Black Hole Thermodynamics: Then and Now, Part II

Edward Witten, IAS

May, 2021

Let us recall where we were at the end of the first lecture. We discussed the Bekenstein bound, which says that a quantum system of entropy E, size R, and entropy S is supposed to obey a universal relation

$$ER \geq kS$$

with some universal constant k. The bound has the property that it is trivial when the usual conditions for validity of thermodynamics are applicable. The bound is only interesting if it is supposed to apply to every quantum state. So we cannot interpret the "S" in the Bekenstein bound as thermodynamic entropy. It has to be some sort of entropy that makes sense for every quantum state.

Such a concept of entropy is the *von Neumann entropy*, which is the entropy of the density matrix ρ that describes the quantum state:

$$S_{\mathrm{vN}} = -\mathrm{Tr}\, \rho \log \rho.$$

If a system is truly in thermal equilibrium,

$$\rho = \rho_{\mathrm{Thermal}} = \frac{1}{Z} \mathrm{Tr}\, e^{-\beta H},$$

then the von Neumann entropy is equal to the usual thermodynamic entropy S_{Thermal}. But S_{vN} is always defined. When we say a system has thermalized, we don't normally mean literally that its density matrix is $\rho = \rho_{\mathrm{Thermal}}$, which is very hard to achieve for a macroscopic system. We mean that simple, accessible measurements do not distinguish ρ from ρ_{Thermal}. In such a situation, the microscopic von Neumann entropy is less than the thermal entropy (possibly much less)

$$S_{\mathrm{vN}} < S_{\mathrm{Thermal}}.$$

The thermal entropy is what you get if you "coarse grain" and approximate the actual ρ as ρ_{Thermal}.

A most basic difference between the two kinds of entropy concerns the second law. Consider an isolated system, say a fluid in a box initially in an inhomogeneous state. It undergoes unitary evolution with a Hamiltonian H:

$$\rho \to e^{-\mathrm{i}Ht}\rho e^{\mathrm{i}Ht}.$$

Such a unitary transformation does not change the eigenvalues of the density matrix, so it doesn't change the von Neumann entropy:

$$\frac{\mathrm{d}S_{\mathrm{vN}}}{\mathrm{d}t} = 0.$$

Now consider the thermodynamic entropy. Initially, simple measurements could reveal the pressure $p(\vec{x})$ and temperature $T(\vec{x})$ as a function of position $\vec{x}$ in the box, and (assuming that is all we can learn from simple measurements) we would define the thermodynamic entropy correspondingly

$$S_{\rm Thermal} = \int {\rm d}\vec{x}\, s(p(\vec{x}), T(\vec{x})).$$

After a while, the fluid "thermalizes" and simple measurements would reveal less, basically only the average temperature and pressure. The themodynamic entropy will have increased

$$\frac{{\rm d}S}{{\rm d}t} > 0$$

because of coarse-graining over microscopic information that is effectively lost when the system "thermalizes."

Let's ask what kind of entropy is the generalized entropy of Bekenstein

$$S_{\rm gen} = \frac{A}{4G\hbar} + S_{\rm out}.$$

In Bekenstein's interpretation, the $A/4G$ represents a coarse-graining over unobserved internal structure of the black hole, so we should interpret this term as thermodynamic entropy. Bekenstein proposed a Generalized Second Law (GSL)

$$\frac{\rm d}{{\rm d}t} S_{\rm gen} \geq 0.$$

Is it true? Remember the classical limit of the statement is the Hawking area theorem of classical General Relativity:

$$\frac{{\rm d}A}{{\rm d}t} \geq 0.$$

Whenever we are near a classical limit in which the black hole is absorbing matter, the GSL is true, because $G\hbar$ is so small; the increase in the $A/4G\hbar$ term overwhelms the decrease in $S_{\rm out}$ because of matter falling in. We made such estimates last time.

To challenge the GSL in a more interesting way, we can consider a stationary black hole, such as a Schwarzschild black hole, such that classically $\mathrm{d}A/\mathrm{d}t = 0$. We let the black hole interact with quantum fields and ask what happens to S_{gen}. What must we mean by S_{out} if the GSL is to be true? If the quantum fields are in a state such that thermodynamic entropy makes sense, we will be back in a situation in which the GSL is easily satisfied. The challenge is to ask if the GSL is true for an *arbitrary* state of the quantum fields outside the black hole. In this generality, the notion of S_{out} that makes sense is the microscopic or fine-grained entropy of the density matrix ρ_{out} that describes the quantum fields outside the horizon:

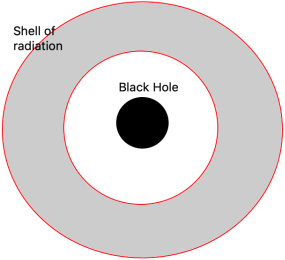
$$S_{\mathrm{out}} = S_{\mathrm{vN,out}} = -\mathrm{Tr}\rho_{\mathrm{out}} \log \rho_{\mathrm{out}}.$$

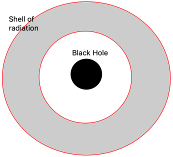
With this interpretation of what we mean by $S_{\rm out}$, both terms in

$$S_{\rm gen} = \frac{A}{4\pi\hbar} + S_{\rm out}$$

are ultraviolet-divergent, but Susskind and Uglum showed (1994) that the divergences cancel and $S_{\rm gen}$ is ultraviolet-finite in perturbation theory. So that again shows what we must mean by $S_{\rm out}$.

Before doing anything fancy, let us just ask what happens to S_{gen} when a black hole, all alone in empty space, is emitting Hawking radiation. The answer is that S_{gen} increases: for a black hole to emit Hawking radiation at temperature T_H into the vacuum, which has temperature $T = 0$, is a thermodynamically irreversible process. A careful estimate was made by D. Page in the 1970's; here is a simple estimate ignoring "gray body factors." During a time $\delta\tau$, the black hole emits thermal radiation that fills a shell of thickness $L = c\delta\tau$:





Let n be the effective number of partial waves in which the outgoing radiation is emitted (because of an angular momentum barrier that we didn't discuss last time, there is an effective cutoff of the angular momentum of an outgoing mode). Effectively the shell contains n modes of $1+1$-dimensional massless chiral fields at temperature T_H, with energy and entropy densities

$$E = \frac{\pi n}{12} T_H^2, \quad S = \frac{\pi n}{6} T_H$$

and hence the change in the energy and entropy of the radiation during time τ are related by

$$\Delta E_{\mathrm{rad}} = \frac{T_H}{2} \Delta S_{\mathrm{rad}}, \quad \Delta S_{\mathrm{rad}} = \frac{2}{T_H} \Delta E_{\mathrm{rad}}.$$

Conservation of energy says that the energy gained by the radiation is the energy lost by the black hole during the same time interval:

$$\Delta E_{\rm rad} = -\Delta E_{\rm BH}.$$

The black hole is emitting radiation adiabatically so the change in its energy and entropy are related by $\mathrm{d}E = T\mathrm{d}S$, or

$$\Delta E_{\rm BH} = T_H \Delta S_{\rm BH}, \quad \Delta S_{\rm BH} = \frac{1}{T_H}\Delta E_{\rm BH}.$$

Comparing this with

$$\Delta S_{\rm rad} = \frac{2}{T_H}\Delta E_{\rm rad},$$

we learn

$$\Delta S_{\rm rad} = -2\Delta S_{\rm BH}.$$

In other words, the entropy gain of the radiation is twice the entropy loss of the black hole, reflecting the fact that for a body to radiate at temperature T_H into the vacuum at temperature 0 is thermodynamically irreversible.

This is encouraging, but can we give a completely general proof of the GSL for a stationary black hole interacting with quantum fields in an *arbitrary* state? Yes, as shown by A. Wall (2011), provided we interpret $S_{\rm out}$ as the von Neumann entropy of the density matrix of the quantum fields outside the horizon.

We are going to use the same *relative entropy* that figured in the discussion of the Bekenstein bound. Recall that the relative entropy between two density matrices ρ, σ, is defined as

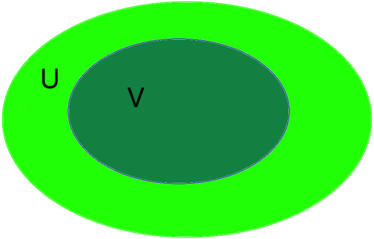
$$S(\rho|\sigma) = \mathrm{Tr}\,(\rho \log \rho - \rho \log \sigma).$$

The key to the Bekenstein bound was the simple fact that

$$S(\rho|\sigma) \geq 0$$

for all ρ, σ, with equality only if $\rho = \sigma$.

For the GSL, the key is a deeper property, closely related to "strong subadditivity" of von Neumann entropy (Lieb and Ruskai, 1973). Let U and $V \subset U$ be two regions of space:

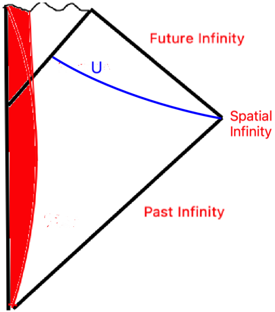


Let ρ_U, σ_U be the density matrices of two states of a quantum field for observations in region U, and let ρ_V, σ_V be the density matrices of the same two states in region V. Then

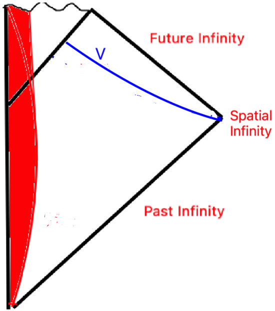
$$S(\rho_V|\sigma_V) \leq S(\rho_U|\sigma_U).$$

(My version of the proof is in section 3 of arXiv:1803.04993. Unfortunately for today we will have to assume this result.)

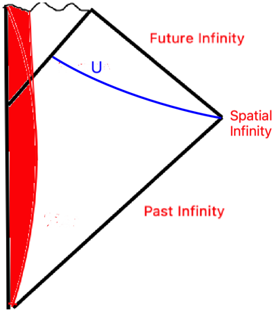
The sort of regions U, V we consider are as follows: We take a "cut" of the horizon at some time, and we define U to be a spacelike surface that goes from the horizon to spatial infinity:



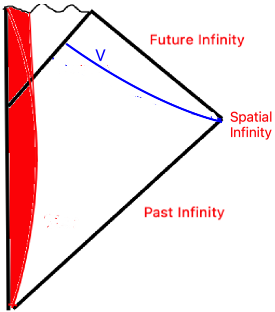
For a smaller region, we take a later cut:



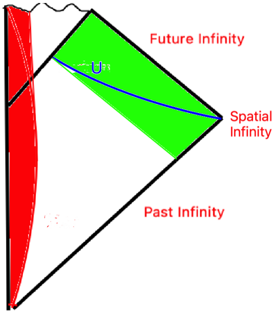
In what sense does the later cut lead to a smaller region?



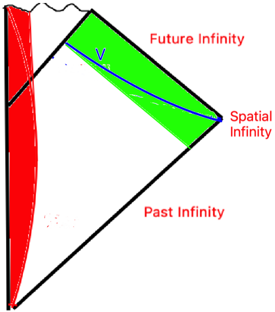
In what sense does the later cut lead to a smaller region?



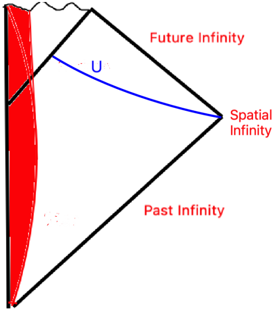
One answer is that the initial value surface with the later cut has a smaller domain of dependence:



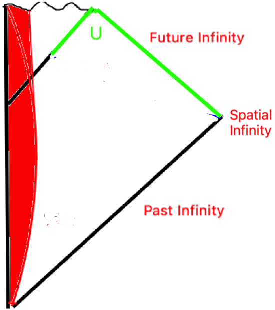
One answer is that the initial value surface with the later cut has a smaller domain of dependence:



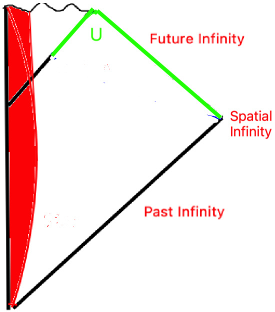
Another answer, which will be technically useful, is to describe the same spacetime region with a different initial value surface, which hugs the horizon plus future infinity outside the horizon. We replace this



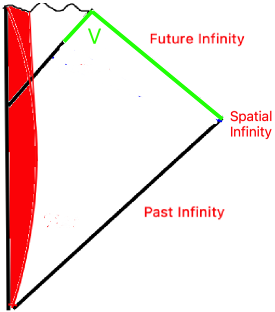
Another answer, which will be technically useful, is to describe the same spacetime region with a different initial value surface, which hugs the horizon plus future infinity outside the horizon. We replace with this



Then the later cut just corresponds to a smaller initial value surface



Then the later cut just corresponds to a smaller initial value surface



So this makes it obvious that we can use the inequality

$$S(\rho_V|\sigma_V) \leq S(\rho_U|\sigma_U).$$

Now, rather as last time, we expand the relative entropy as a sum of two terms:

$$S(\rho|\sigma) = \mathrm{Tr}\,\rho(\log\rho - \log\sigma) = -S(\rho) + E$$

where

$$E = -\mathrm{Tr}\rho\log\sigma$$

We are going to pick ρ to be the quantum state in which we are trying to prove the GSL. So $S(\rho)$ is just going to be S_{out} and the inequality will be

$$S_{\mathrm{out},V} - E_V \geq S_{\mathrm{out},U} - E_U.$$

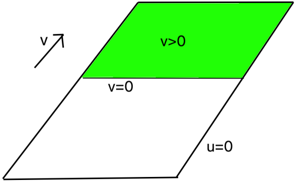
What we want is

$$\frac{\Delta A}{4G\hbar} + \Delta S_{\mathrm{out}} \geq 0,$$

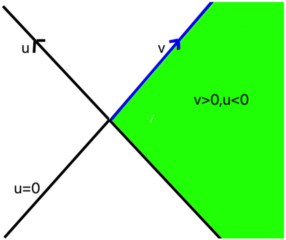
where $\Delta S_{\mathrm{out}} = S_{\mathrm{out},V} - S_{\mathrm{out},U}$, and $\Delta A = A_V - A_U$. So the inequality we have will be the one we want if we can pick σ so that – with the help of the Einstein equations –

$$E_U - E_V = -\frac{\Delta A}{4G\hbar}.$$

What we need to do is to understand quantum fields on a null plane. Let us start with Minkowski space with $\mathrm{d}s^2 = -\mathrm{d}u\mathrm{d}v + \mathrm{d}\vec{x}^2$ and let N be the null plane $u = 0$. "Cut" N at $v = 0$:



The "domain of dependence" (roughly) of the portion N_+ above the cut is (above two dimensions) a Rindler space

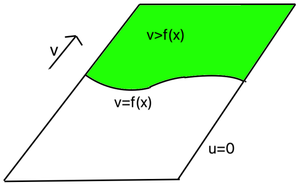


Suppose that we are only able to make observations along N above the cut, or equivalently in the right Rindler space. We learned last time that the density matrix that describes such observations *in the vacuum state* $|\Omega\rangle$ is $\sigma = \exp(-2\pi K)$ where K is the Lorentz boost generator, which can be written on the null plane as

$$K = \int_0^\infty \mathrm{d}v \int \mathrm{d}\vec{x}\, v T_{vv}.$$

(To be precise, this is a partial Lorentz boost operator that boosts fields in the right Rindler space, does nothing in the left Rindler space, and does something complicated in the future and past wedges.)

We want to generalize this for a general cut of the null plane N of the form $v = f(\vec{x})$ for an arbitrary function f:



Let $N_{+,f}$ be the region $v \geq f(\vec{x})$, The domain of dependence of $N_{+,f}$ is a wiggly generalization of the Rindler region, which I will not try to draw.

To find the density matrix that describes measurements in $N_{+,f}$ (or its domain of dependence), we will use the symmetries of operators on the null plane. Consider first the operator P that generates translations in v:

$$P = \int_H \mathrm{d}v \mathrm{d}\vec{x}\, T_{vv}.$$

It annihilates the vacuum $P|\Omega\rangle = 0$ and it is strictly positive on all other states. It generates translations of operators on the null plane (and all other operators, of course):

$$\exp(\mathrm{i}fP)\mathcal{O}(v,\vec{x})\exp(-\mathrm{i}fP) = \mathcal{O}(v+f,\vec{x}).$$

Now instead consider the operator

$$P_f = \int_N \mathrm{d}v \mathrm{d}\vec{x} f(\vec{x}) T_{vv}(v, \vec{x})$$

for a general $f(\vec{x})$. We are going to use this operator to map the cut at $v = 0$ to the cut at $v = f(\vec{x})$. P_f is not a symmetry operator, it does not commute with the Hamiltonian, and off of the null plane N it probably does very complicated things. But on the null plane it is simple. First, the null translation P and the Lorentz boost K have simple commutators with P_f:

$$[P, P_f] = 0, \quad [K, P_f] = P_f.$$

These imply that P_f annihilates the vacuum:

$$P_f |\Omega\rangle = 0.$$

($P_f|\Omega\rangle$ must be a multiple of $|\Omega\rangle$ because $|\Omega\rangle$ is the only state of $P = 0$, and if $P_f|\Omega\rangle = \lambda|\Omega\rangle$, we must have $\lambda = 0$ because of Lorentz invariance.)

The key point is that $e^{\mathrm{i}P_f}$ conjugates operators on the region N_+ on the null plane above the $v = 0$ cut to operators on the region $N_{+,f}$ above the $v = f$ cut. If f is constant, we have $P_f = fP$ and

$$\exp(\mathrm{i}P_f)\mathcal{O}(v, \vec{x}) \exp(-\mathrm{i}P_f) = \mathcal{O}(v + f, \vec{x})$$

while if f is not constant, there are additional terms involving derivatives of f:

$$\exp(\mathrm{i}P_f)\mathcal{O}(v, \vec{x}) \exp(-\mathrm{i}P_f) = \mathcal{O}(v+f, \vec{x})+\sum_i \partial_i f\, \mathcal{O}^i+\sum_{i,j} \partial_i\partial_j f\, \mathcal{O}^{ij}+\cdots .$$

Assuming the operator $\mathcal{O}$ has finite dimension, this is a finite series since after finitely many steps we would be getting operators of negative dimension. The important point is that conjugation by $\exp(\mathrm{i}P_f)$ maps observables in the half-null-plane N_+ into observables in the above-the-cut region $N_{+,f}$.

We have learned that

- The operator $\exp(\mathrm{i}P_f)$ leaves the vacuum state Ω invariant.
- Conjugation by $\exp(\mathrm{i}P_f)$ is an isomorphism between the algebra of observables in region N_+ and the algebra of observables in $N_{+,f}$.

Therefore, conjugation by $\exp(\mathrm{i}P_f)$ will transform the density matrix $\sigma_0 = \exp(-2\pi K)$ for observations in region N_+ to the density matrix σ_f for observations in region $N_{+,f}$:

$$\exp(\mathrm{i}P_f)\sigma_0 \exp(-\mathrm{i}P_f) = \sigma_f.$$

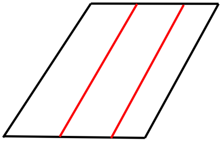
With

$$\log \sigma_0 = -2\pi \int_{v\geq 0} \mathrm{d}v\mathrm{d}\vec{x}\, vT_{vv} = -2\pi K,$$

this gives

$$\log \sigma_f = -2\pi \int_{v\geq f(\vec{x})} \mathrm{d}v\mathrm{d}\vec{x}(v - f(\vec{x}))\, T_{vv} = -2\pi K_f.$$

A small generalization of this formula plus some discussion of Einstein's equations will lead to the GSL. But before going there, I want to explain a heuristic explanation of this result, due to Wall. The idea is that there are no causal relations between operators that are on different null lines on the null plane



Operators on different null lines are spacelike separated and commute. As far as the construction of its Hilbert space is concerned, the theory on the null plane is like an infinite tensor product of 1+1-dimensional theories, parametrized by $\vec{x}$. In each of those theories, at any given $\vec{x}$, f is a constant and $\exp(\mathrm{i}P_f)$ is a symmetry operator that conjugates the density matrix by $v \to v + f$, mapping K to K_f. Then we get $\log \sigma_f$ by integrating the resulting formula over K_f. As Wall also explains, if it is possible to make a lattice regularization of the theory in the $\vec{x}$ directions while maintaining $1 + 1$-dimensional Poincaré invariance, then this argument becomes rigorous.

Now we need to navigate toward the application to black hole horizons. The black hole horizon is, first of all, a null surface, swept out by lightlike geodesics that are called the horizon generators. The general form of its metric is

$$\mathrm{d}s^2 = \sum_{a,b} g_{ab}(\vec{x}, v)\mathrm{d}x^a\mathrm{d}x^b + 0 \cdot \mathrm{d}v\, \mathrm{d}x^a + 0 \cdot \mathrm{d}v^2,$$

where the horizon generators are parametrized by v at fixed $\vec{x} = x^1 \cdots, x^{D-2}$. We can also stipulate that v is an affine parameter along each null geodesic. The metric $g_{ab}(\vec{x}, v)$ is not arbitrary but must obey a constraint equation which is the null analog of the Raychaudhuri equation. This is just the vv component of the Einstein equations, that is it is the equation $G_{vv} = 8\pi G T_{vv}$. To analyze it, it is useful to set $\mathbf{a} = \sqrt{\det g}$, and $\theta = \mathbf{a}^{-1}\partial_v \mathbf{a}$. The Einstein-Raychaudhuri-Sachs constraint equation is then

$$\frac{\partial \theta}{\partial v} = -\frac{\theta^2}{D-2} - \sigma_{ab}\sigma^{ab} - 8\pi G T_{vv}$$

(where σ is the traceless part of $\partial_v g_{ab}$).

Before applying this constraint equation to the GSL, I am going to briefly explain how it leads at the classical level to the Hawking area theorem – the classical limit of the GSL. The point is that everything on the right hand side is negative:

$$\frac{\partial\theta}{\partial v} = -\frac{\theta^2}{D-2} - \sigma_{ab}\sigma^{ab} - 8\pi G T_{vv},$$

by virtue of which one can prove that if there is anywhere on the horizon with $\theta < 0$, then one gets $\theta \to -\infty$ at finite v, which one can show is a contradiction. So it must be that $\theta \geq 0$ everywhere on the horizon. Since $\theta = \mathbf{a}^{-1}\partial_v \mathbf{a}$, $\theta \geq 0$ is equivalent to $\partial_v \mathbf{a} \geq 0$. But with $\mathbf{a} = \sqrt{\det g}$, the black hole area is

$$A = \int \mathrm{d}^{D-2}x\,\mathbf{a}$$

and $\partial_v \mathbf{a} \geq 0$ means that the area is everywhere increasing. That is the Hawking area theorem.

Now in our study of the GSL, we want to consider a horizon that classically has constant area, and study the quantum effects. To make the area constant, we want $\theta = 0$ and for this to be true as v increases into the future, the constraint equation

$$\frac{\partial \theta}{\partial v} = -\frac{\theta^2}{D-2} - \sigma_{ab}\sigma^{ab} - 8\pi G T_{vv}$$

tells us we need $T_{vv} = 0$ in the classical limit, and also $\sigma_{ab} = 0$. But $\theta = \sigma_{ab} = 0$ means that $\partial_v g_{ab} = 0$ so the horizon metric is "stationary" – independent of the null time v:

$$\mathrm{d}s^2 = \sum_{a,b} g_{ab}(\vec{x})\mathrm{d}x^a \mathrm{d}x^b.$$

Now we want to turn on effects of order $\hbar$. In the constraint equation,

$$\frac{\partial\theta}{\partial v} = -\frac{\theta^2}{D-2} - \sigma_{ab}\sigma^{ab} - 8\pi G T_{vv}$$

we need to consider effects of order $\hbar$ in T_{vv}. On the right hand side, since we are considering a situation in which θ and σ are zero classically, they are order $\hbar$ and those terms in the equation are order $\hbar^2$ (plus what can be regarded as a graviton contribution to T_{vv}). Since $\mathbf{a}$ is constant classically, $\partial_v\mathbf{a}$ and $\partial_v^2\mathbf{a}$ are of order $\hbar$. Hence $\partial_v\theta = \partial_v(\mathbf{a}^{-1}\partial_v\mathbf{a}) = \mathbf{a}^{-1}\partial_v^2\mathbf{a} + \mathcal{O}(\hbar^2)$. So the expectation value of the constraint equation becomes

$$\frac{\partial^2\mathbf{a}_\hbar}{\partial v^2} = -8\pi G\langle T_{vv}\rangle\,\mathbf{a}_{\rm cl},$$

where $\langle T_{vv}\rangle$ is the expectation value in whatever quantum state the system is in. On the right hand side, we can use the classical, v-independent value for $\mathbf{a}$, which I call $\mathbf{a}_{\rm cl}$, since $\langle T_{vv}\rangle$ is $\mathcal{O}(\hbar)$. $\mathbf{a}_\hbar$ is the $\mathcal{O}(\hbar)$ correction.

The solution with sensible behavior for $v \to \infty$ is

$$\mathbf{a}_\hbar(\vec{x}, v) = -8\pi G \int_{v' \geq v} \mathrm{d}v' \mathrm{d}\vec{x} \sqrt{\det g}(v' - v) \langle T_{vv}(\vec{x}, v') \rangle,$$

where we drop a possible additive constant that won't be important. Integrating over the transverse coordinates, we learn that the order $\hbar$ correction to the area on a cut at $v = f(\vec{x})$ is

$$A_{f,\hbar} = -8\pi G \left\langle \int_{v \geq f(\vec{x})} \mathrm{d}v \mathrm{d}\vec{x} (v - f(\vec{x})) T_{vv}(\vec{x}, v) \right\rangle.$$

Hence, the order $\hbar$ contribution to $A/4G$ along that cut is

$$\frac{A_{f,\hbar}}{4G} = -2\pi \left\langle \int_{v \geq f(\vec{x})} \mathrm{d}v \mathrm{d}\vec{x} (v - f(\vec{x})) T_{vv}(\vec{x}, v) \right\rangle.$$

Now suppose that there is a state whose density matrix σ_f in the region $N_{+,f} : v \geq f(\vec{x})$ satisfies

$$\log \sigma_f = -2\pi \int_{v \geq f(\vec{x})} \mathrm{d}v \mathrm{d}\vec{x} (v - f(\vec{x})) T_{vv}(\vec{x}, v).$$

We will find that state in a moment. Then the $\mathcal{O}(\hbar)$ contribution to the black hole entropy measured on the cut is

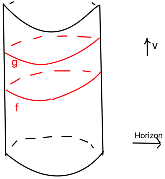
$$\frac{A_{f,\hbar}}{4G\hbar} = \mathrm{Tr} \rho_f \log \sigma_f,$$

where ρ is the density matrix in $N_{+,f}$ for the state that the system is actually in, as opposed to σ which is the density matrix of the handy state with the nice formula for $\log \sigma_f$.

The relative entropy between the two states ρ and σ for the region $N_{+,f}$ is then

$$S(\rho_f|\sigma_f) = \mathrm{Tr}\left(\rho_f \log \rho_f - \rho_f \log \sigma_f\right) = -S(\rho_f) - \frac{A_{f,\hbar}}{4G\hbar}.$$

Now suppose we have two "cuts" g, f with g to the future of f:



Then monotonicity of relative entropy gives

$$S(\rho_f|\sigma_f) \geq S(\rho_g|\sigma_g),$$

which is the Generalized Second Law

$$\frac{A_{g,\hbar}}{4G\hbar} + S(\rho_g) \geq \frac{A_{f,\hbar}}{4G\hbar} + S(\rho_f).$$

To finish the argument, then, for a general stationary horizon

$$\mathrm{d}s^2 = \sum_{a,b} g_{ab}(\vec{x})\mathrm{d}x^a\mathrm{d}x^b$$

we need to find a state σ such that for any region $N_{+,f} : v \geq f(\vec{x})$, the reduced density matrix satisfies

$$\log \sigma_f = -2\pi \int_{N_{+,f}} \mathrm{d}v\mathrm{d}\vec{x}(v - f(\vec{x}))T_{vv}.$$

We have done this already for a Rindler horizon, which is the case that the x^a parametrize $\mathbb{R}^{D-2}$ with a flat metric. Now we need to extend to the general case. This does not require much.

Although physically, we are only assuming that the horizon metric takes the form

$$\mathrm{d}s^2 = \sum_{a,b} g_{ab}(\vec{x})\mathrm{d}x^a\mathrm{d}x^b$$

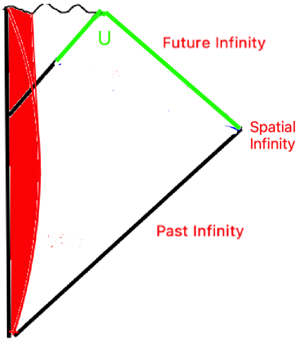
sufficiently far in the future, we can imagine a spacetime in which the metric on some null surface has this form at all times. In fact, let H be a cut of the horizon at $v = f(\vec{x})$ for some f; regardless of f, the metric on H is $\sum_{ab} g_{ab}(\vec{x})\mathrm{d}x^a\mathrm{d}x^b$. Consider the product $\mathbb{R}^{1,1} \times H$ with metric

$$\mathrm{d}s^2 = -\mathrm{d}u\mathrm{d}v + \sum_{a,b} g_{ab}(\vec{x})\mathrm{d}x^a\mathrm{d}x^b$$

and let N be the null surface $u = 0$ in this spacetime. Let Ω be the ground state of our QFT in this spacetime. In our analysis of Rindler horizons, we were in $\mathbb{R}^{1,D-1} = \mathbb{R}^{1,1} \times \mathbb{R}^{D-2}$, with D-dimensional Poincaré symmetry. But the arguments only used two-dimensional Poincaré symmetry. So we can repeat the whole discussion and argue that as in Minkowski space

$$\log \sigma_f = -2\pi \int_{N_{+,f}} \mathrm{d}v\mathrm{d}\vec{x}(v - f(\vec{x}))\, T_{vv}.$$

There is maybe one more important detail to fill in. In this picture

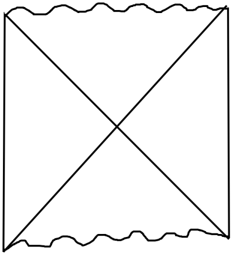


the initial value surface has two pieces, one on part of the null surface N and one at future infinity $\mathcal{I}_+$. Along with the state σ on the null surface N, we could take any state σ' on $\mathcal{I}_+$, so the overall state is $\sigma \otimes \sigma'$. When we compute the relative entropies, there is an extra term from $\mathcal{I}_+$, but it does not depend on the choice of a cut and so doesn't contribute in the discussion of the GSL.

By now we've spent the better part of two lectures discussing the Bekenstein-Hawking formula, "entropy equals horizon area." As was clear from the beginning, this is a kind of thermodynamic entropy. It reflects coarse-graining over the black hole interior, and is appropriate if one is only making observations outside the black hole. But we found that to understand more deeply the thermodynamics of a black hole interacting with quantum fields, we need to consider the "entropy" of the quantum fields to be the microscopic or "fine-grained" von Neumann entropy. This question might motivate one to ask: is there a geometrical formula for some sort of fine-grained version of the Bekenstein-Hawking entropy itself?

Such a formula was found by Ryu and Takayanagi (2006), with later improvements by several others (including Hubeny, Rangamani, and Takayanagi 2007, and Faulkner, Lewkowycz and Maldacena 2013). Note that the Ryu-Takayanagi work preceded Casini on the Bekenstein bound (2008) and Wall on the Generalized Second Law (2011).

To motivate the formula, we can consider a maximally analytically continued black hole solution, describing an entangled state of two universes:



I've drawn the Penrose diagram for the Anti de Sitter version, but I was tempted to illustrate the same point using the Minkowski space version. According to Gibbons and Hawking from the 1970's, as reinterpreted in the AdS context by Maldacena, this picture represents two entangled worlds in the "thermofield double" state

$$\Psi = \frac{1}{\sqrt{Z}} \sum_i \exp(-\beta E_i) |i\rangle \otimes |i\rangle$$

in a tensor product Hilbert space $\mathcal{H}_\ell \otimes \mathcal{H}_r$, one factor for the left universe and one for the right one.

If one constructs from

$$\Psi = \frac{1}{\sqrt{Z}} \sum_i \exp(-\beta E_i/2) |i\rangle \otimes |i\rangle$$

a density matrix for just the right or left universe, it is

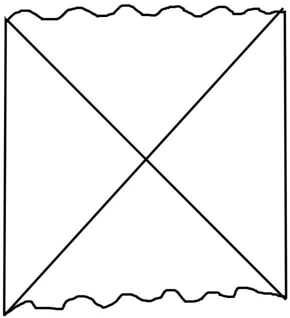
$$\rho_r = \mathrm{Tr}_{\mathcal{H}_\ell} |\Psi\rangle\langle\Psi| = \frac{1}{Z} \sum_i \exp(-\beta E_i) |i\rangle\langle i|,$$

or similarly

$$\rho_\ell = \mathrm{Tr}_{\mathcal{H}_r} |\Psi\rangle\langle\Psi| = \frac{1}{Z} \sum_i \exp(-\beta E_i) |i\rangle\langle i|,$$

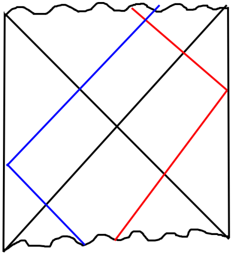
that is, a standard thermal density matrix.

According to Gibbons and Hawking (as adapted to the AdS context) the entropy of this density matrix is the horizon area, and for this ideal solution it does not matter which horizon one picks or where one measures it



However, suppose two observers, one living on the left, and one on the right, decide to disturb this system. The left observer can apply a unitary operator to $\mathcal{H}_\ell$ and the right observer can apply a unitary operator to $\mathcal{H}_r$.

The geometry changes a lot:



The corresponding state changes also, by

$$\Psi \to U_\ell \otimes U_r \Psi \in \mathcal{H}_\ell \otimes \mathcal{H}_r.$$

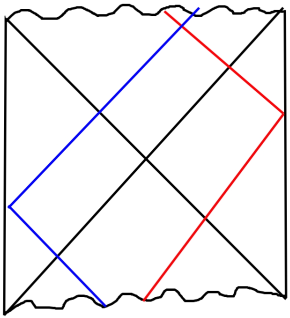
The two observers can do a lot to the system, but there is something they cannot change: they can only change the left or right density matrices by conjugation

$$\rho_\ell \to U_\ell \rho_\ell U_\ell^{-1}, \qquad \rho_r \to U_r \rho_r U_r^{-1}$$

and this will not change the entanglement or fine-grained entropy

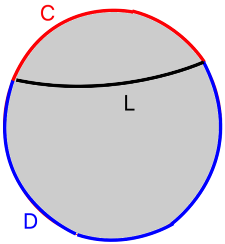
$$S = -\mathrm{Tr}\, \rho_\ell \log \rho_\ell = -\mathrm{Tr}\, \rho_r \log \rho_r.$$

So if this entanglement entropy can be represented by something in geometry, it will be something that observers on the two sides cannot change. Ryu and Takayanagi (with later elaborations) identified such a quantity, namely $A/G\hbar$, where A is the area of an extremal surface that is homologous to a Cauchy hypersurface of either of the two boundaries:



In this example, it is the bifurcation surface where the two horizons cross; the observers on left and right can do nothing to change it.

Ryu and Takayanagi did not originally formulate their idea primarily for the case of such a pair of entangled universes. They considered a single asymptotically AdS spacetime and they divided a Cauchy hypersurface of the boundary into two parts C and D. Then they considered the reduced density matrices ρ_C, ρ_D of the vacuum state Ψ (for example) in regions C or D. Their proposal was that the entanglement entropy $S(\rho_C) = S(\rho_D)$ is the area of a minimal area surface L in the bulk that separates the two boundary regions:



Note that the entanglement entropy is ultraviolet divergent in quantum field theory, as we discussed in the first lecture, and the area (or in the picture, the length) of L is also divergent. The claim is that cutoff versions of the entropies match cutoff versions of the areas or lengths.

The RT formula had a big impact from the beginning, and its impact has only grown as the importance of fine-grained or microscopically defined entropy has come to be further appreciated, partly because of the developments that I reviewed in the two lectures. Unfortunately there isn't really time to tell about further developments that resulted from the RT formula.

In summary, today we discussed the Generalized Second Law and we learned that, like the Bekenstein bound, to understand it well requires interpreting matter entropy to be the fine-grained or von Neumann entropy of a density matrix. And I at least gave a few hints about the fine-grained version of the Bekenstein-Hawking formula, of which the earliest version is due to Ryu and Takayanagi.