle + |a_2|^2 \langle 2|A|2\rangle + a_2^* a_1 \langle 2|A|1\rangle + a_1^* a_2 \langle 1|A|2\rangle$$

Wheras for the classical mixture we have

$$\langle A\rangle = |a_1|^2 \langle 1|A|1\rangle + |a_2|^2 \langle 2|A|2\rangle$$

So the difference is in the cross terms. The classical mixture is usually represetned by a density matrix

$$p = |a_1|^2 |1\rangle\langle 1| + |a_2|^2 |2\rangle\langle 2|$$

And the way you find an expectation value is compute the $\text{tr}(\rho A)$. Of course you could also compute a density matrix for a pure state

$$p_{\text{pure}} = |\psi\rangle\langle\psi|$$

And notice that this pure state density matrix $p_{\text{pure}}^2 = p_{\text{pure}}$. This property differentiates pure states from mixed states. Now we want to ask how close are pure and mixed states. Lets add a little bit of physics to our situation. Say we have a system with discrete energly level. The mean energy is E_0. We want to consider energy in the region

$$E_0 - \Delta \leq E_i \leq E_0 + \Delta$$

Where Δ is the spread. We could consider a state

$$|\psi\rangle = \sum_{i=1}^{w} a_i |E_i\rangle$$

Where w is the number of eignetstates in the interval $[E_0 - \Delta, E_0 + \Delta]$. This w is nothing but e^S, the exponential of the entropy. However in stat mech we don't usually consider pure states like this but mixed states with a density matrix

$$\rho_{\text{micro}} = \frac{1}{w} \sum_i |E_i\rangle\langle E_i|$$

This density matrix is the microcanonical density matrix and what this is doing is saying is that you have equal probability to be in one of the corresponding eigenstate. Now the question we wnat to ask is clear. How close is

$$\sum a_i |E_i\rangle \qquad \text{To} \qquad \rho_{\text{micro}}$$

Well in a Hilbert space these two things are orthogonal(there exists a observable where the EV of the observable with $|\psi\rangle$ is 0 and with the density is 1). But if we take a "typical state" $|\psi\rangle$ we find that these are "extremley close" to ρ_{micro}. Lets define typical state an extremley close. We want a physical notion of cloesness. Lets say we have some observer which can make physical observations. From the POV of physical observations how easy or difficult is it to distinguish the two? Physical observations always have to do with probabilities of different outcomes. And these probabilities come from p = projector and trying to find

$$\langle\psi|p|\psi\rangle$$

The physical significance of this statement is consider some observable $\mathcal{O}$ with spectracl composition $\mathcal{O} = \lambda p + \sum_i \lambda_i p_i$. Upone measuring $\mathcal{O}$ what is the probability of getting λ. The expectation value has information on the average value but also λ. Lets say we take

$$\mathcal{O} \to 1000\mathcal{O} \Rightarrow \lambda \to 1000\lambda$$

In this rescaling the probabilities are invariant. So if we have a typical pure state and some projector p how does $\langle\psi|p|\psi\rangle$ compare with $\text{tr}(\rho P)$ where p is some micro mixed state. This is a more physical way of defining similarlity. Now we'll finally define a typically state. Recall some properties of the hilbert space

$$\sum a_i |E_i\rangle \qquad \sum |a_i|^2 = 1$$

One way to think about this state is to think about it as a sphere. The hilbert space is a sphere in very high dimensions and we're just picking points on the surface of the sphere. Note we still have som ambiguity in phase ($|\psi\rangle \to E^{i\phi}|\psi\rangle$) The question now is lets say we have the surface of the sphere if we pick a random point on the sphere and ask how close is

$$\langle\psi|p|\psi\rangle \qquad \text{to} \qquad \text{tr}(\rho P)$$

We'll do this by defining a volume measure on the sphere. What is a natural probability to shove onto a sphere. Well first we define a measure on a hilbert space

$$d\mu_\psi = \frac{1}{V}\delta\Big(\sum |a_i|^2 - 1\Big)\pi d^2 a_i \Rightarrow \int d\mu_\psi = 1$$

The $d^2 a_i$ is for complex numbers, δ function is for normalizaiton and $\frac{1}{V}$ is another normalziation factor. Now lets define

$$\delta = \langle\psi|p|\psi\rangle - \mathrm{tr}(\rho p)$$

We want to compute $\langle\delta\rangle_M$ and $\left\langle\delta^2\right\rangle_M$. Now lets do some computations

$$\langle\psi|p|\psi\rangle = \sum_i |a_i|^2 \langle E_i|p|E_i\rangle + \sum_{i\neq j} a_i a_j^* \left\langle E_j|p|E_i\right\rangle$$

We need to average this over all possible pure states using the measure we defined above $\int \langle\psi|p|\psi\rangle d\mu_\psi$. To compute this we need $\int |a_i|^2 d\mu_\psi$ and $\int a_i a_j^* d\mu_\psi$. This is easy to work out because $\left\langle |a_i|^2\right\rangle$ cannot depend on i. This means we can just as easily compute

$$\frac{1}{w}\left\langle\sum |a_i|^2\right\rangle_\mu = \left\langle |a_i|^2\right\rangle_\mu = \frac{1}{w}$$

Similarly

$$\left\langle a_i a_j^*\right\rangle_\mu = \delta_{ij}$$

There is no correlcation betwen a_i and a_j because we're picking things randomly. All of this gives us

$$\boxed{\int \langle\psi|p|\psi\rangle d\mu_\psi = \frac{1}{w}\sum_i \langle E_i|p|E_i\rangle = \mathrm{tr}(\rho p) \Rightarrow \langle\delta\rangle = 0}$$

This does not end the story. We need to now compute $\left\langle\delta^2\right\rangle$. It's a bit more involved but we can do it.

$$\begin{aligned}\left\langle\delta^2\right\rangle &= \int [\langle\psi|p|psi\rangle - \mathrm{tr}(p\rho)]^2 d\mu \\ &= \int \sum_{i,j,k,l} \left\langle E_j|p|E_i\right\rangle (a_i a_j^* - \delta ij/2)\langle E_l|p|E_k\rangle (a_k a_l^* - \delta_{kl}/w) d\mu\end{aligned}$$

We need $\int a_i a_j^* a_k a_l^* d\mu = \dfrac{\delta_{ij}\delta_{kl} + \delta_{il}\delta_{jk}}{w(w+1)}$

If we're careful we can notice that

$$\delta^2 \leq \sum_{i,j} \frac{1}{w(w+1)} \left\langle E_i|p|E_j\right\rangle \left\langle E_j|p|E_i\right\rangle \leq \sum_i \frac{1}{w(w+1)} \left\langle E_i|p^2|E_i\right\rangle \leq \frac{1}{w+1}$$

The second relation comes from notion that E_j is almost but not a complete basis. The above result tells us

$$\left\langle\delta^2\right\rangle_\mu \leq \frac{1}{w+1}$$

This is significant since $w \propto e^S$ meaning that average deviation are of size $e^{-\frac{S}{2}}$. So pure states not only on average look like mixed states but also the deviation is small. To summarize

(a) For a given observable for most pure states look exponentially to the maxiamlly mixed state

(b) Volume of "atypical state"(a state where δ is large) is exponentially small. (this results comes from $\left\langle \delta^2 \right\rangle_\mu \propto \frac{1}{e^S}$

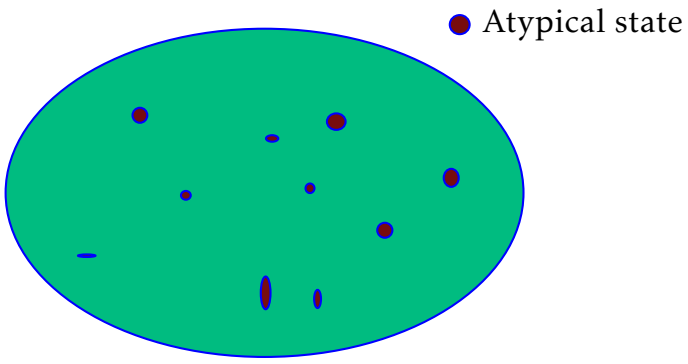


Figure 5: Illustration of the Hilbert state and atypical state

Now lets talk about things that coudl cause confusion

(a) basis vs typical states: We have n qubits which can be spanned by basis

$$|00\ldots0\rangle$$
$$|10\ldots0\rangle$$
$$|01\ldots0\rangle$$

We have σ_3^i that has definite values in all these states. Basis states occupy 0 volume.

(b) ETH = Eigenstate Thermlization Hypothesis: Our result is kinematical. We never used ETH.

(c) Entanglement entropy. $S = 0$ for pure state and $S = \ln W$ for mixed states. You might say that engatnelment entropy differentiates pure state and mixed state. But the entanglement entropy isa fine grained observable and the expecation value of observavble. $S \neq \langle A \rangle$. The reason for this simple. One can always find a basis of states for which $S = 0$. If S was the expectation value for an operator then it's the same for all basis. Thus S is not an expectation value of any operator. However S is a function of correlation functions.

So how does this all apply to hawking original paradox.

$$\left\langle a_{\omega,\ell} a^\dagger_{\omega,\ell} \right\rangle = \frac{1}{1-e^{-\beta\omega}} \qquad \left\langle a_{\omega,\ell} a^\dagger_{\omega',\ell} \right\rangle = 0 \text{ for } \omega \neq \omega'$$

This lead to the conclusion that final state was mixed. However if the had found

$$\left\langle a_{\omega,\ell} a^\dagger_{\omega,\ell} \right\rangle = \frac{1}{1-e^{-\beta\omega}} + O\left(e^{-\frac{S}{2}}\right) \qquad \left\langle a_{\omega,\ell} a^\dagger_{\omega',\ell} \right\rangle = O\left(e^{-\frac{S}{2}}\right) \text{ for } \omega \neq \omega'$$

(The results by hawking aren't exact) then we would have found that perfectly consistent with pure states. What we're saying is that Hawking's original paradox is not a paradox. In fact it would have been weird if we hadn't found a mixed state. In fact the accuracy of our computation $\ll e^{-S/2}$. Consider the following. What if we formed a b.h. in a micro ensemble or in canonical ensemble. Observables in these two are different by an amount we can compute $\frac{1}{\sqrt{S}} \gg e^{-\frac{S}{2}}$. The fact that simple correlators look thermal does not imply that the final state is thermal. As we have just shown pure states and thermal states can look expoentially close to eachother. The punchline is this

Hawking's original coputation is not precise enough to be a paradox

There are some things we have not answered here. We can refine the paradox. Also what about the "principle of ignorance" that came from looking at the causal structure of spacetime.

Lecture 7: Trying to Localize Operators behind the Horizon

February 03, 2021

There were two parts to Hawking's original paradox

(a) concrete computation

$$\left\langle a_{\omega,\ell}, a^{\dagger}_{\omega,\ell} \right\rangle = \frac{1}{1-e^{-\beta\omega}}$$

(b) Some intuitional: Hawking looked at the casual structure of the black hole geometry. This is Figure 4. So because the causal structure is different we clearly don't have information on the insdide. This is what gives us the "principle of ignorance" where the observer doesn't know what goes on inside the black hole.

Last time we talked about (a) with stat mech. If $|\psi\rangle$ is a typical state in a large hilbert space from a given energy band

$$\left| \langle \psi | A | \psi \rangle - \frac{1}{e^S} \mathrm{tr}(A) \right| \propto e^{-\frac{s}{2}}$$

Basically mixed a pure states are very close. One thing Suvrat wanted to emphesize last time is that black hole has has a natural perturbative parameter

$$T \propto \frac{1}{r_H} \qquad G_N T^{d-1} = \text{natural perturbative parameter} \propto \frac{1}{S}$$

No one has worked out final state entropy non-perturbativley and this is the only way we could transform hawking's original paradox to an actual paradox.

Someone could ask what about the "principle of ignorance." First lets try to frame this more precisley. An observer outside the blackhole doesn't have "no information" about a black hole. There is a "no hair" theorem. A blackhole is characterized by mass, angular momentum, and charge. A observer can observe these things. However this information isn't enough to get all the information about the inital state. We should emphesize that the no hair theorem is a *classical result*. There is no analagous quantum result. This is not some technical point about lack of proof. This is a substanstive point about the difference between quantum and classical. We

will claim that we shall never expect a quantum no-hair theorem.

In classical physics, why do we always know the mass of a black hole? It's because classically, gravity obeys a gauss's law (aparently this guy learned this in high school???) We have an expresssion

$$H = \frac{1}{16}\int (\partial^i \underbrace{h_{ij}}_{\text{deviation from flat metric}} - \partial_j h_{ii}) n^j d^2 s$$

We'd also like to expect something like this to will work in QM as well. Lets try to see this

$$|\psi\rangle = \sum_i a_i |E_i\rangle$$

Note that a blackhole formed by collappse and evaporates it cannot be an energy eigenstate since energy eigenstates don't evolve. Thus it must be a superposition of distinct energies(these energy eigenstates are wrt Hamiltonian of some imaginary theory of quantum gravity.) From here we can measure

$$\langle\psi|H|\psi\rangle = \sum |a_i|^2 E_i$$

Quantum mechanically we can measure more than the value of $\langle H\rangle$ but can also measure other observables like

$$\left\langle\psi|H^2|\psi\right\rangle = \sum_i |a_i|^2 E_i^2 \neq \langle\psi|H|\psi\rangle^2$$

We do expect spreads in energy. We also have information about

$$p(E, E+\Delta) = \sum_{E<E_i<E+\Delta} |a_i|^2$$

There is no classical analouge of this. Quantum mechanically we have more information. But this is not the end of the story. Lets say $\mathcal{O}$ is someother observable without a conserved charge. Like $\mathcal{O} \to \phi$ as $r \to \infty$. We can ask about correlators

$$\langle H\mathcal{O}\rangle \neq \langle H\rangle\langle\mathcal{O}\rangle$$

Lets say someone was trying to prove a theorem(a quantum version of no-hairtheroem)

$$\langle H\mathcal{O}\rangle \to \text{universal value at late time}$$

Is this accurate up to $e^{-\frac{S}{2}}$ corrections. These exponentially small corrections is where the information is. So quantum mechanically the obstacles to no-hair theorem are much more numerous than in classical mechanics. This is why there is no quantum no hair theorem. The spirit of the no-hair theroem is that the geometry settles down. So proving $\langle H\mathcal{O}\rangle$ would be proving no-hair in spirit.

Lets look at this another way. Recall Figure 4 Lets say we draw some slice through the horizon. In local QFT it's possible to change something inside the horizon without changing anything outside the horizion. Now lets say we want to prove there is a principle of ingnraocne. Then the person had to say "hey there's two kinds of states in the black hole" and that no observer outside the horizon can know if the field is excted or not inside the horizon. For the principle

of ignorance too hold there has to exists a unitary U such that U commutes with all operators outside but changes the inside.

$$\left\langle \psi | U^\dagger \mathcal{O}(\text{out}) U | \psi \right\rangle = \langle \psi | \mathcal{O}(\text{out}) | \psi \rangle$$

So it must be the case that $U^\dagger H U = H$. So U must commute with the hamiltonian. But know we are in trouble because if U commutes with the Hamiltonian, U cannot be localized to the black hole interior. And if it is not purely localized insie the interior then it must change some observable outside the horizion. This sometimes goes under the name "there are no gauge invariant quantum operators in quantum gravity." To summarize. If someone wanted to make hawking's princile of ignoracne precise then there would have to be at least one unitary U so that this unitary U commmutes with all observables outside but change something inside. But this would mean that U commutes with the Hamiltonian. In this case then U would have zero energy and thus could not be localized. If it is not localized in the interior then it must fail to commute with some operator on the outside.

Lets look at yet another argument. In any gauge theory lets say we wanted to put a charge somewhere. An operator with charge isn't gauge invaraint. We neeed to dress this operator up with a wilson line to infinity (or wilson loops) to make this gauge invariant. In graivty the "Wilson line" run to infinity. There's another way to say this. Lets say we want to act with some operator and create an excitation insie the black hole.

$$\phi(x) | \psi \rangle$$

You might have thought that ϕ is a scalar field and thus is gauge invarint. However if you consider

$$x^\mu \to x^\mu + \epsilon^\mu \Rightarrow \phi \to \phi + \partial_\mu \epsilon^\mu$$

How do you make this difeomrophism invariant? well we could define the observable relationally to infinity. So you can define a notion of "dressing" in gravity by starting at the boundry of the metric, drop some geodesic, and them measure the operator at the end of the geodesic. So in gravity there is no such thing as a localized gauge invariant operator. This is actually a well known result that there is no such thing as local gauge-invariant operator in quantum-gravity. Mathematically we could also just say "we fix a gauge."

So in this lecture we have gone over three intuitional arguments for why we can't change the interior without knowing outside the horizon. These refute the "prinicple of ignorance" that Hawking argued for. The error in Hawking's argument we found are

(a) Computation of low-pt correlatiors are insuffiecnt to conlcude that final states are thermal

(b) Intuition of principle of ignorance is very sublte.

So both of these tell you that we have information loss have been refuted. Next time we'll talk about a more formidable paradox. Some paradoxes by Mathur and AMPS firewall/fuzzball paradox

Lecture 8: Some Results from Quantum Information

February 04, 2021

Today we're going to go on a detour. We're going to talk about some simple techniques and questions from quantum informtion.

The argument of Mathur/AMPS:

(a) if you have a smooth horizon then the modes inside and outside the horizon must be entangled if the horizon is smooth. We encountered this result in some sense when we looked at correlators accross the null surface.

(b) Typicality requires entangleemnt between modes that are close to the horizon (near-horizon modes) and far-away modes at late times.

(c) These two requirements are inconsistent because entanglement in monogamous.

At this point we're not in a place to make this paradox precise. Lets first talk about entanglement. He wants to emphasize that these arguments look at the inside (makes quantitative reference to the interior.)

So what are the original paradoxes used entanglement entropy. We'll use Bell corrleators and their refinement CHSH correlators to phrase these paradoxes, monogomy, and entanglement. These correlators are well defined in a theory of gravity.

Definition 1: (CHSH Correlators) Lets say we have two distinct systems A_1, A_2, B_1, B_2. Now lets say each observable takes values $\in [-1, 1]$. First lets look at the classical case. Consider the situtaiton where A and B two coins.

$$A_1 = \begin{cases} +1 & \text{heads} \\ -1 & \text{tails} \end{cases} \qquad A_2 = \begin{cases} +1 & \text{dime} \\ -1 & \text{quarter} \end{cases}$$

And same idea for B. We're not going to put any constraints on the probabilty distribution. The point is to define a joint observable, the CSSH observable

$$C_{AB} = A_1(B_1 + B_2) + A_2(B_1 - B_2)$$

In the classical case it should be clear that $|C_{AB}| < 2 \Rightarrow |\langle C_{AB}\rangle| \leq 2$. In the quantum mechanical case there is something case Tsirelson's bound which says

$$|\langle C_{AB}\rangle| \leq 2\sqrt{2}$$

In Q.M we again have

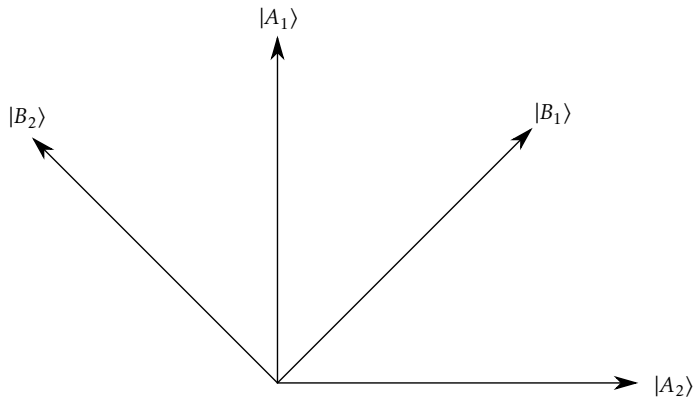
$$C_{AB} = A_1(B_1 + B_2) + A_2(B_1 - B_2) \Rightarrow |A_i|, |B_i| \leq 1$$

The statement that the operators probe different systems is that $[A_i, B_j] = 0$. So how does this Tsirelson's bound happen? Lets look at a configuration where this happens. The flaw in our CM system we assigned all values simulatenously. But now if we have some spinor and A_1 was measuring the z component and A_2 was measuring the x component then we could violate the correlator.

$$\text{If } 2 \leq |\langle C_{AB}\rangle| \leq 2\sqrt{2}$$

Then the system is said to be entangled(correlators between systems exceed maximum allowed classical values.)

Say we have a joint state $|\psi\rangle$. Lets see how we'd saturate Tsirelson's bound. Lets use the notation $|A_i\rangle = A_i|\psi\rangle$ and $|B_i\rangle = B_i|\psi\rangle$. And now we have a very simple way to sataurate the bound. Consider



$|B_1\rangle+|B_2\rangle = \sqrt{2}|A_1\rangle$ $|B_1\rangle-|B_2\rangle = \sqrt{2}|A_2\rangle \Rightarrow \langle C_{AB}\rangle = \langle A_1|(|B_1\rangle+|B_2\rangle)+\langle A_2|(|B_1\rangle-|B_2\rangle) = 2\sqrt{2}$

Say $A_i^2 = B_i^2 = 1$. Then

$$C_{AB}^2 = 4 - [A_1, A_2][B_1, B_2]$$

Now lets talk about the monogamy of entanglement. Contuining from our definition above now lets say we have a system C where

$$[C_i, A_j] = [C_i, B_j] = 0$$

From here we can define an observable

$$C_{AC} = A_1(C_1 + C_2) + A_2(C_1 - C_2)$$

From our notion of entanglement defined above we can again see how entangled things are. Something else to note is that there is a remarkable inequality

$$\langle C_{AB}\rangle^2 + \langle C_{AC}\rangle^2 \leq 8 \Rightarrow \{|\langle C_{AB}\rangle| < 2 \Rightarrow |\langle C_{AC}\rangle| < 2\} \text{ and converse}$$

"This inequality shows that classical correlators can be shared but quantum correlators cannot." This also shows that correlators are monogomous[1].

We can now talk about average entanglement between subsystesms. We have two systems $e^{S'}$ and e^S in a big system with hilbert space dimension $e^{S+S'}$. We'll also assert that $e^{S'} \ll e^S$. This turns out to not be that strong a statement and really is here to make our life easier. The reason this isn't a very stron statement is because black holes have lots of entropy. Say one system has 10^{10} qubits and the other system has $(1 - 10^{-6}) \times 10^{10}$. That's a small difference in qubits but a huge difference in e^S and $e^{S'}$. Now we want to ask in some precise sense "if we take a typical state of the large system, how are the subsystesms entangled." There are two answers to this question. The answer relies on considering the *density matrix*. Recall that we defined before as taking the Haar measure on the big system. We'll make the following claims

[1] cute

(a) Given operators with simple properties in smaller subsystem then we can find some operators in the larger subsystem to saturate Tsirelson's bound

(b) The entanglement entropy between the subsystem obeyes a "Page curve."

Lets try to prove property (a). We can write a typical state as

$$|\psi\rangle = \sum_{n=1}^{e^S}\sum_{m=1}^{e^{S'}} a_{mn}|m,n\rangle \qquad d\mu_\psi = \pi d^2 a_{mn}\delta\left(\Sigma|a_{mn}|^2 - 1\right)$$

The density matrix of the smaller subsystem is

$$\rho_{mn'} = \sum_{n=1}^{e^S} a_{mn}a^*_{m'n}$$

It should be emphasized that the eigenvalues of the larger density matrix are **the same**. One can always choose a basis to "diagonalize" a_{mn} (actually since this is rectangular it's more precise to do a singular value decompositiion) meaning that you can always write (after some change of basis in smaller and larger system)

$$|\psi\rangle = \sum_{\alpha=1}^{e^{s'}} \sqrt{\rho_\alpha}|\alpha,\alpha\rangle$$

This tells us that the rank of the larger density matrix at most have $e^{S'}$. So the first question we can ask is what is $\langle\rho_{mm'}\rangle$ (wrt to the Haar measure). We can fidn this with

$$\langle\rho_{mm'}\rangle = \int \sum_n a_{mn}a^*_{m'n}d\mu_\psi = \frac{1}{e^{S+S'}}\sum_n \underbrace{\delta_{mm'}}_{e^S}\delta_{nn} = \frac{\delta_{mm'}}{e^{S'}}$$

There is a more useful thing we could ask

$$\left\langle \sum_{mm'}\left|\rho_{mm'} - \frac{1}{e^{s'}}\delta_{mm'}\right|^2\right\rangle = \frac{e^S + e^{S'}}{e^{S+S'}+1} - \frac{1}{e^{S'}} = \delta e^{S'}$$

This computation is done in the lecture value. To get a good notion of the average, we should divide everything by $1/e^{S'}$. At large S and S' we have

$$\delta \approx \frac{1}{e^{S+S'}}$$

This δ is telling us about the average deviation squared and $\sqrt{\delta}$ is the size of the average deviation of eigenvalues.

$$\sqrt{\delta} = \frac{1}{e^{\frac{S+S'}{2}}} \ll \frac{1}{e^{S'}}$$

This tells us that the average density matrix is close to the identity [normalized]. Notice by the way that this does not work if we trtyto use this argument for the larger system. We can now choose a orhtonormal (schmidt?) basis such that

$$|\psi\rangle = \frac{1}{e^{S'/2}} = \sum_{m=1}^{e^{S'}}|m,\tilde{m}\rangle$$

Note that not all states can be written in this way but for a typical state we should be fine. Say we have a pseudospin operator in the smaller subsystem. What we mean is that these operators should behave

$$A_1^2 = A_2^2 = 1 \Rightarrow (A_1 + A_2)^2 = (A - 1 - A_2)^2 = 2$$

This is our way of saying we're looking at a qubit. We can find such operators in QFT as well, it's not only qubit operators. The result we claimed in the beginning is that we can find some operator in the larger subsytem to saturate Tsirelson's bound. Lets show that. Say we have some state. Lets define the matrix elements. Let $|m\rangle$ be the basis that appears inthe schmidt decomposition

$$A_1 |m\rangle = \sum_q (A_1)_{mq} |q\rangle \quad A_2 |m\rangle = \sum_q (A_2)_{mq} |q\rangle$$

The point about finding operators taht are entangled with these operators is that given these operators we can define another operator $\tilde{A}_1 |\tilde{m}\rangle$ that actsin the larger subsystem.

$$\tilde{A}_1 |\tilde{m}\rangle = \sum_q (A_1)_{qm} |\tilde{q}\rangle \leftarrow \text{transpose of above}$$

The matrix elements are the transpose of the smaller subsystem. And we have something similar for $\tilde{A}_2$. It should be clear that

$$|\tilde{A}_1| = |\tilde{A}_2| = 1$$

$$\tilde{A}_1 |\psi\rangle = \frac{1}{e^{S'/2}} = \sum_{m=1}^{S'} \tilde{A}_1 |m, \tilde{m}\rangle = \frac{1}{e^{S'/2}} = \sum_{mq} (A_1)_{qm} |m, \tilde{q}\rangle$$

Renaming $q \leftrightarrow m$ we have

$$\tilde{A}_1 |\psi\rangle = \frac{1}{e^{S'/2}} = \sum_{m,q} (A_1)_{mq} |q, \tilde{m}\rangle = A_1 |\psi\rangle$$

And similarly we can find $\tilde{A}_2 |\psi\rangle = A_2 |\psi\rangle$. Now define

$$B_1 = \frac{1}{\sqrt{2}} \left(\tilde{A}_1 + \tilde{A}_2\right) \quad B_2 = \frac{1}{\sqrt{2}} \left(\tilde{A}_1 - \tilde{A}_2\right) \quad B_1^2 = B_2^2 \quad C_{AB} = \text{same as above}$$

This gives us

$$\langle \psi | C_{ab} | \psi \rangle = \sqrt{2} \left\langle \psi | A_1^2 | \psi \right\rangle + \sqrt{2} \left\langle \psi | A_2^2 | \psi \right\rangle = 2\sqrt{2}$$

To summarize: what we hae argued here is that we had some large system that we devided into two parts (where the dimension of smaller much smaller than alrger.) We then showed that the density matrix of the smaller system is nearly diagonal. Now focusing on an arbitrary operator we showed that we construct a operator in the larger system that is maximally entangled.

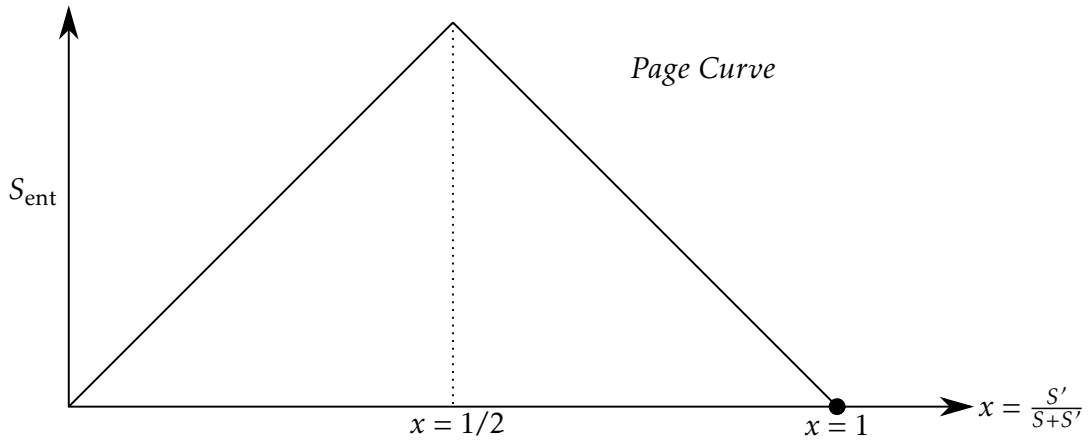
We'll now talk about the page curve. Consider the entanglement entropy of two systems. Lets try to find the exepctation value of this wrt the haar measure

$$\langle -\text{tr}(\rho \ln \rho) \rangle_{\mu_\psi} \approx S'$$

This is true because the density matrix is almost identity. This is also true for the larger system. So we have an assymetry. In general

$$-\mathrm{tr}(\rho \ln \rho) = \min(S', S) \text{ provided } |S' - S| \gg 1$$

The statement is that theentaglement entropy is the dimension of the smaller ofthe two systems. We can now plot the entanglement entropy as a function of the fraction of system size.



The peak is very sharp in a thermosynamic ssystem. We'll apply this to Black Holes soon.