Introduction to Black Hole Thermodynamics

Edward Witten, IAS

Madan Lal Mehta Lecture, April 19, 2021

In the first half of this talk, I will explain some key points of black hole thermodynamics as it was developed in the 1970's.

In the second half, I will explain some more contemporary results, though I will not be able to bring the story really up to date.

I cannot explain all the high points in either half of the talk, unfortunately, as time will not allow.

Black hole thermodynamics started with the work of Jacob Bekenstein (1972) who, inspired by questions from his advisor John Wheeler, asked what the Second Law of Thermodynamics means in the presence of a black hole.

The Second Law says that, for an ordinary system, the “entropy” can only increase. However, if we toss a cup of tea into a black hole, the entropy seems to disappear. Bekenstein wanted to “generalize” the concept of entropy so that the Second Law would hold even in the presence of a black hole. For this, he wanted to assign an entropy to the black hole.

What property of a black hole can only increase? It is *not* true that the black hole mass always increases. A rotating black hole, for instance, can lose mass as its rotation slows down. But there is a quantity that always increases: Stephen Hawking had just proved the "area theorem," which says that the area of the horizon of a black hole can only increase. So it was fairly natural for Bekenstein to propose that the entropy of a black hole should be a multiple of the horizon area. For example, for a Schwarzschild black hole of mass M

$$\mathrm{d}s^2 = -\left(1 - \frac{2GM}{r}\right)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{1 - \frac{2GM}{r}} + r^2 \mathrm{d}\Omega^2,$$

the horizon is at

$$R = 2GM$$

and the horizon area is

$$A = 4\pi R^2 = 16\pi^2 G^2 M^2.$$

Since entropy is dimensionless, to relate the entropy of a black hole to its area, one requires a constant of proportionality with dimensions of area. From fundamental constants $\hbar, c$ and $G = \mathrm{Newton's\ constant}$, one can make the Planck length $\ell_P = (\hbar G/c^3)^{1/2} \cong 10^{-33}$ cm, and the Planck area ℓ_P^2. In units with $\hbar = c = 1$, Bekenstein's formula for the entropy was

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

where the constant $1/4$ was not clear in Bekenstein's work and was provided by Stephen Hawking a few years later, in a way that I will explain. For a Schwarzschild black hole

$$S = 4\pi^2 GM^2.$$

Bekenstein's idea was that the entropy of a black hole was supposed to capture the information lost when the black hole was formed – he interpreted it as the logarithm of the number of possible ways the black hole might have formed. Bekenstein proposed a "generalized second law" saying that the "generalized entropy"

$$S_{\mathrm{gen}} = \frac{A}{4G} + S_{\mathrm{out}}$$

always increases. Here S_{out} is the ordinary entropy of matter and radiation outside the black hole. The claim is that when something falls into the black hole, S_{out} may go down but $A/4G$ increases by more.

Bekenstein made a few tests of the generalized second law. Here is one. Shine photons with a wavelength λ and (therefore) energy $E = 1/\lambda$ on the black hole. The entropy of a single photon is of order 1, for example because the photon has two polarization states. When the black hole absorbs one photon, its mass shifts by

$$\Delta M = \frac{1}{\lambda}$$

so its entropy $S_{\rm bh} = 4\pi^2 GM^2$ shifts by

$$\Delta S_{\rm bh} = 4\pi^2 G((M + 1/\lambda)^2 - M^2) \cong 8\pi^2 G \frac{M}{\lambda}.$$

Bekenstein wanted $\Delta S_{\rm bh} > \Delta S_{\rm out} \cong 1$. He observed that if the black hole is capturing a photon of size smaller than the Schwarzschild diameter $2R = 4GM$ of the black hole, say

$$\lambda << 4GM$$

then

$$\Delta S_{\rm bh} >> 2\pi^2$$

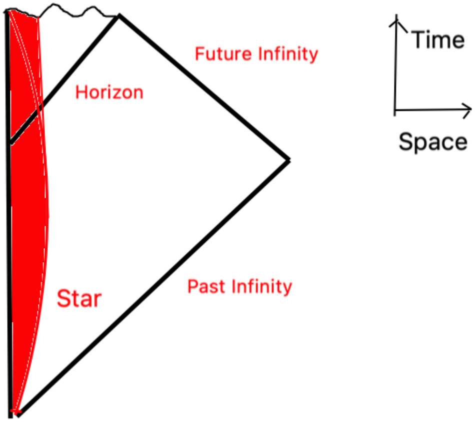
which is satisfactory. (He did a more complete calculation for a rotating black hole and got a smaller but still satisfactory result).

However, Bekenstein did not really get a satisfactory answer if the black hole is absorbing photons of wavelength *larger* than the black hole size – which can happen, though not very efficiently. This question really does not have a satisfactory answer in the framework that Bekenstein was assuming, which was that whatever falls behind the black hole horizon stays there forever. In thermodynamic terms, since Bekenstein assumed that the black hole does not radiate, one would have to assign it a temperature of 0. Thermodynamics says that at equilibrium the changes in energy E and entropy S of a system are governed by

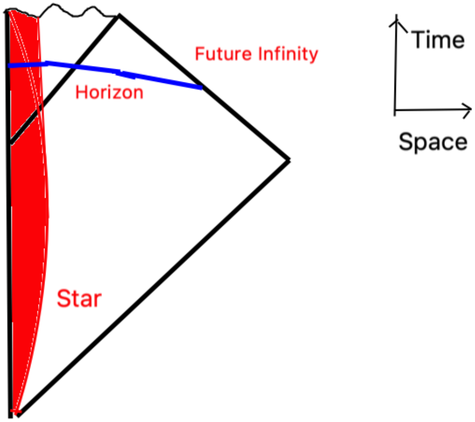
$$\mathrm{d}E = T\mathrm{d}S$$

or $\mathrm{d}S = \mathrm{d}E/T$, so a system with $T = 0$ should have $\mathrm{d}S = \infty$ if $\mathrm{d}E \neq 0$. But Bekenstein wanted to attribute a finite, not infinite, entropy to the black hole. One cannot analyze the absorption of very long wavelength photons by the black hole while ignoring the fact that the black hole is strongly emitting such photons.

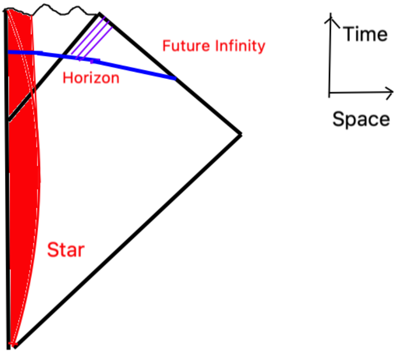
Famously, Stephen Hawking discovered in 1974 that at the quantum level, a black hole is not really black – it has a temperature proportional to $\hbar$. Hawking discovered this by analyzing the behavior of quantum fields in a black hole geometry:



Measurements that an observer at future null infinity will make can be traced back to initial conditions of the quantum field on a Cauchy hypersurface. It is convenient to pick a hypersurface that crosses the horizon to the future of the collapsing star:

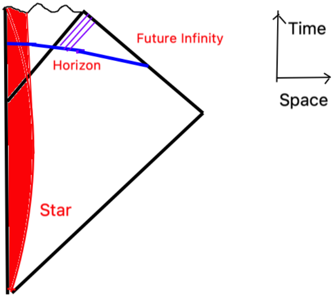


This picture shows signals propagating out at the speed of light from the initial value surface to the observer at infinity:



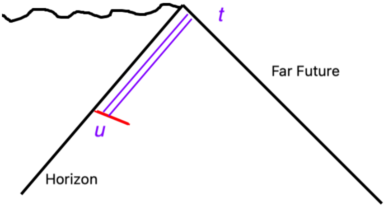
What will the observer see? Part of Hawking's insight was that although the full details of exactly what the observer will see depend on the details of the collapsing star, if we ask what the observer will see *in the far future* after transients die down, we will get a universal answer.

The most important point about this picture is that a signal that is received very late



originated from very close to the horizon. This means that observations made at late times depend on measurements of the state of the quantum fields at short distances. But every state is equivalent to the vacuum at short distances. So the late time observations of the observer probe the vacuum state near the horizon at short distances. That is why Hawking got a universal answer for the late time behavior, regardless of exactly how the black hole formed.

Let u be a coordinate function that vanishes on the horizon on some particular Cauchy slice - it doesn't matter precisely how u is defined.



And let t be the time at which a signal is detected by an observer at infinity. The relation between u and t is

$$t = 4GM \log \frac{1}{u} + C_0 + \mathcal{O}(u),$$

where C_0 is an integration constant. One finds this formula by solving the geodesic equation for an outgoing null geodesic. Rescaling u will only shift the unimportant constant C; nonlinear redefinitions of u will affect the unimportant $\mathcal{O}(u)$ terms.

We can solve the equation $t = 4GM \log \frac{1}{u} + C_0 + \mathcal{O}(u)$ for u:

$$u = C_1 \exp(-t/4GM).$$

At late times, that is if t is large, u is exponentially small. Moreover, $\mathrm{d}u/\mathrm{d}t$ is also exponentially small, which means that a mode observed at infinity will have undergone an exponentially large redshift on its way. A mode of any given energy E that is observed at a sufficiently late time will have originated from a very high energy mode near the horizon. That is why there is a simple answer. A mode of very high energy propagates freely, along the geodesics that I've been drawing.

The observer at infinity probes the radiation by measuring a quantum field $\psi(t)$. A typical observable is a two-point function

$$\langle\psi(t)\psi(t')\rangle.$$

Near the horizon, if the field ψ is for simplicity a free fermion with dimension $1/2$ in the $1+1$-dimensional sense, one would have had

$$\langle\psi(u)\psi(u')\rangle = \frac{(\mathrm{d}u\,\mathrm{d}u')^{1/2}}{(u-u')}.$$

Setting $u = C_1\exp(-t/4GM)$, we see that for the observer at infinity, this translates to

$$\langle\psi(t)\psi(t')\rangle = \frac{(\mathrm{d}t\mathrm{d}t')^{1/2}}{\exp((t-t')/8GM) - \exp(-(t-t')/8GM)}.$$

This is antiperiodic in imaginary time, that is it is odd under $t \to t + 8\pi GM\mathrm{i}$. That antiperiodicity corresponds to a thermal correlation function at a temperature $T_H = 1/8\pi GM$, which is the Hawking temperature.

In other words, a black hole, after transients that depend on how it was created die down, radiates thermally at a temperature $T_H = 1/8\pi GM$. This explains why Bekenstein had had trouble making sense of the interaction of the black hole with low energy photons. It also lets us confirm the value of the entropy: using

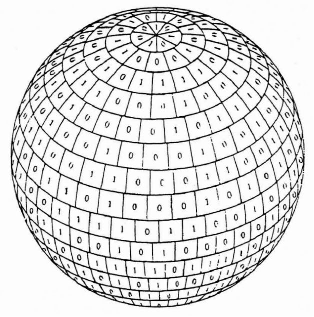
$$\mathrm{d}E = T\mathrm{d}S$$

where $E = M$ and $T = 1/8\pi GM$ gives $\mathrm{d}S = 8\pi GM\mathrm{d}M$ so $S = 4\pi GM^2$. The area of a Schwarzschild black hole is $A = 16\pi G^2M^2$ so the entropy is

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

This is how Hawking confirmed Bekenstein's ansatz and determined the constant that was unclear in Bekenstein's work.

Many researchers have thought that, somehow, the entropy $S = A/4G$ means that the black hole can be described by some sort of degrees of freedom that live at its surface – one bit or qubit for every $1/4G$ of area. For example, in a famous article in 1992, John Wheeler illustrated that idea with this picture:



A fundamental point about Hawking radiation is that the radiation appears to be thermal even though the black hole could have formed from a pure state. The reason that this happened is that the observations of the distant observer amount to observing a quantum field that lives in $1+1$ dimensions on only half of space:

Think of the ground state of a quantum field as a function $\Psi(\phi(x))$ where $\phi(x)$ is the field on the real line. Further, think of $\phi(x)$ as a pair $\phi_\ell(x)$, $\phi_r(x)$, where ϕ_ℓ is defined on the left half of the line and ϕ_r on the right half. So the ground state is a function $\Psi(\phi_\ell, \phi_r)$. Let us discuss how to make a "density matrix" appropriate for observations of ϕ_r only.

Let us remember the general idea of a density matrix: We start with a pure state $\psi_{\rm AB}$ in a tensor product Hilbert space $\mathcal{H}_{\rm A} \otimes \mathcal{H}_{\rm B}$. We first make the "pure state" density matrix

$$\rho_{\rm AB} = |\psi_{\rm AB}\rangle\langle\psi_{\rm AB}|$$

The expectation value of any operator $\mathcal{O}_{\rm AB}$ is

$$\langle\psi_{\rm AB}|\mathcal{O}_{\rm AB}|\psi_{\rm AB}\rangle = {\rm Tr}_{\rm AB}\,\mathcal{O}_{\rm AB}\rho_{\rm AB}.$$

Note that $\rho_{\rm AB}$ is the orthogonal projection operator on the state $\psi_{\rm AB}$; in particular, it is hermitian, non-negative, satisfies

$${\rm Tr}_{\rm AB}\,\rho_{\rm AB} = 1$$

and has rank 1.

Now suppose we are only going to observe the subsystem A. That means that we consider only operators of the form $\mathcal{O}_{\mathrm{AB}} = \mathcal{O}_{\mathrm{A}} \otimes 1_{\mathrm{B}}$. The expectation of this operator in the state ψ_{AB} is

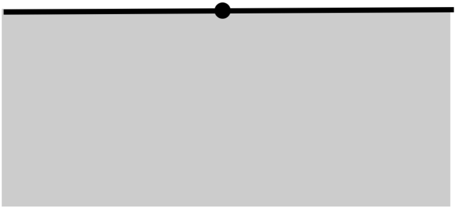
$$\mathrm{Tr}_{\mathrm{AB}}\,(\mathcal{O}_{\mathrm{A}} \otimes 1_{\mathrm{B}})\rho_{\mathrm{AB}} = \mathrm{Tr}_{A}\,\mathcal{O}_{\mathrm{A}}\rho_{\mathrm{A}}$$

where

$$\rho_{\mathrm{A}} = \mathrm{Tr}_{\mathrm{B}}\,\rho_{\mathrm{AB}}.$$

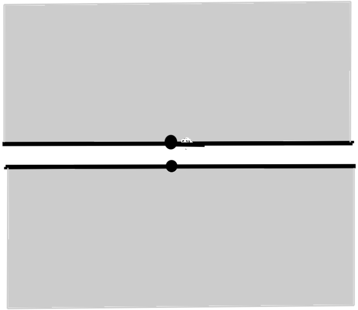
In other words, for measurements on system A only, we can use the density matrix ρ_{A} which is obtained from ρ_{AB} by taking a "partial trace" on $\mathcal{H}_{\mathrm{B}}$. The definition of ρ_{A} ensures that it is hermitian, non-negative, and has trace 1, just like ρ_{AB}, but it does *not* necessarily have *rank* 1. If ψ_{AB} is an "entangled" state of the subsystems A and B, then ρ_A has rank greater than 1. An entangled state is just a state that is not a tensor product $\psi_{\mathrm{A}} \otimes \widetilde{\psi}_{\mathrm{B}}$ of separate states of the two subsystems. So almost every quantum state of the combined system is entangled.

To imitate this in field theory, we first need a convenient representation of the ground state wavefunction $\Psi(\phi_\ell, \phi_r)$. It is given by a path integral on the lower half-plane:



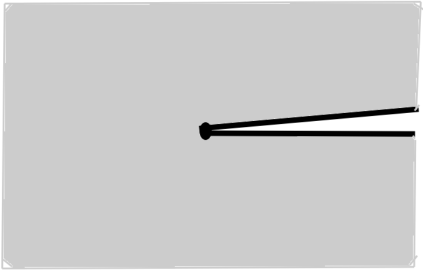
with boundary conditions in which the boundary values ϕ_ℓ, ϕ_r on the boundary are specified. The path integral computes a wavefuncetion $\Psi(\phi_\ell, \phi_r)$.

Similarly, to make the pure state density matrix $|\Psi\rangle\langle\Psi|$ for the whole line, we multiply a path integral on the lower half-plane by a similar path integral on the upper half-plane:



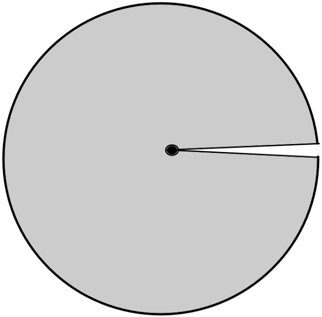
Now the "partial trace" that gives a density matrix for the right half only is accomplished by setting the fields ϕ_ℓ in the left halves of the picture to be equal and integrating over ϕ_ℓ. This glues together the left halves of the upper and lower boundaries.

Now the path integral that we are doing looks like this:



It is a path integral on the whole plane with a cut on the positive real axis. The path integral computes a function $\rho(\phi_r, \phi_r')$, where ϕ_r' are the boundary values above the cut and ϕ_r are boundary values below the cut. This is our density matrix.

However, this path integral can be understood in another way:



I've drawn the same thing, but in a way that emphasizes the rotational symmetry. One can generate the cut plane by starting with a half-line and rotating it by a 2π angle around its endpoint. This gives us a formula for the density matrix. If R is the operator that generates the rotation, the density matrix is

$$\rho = \exp(-2\pi R).$$

It is useful to write this formula in Lorentz signature. In Lorentz signature, R becomes the Lorentz boost operator K so

$$\rho = \exp(-2\pi K).$$

However, since ρ is a density matrix acting on a state defined only on the positive half-space, say $x > 0$, likewise here K is a Lorentz boost generator defined in the region $x > 0$:

$$K = \int_0^\infty \mathrm{d}x \, x T_{00}(x).$$

As usual for a charge generator, the integral is over space (here a half-space) at a fixed time, say $t = 0$.

Let us go back to our formula

$$u = C_1 \exp(-t/4GM). \qquad (*)$$

We can now give a new explanation of why this formula led to a thermal density matrix. Observations near future infinity are equivalent to observations made near the horizon with $u > 0$. That means that our discussion applies and in terms of u, the density matrix is $\exp(-2\pi K)$. Let us convert that to the way it would be viewed by an observer at infinity. The Lorentz boost operator acts on u by $u\partial_u$. Using the formula $(*)$ we see that $u\partial_u = -4GM\partial_t$. The generator ∂_t corresponds to the Hamiltonian H. So the mapping from u to t maps K to $4GMH$, and the density matrix $\rho = \exp(-2\pi K)$ becomes

$$\rho = \exp(-8\pi GMH),$$

which is a thermal density matrix at $T_H = 1/8\pi GM$.

With a little more time, I would go on now to explain the Euclidean picture of black hole thermodynamics that was developed by Gibbons and Hawking, among others. Unfortunately that would come at the cost of not being able to give at least a taste of some of the modern results. Progress in the 21st century has largely depended on understanding the meaning of quantum entropy at a deeper level. I am really only going to be able to explain one result in this direction, which was by H. Casini (2008).

We need to begin by discussing more thoroughly what entropy means at the quantum level. The original definition of entropy in terms of microphysics was by Boltzmann in the 19th century. Consider a system of N particles in a box with positions $\vec{x}$ and momenta $\vec{p}$. As a classical physicist, Boltzmann assumed that at a given time, $\vec{x}$ and $\vec{p}$ have definite values even if we do not know them. He described the state of our knowledge by a probability distribution function $\rho(\vec{p}, \vec{x})$ that encodes our knowledge, and – after great labor – defined the entropy as the phase space integral of $-\rho \log \rho$:

$$S = \int \mathrm{d}\vec{p}\, \mathrm{d}\vec{x}\, (-\rho(\vec{p}, \vec{x}) \log \rho(\vec{p}, \vec{x}))\,.$$

For a quantum system, the *density matrix* ρ, which we introduced earlier, is the closest analog of the classical probability distribution function. Recall that if a system is described by density matrix ρ, then the expectation value of any observable $\mathcal{O}$ is

$$\langle \mathcal{O} \rangle = \mathrm{Tr}\, \mathcal{O} \rho.$$

After diagonalizing the hermitian matrix ρ as $\rho_{ij} = \delta_{ij} \lambda_j$ for some $\lambda_1, \cdots, \lambda_N$, we have therefore

$$\langle \mathcal{O} \rangle = \sum_{i=1}^{N} \lambda_i \mathcal{O}_{ii},$$

as if the system is in state i with probability λ_i. This interpretation makes sense, since, because ρ is nonnegative and $\mathrm{Tr}\, \rho = 1$, the λ_i are nonnegative and their sum is 1.

So the quantum analog of Boltzmann's integral

$$S = \int \mathrm{d}\vec{p}\,\mathrm{d}\vec{x}\,(-\rho(\vec{p},\vec{x})\log\rho(\vec{p},\vec{x}))$$

is the *von Neumann entropy* of a density matrix

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

In the classical limit, the von Neumann entropy goes over to the classical entropy (times a constant that is poorly defined classically). If the eigenvalues of ρ are $\lambda_1, \lambda_2, \cdots, \lambda_N$, then

$$S = -\sum_{i=1}^{N} \lambda_i \log\lambda_i.$$

The state of a system is known with certainty if one of the λ_i is 1 and the others 0 – which means that ρ has rank 1. In that case, the formula

$$S = -\sum_{i=1}^{N} \lambda_i \log \lambda_i$$

gives $S = 0$. Otherwise every nonzero λ makes a positive contribution in the sum so

$$S > 0.$$

If a system is entangled with something else, ρ has rank greater than 1 and therefore $S > 0$.

However, there is a fundamental difference between the classical and quantum cases. Classically, one assumes that $\vec{x}$ and $\vec{p}$ have actual values, and one is describing a system by a distribution function $\rho(\vec{p}, \vec{x})$ because of lack of microscopic knowledge. Quantum mechanically, a system can have a microscopic or "fine-grained" entropy even if we have as full a description of its state as quantum mechanics allows. That happens because of entanglement. Suppose that A and B are two quantum systems in an overall pure state ψ_{AB}. If ψ_{AB} is entangled, then the density matrix ρ_{A} of system A has rank greater than 1 and system A has a positive von Neumann entropy. If A and B are entangled, this can be verified experimentally and there is no way to describe system A by a pure state density matrix with zero entropy.

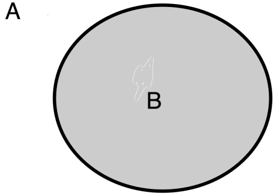
In general, just as classically, our knowledge of the state of a system might be less complete than quantum mechanics allows. In that case, we describe a system by a density matrix ρ that reflects our knowledge. For example, if we know nothing about the state of a system, we would take ρ to be a multiple of the identity, even if someone else (who maybe knows how the system was prepared) would describe it by a pure state. If we know nothing about a system except its temperature, we use $\rho = \frac{1}{Z} e^{-\beta H}$, where Z is chosen so $\mathrm{Tr}\,\rho = 1$. In such a situation, the von Neumann entropy

$$S = -\mathrm{Tr}\,\rho \log \rho$$

is the thermodynamic entropy, a familiar concept classically.

What is different about quantum mechanics is the entropy that remains if our knowledge of the state of a system is as complete as quantum mechanics allows. This has been called “entanglement entropy” – because it results from the entanglement of a system with some other system – and it has also been called “fine-grained entropy.”

The idea that the Bekenstein-Hawking entropy of a black hole should be understood in terms of entanglement entropy was apparently first put forward by R. Sorkin in 1983 (in a paper that attracted only modest attention at the time). The idea was just the following. In a quantum field theory, divide space into two regions A and B

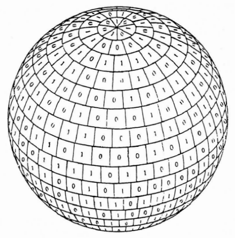


Let Ψ be the vacuum state, and ρ_{A} the "reduced density matrix" of the vacuum for the state Ψ. One can try to calculate the fine-grained entropy S_{A}. One finds that it is ultraviolet divergent but the coefficient of the divergence is proportional to the area A of the boundary between regions A and B.

Sorkin's idea, in modern language, was that somehow gravity cuts off the ultraviolet divergence, leaving an entanglement entropy in the vacuum between the two regions that is the Bekenstein-Hawking entropy $A/4G$, where A is the area of the boundary between them. This makes a lot of intuitive sense, as it matches two ideas:

(1) $A/4G$ is the irreducible entropy of the system for someone who has access only to the region outside the horizon

(2) the divergence in the entanglement entropy is proportional to A because it comes from short wavelength modes near the "horizon," as if (after cutting off the divergence) the density of quantum degrees of freedom on the horizon per unit area is $1/4G$ as in Wheeler's picture:



Twenty-first century developments have supported the intuition in these statements, though leaving us with plenty of mysteries.

I will use the remaining time to explain just one example where something really informative has been said which depends on a better understanding of what we should mean by "entropy." This involves the work of H. Casini (2008) on the "Bekenstein bound."

Bekenstein (1980) considered whether the Generalized Second Law (GSL) is obeyed when a black hole of mass M and therefore radius $R = 2GM$ absorbs a body of size $\mathcal{R}$, energy E, and entropy S. The entropy S of the body disappears, so the GSL says that the increase in the black hole entropy must be more than that. The black hole entropy $A/4G$ changes, as we actually computed in an earlier discussion of Bekenstein's work, by $8\pi^2 GME$. We therefore would like

$$8\pi^2 GME > S$$

if the black hole can absorb the given body. If one naively says that a black hole of size $2GM$ can only absorb a body of size $\mathcal{R}$ if $\mathcal{R} < 2GM$, then the desired inequality becomes

$$4\pi^2 \mathcal{R} E > S.$$

Note that this formula makes no mention of gravity or Newton's constant – it is potentially just a statement about ordinary quantum physics without gravity.

My explanation did not do justice to Bekenstein's argument, which involved considering a highly rotating black hole and gave a more convincing explanation with better constants. Let us interpret the "Bekenstein bound" to be the statement that in quantum field theory without gravity, there should be a universal inequality between the size, energy, and entropy of a system, of the general form

$$\mathcal{R}E > kS,$$

with a universal constant k.

Faced with a conjecture like this, two questions come to mind: (1) Is it true? and (2) If true, is it interesting? For years, there seemed to be ample basis for skepticism on both counts.

On the question of whether the Bekenstein bound was true, there seemed to be a trivial argument that it could not be true, at least not as a universal statement about all quantum field theories. Consider a theory that has N elementary particles all of the same mass. Such theories exist, for instance free field theories. Consider a system consisting of a box with one of these particles inside it. The entropy seems to be at least $\log N$, from the choice of one of N possible particles to put in the box. On the other hand, the size and energy of the box do not depend on N. So it seemed clear that the Bekenstein bound could not be a universal statement about quantum field theory.

On the other hand, if true, is the Bekenstein bound interesting? Can we think of a system where the Bekenstein bound is close to being violated? Consider a system made of massless particles. (Making the particles massive increases the energy without increasing the entropy so it goes in the wrong direction.) For example, consider a box of size ρ containing a gas of massless particles at temperature T. In three dimensions, the number of particles in the box is of order $(\rho T)^3$ and the entropy is of the same order

$$S \sim (\rho T)^3.$$

On the other hand, the energy is

$$E \sim \rho^3 T^4.$$

So the Bekenstein ratio $S/\rho E$, on which we want an upper bound, is of order

$$\frac{S}{\rho E} \sim \frac{1}{\rho T}.$$

If $\rho T >> 1$, the notion of a thermal gas in a box makes sense. In that case, $S/\rho E << 1$ and the Bekenstein bound is satisfied by a wide margin.

How do we find a system for which the Bekenstein bound is interesting as well as true? Keeping still many particles in the box, but assuming the particles are not in thermal equilibrium, makes matters worse since it lowers the entropy S for given energy E. To make the Bekenstein bound challenging, we need to reduce the number of particles, which we can do by lowering the temperature.

The best case is that the box just contains 1 or a few massless particles, or equivalently $T \leq 1/\mathcal{R}$. But a box containing just 1 or a few massless particles is going to weigh much more than the particles that it contains. In estimating the Bekenstein ratio, we cannot ignore the mass of the box (which might have been unimportant when there are many particles in the box but is important if the box is almost empty). Including the mass of the box goes in the wrong direction again, and makes the Bekenstein bound uninteresting.

So in other words, the case that the Bekenstein bound is interesting is the case of a single particle without a box. But what sense does the Bekenstein bound make in that case? In other words, what is the entropy of a single particle? And for that matter, what is the size of a system consisting of a single particle? In other words, when we try to go to a regime in which the Bekenstein bound is interesting, the statement of the Bekenstein bound does not seem to make sense.

Amazingly, Casini was able to give a precise definition to the terms in the Bekenstein bound and to prove it. To explain this, I need to tell you one more thing about density matrices. Let σ, ρ be two density matrices on the same Hilbert space $\mathcal{H}$. The *relative entropy* between them is defined as

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

It measures, in a sense that turns out to be useful, how different are ρ and σ. For now, what we care about is just this: $S(\rho||\sigma)$ is non-negative, and vanishes only if $\sigma = \rho$. To prove this, one lets $\sigma(t) = (1-t)\rho + t\sigma$ for $0 \leq t \leq 1$ and one considers the function

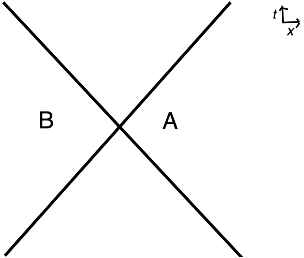
$$f(t) = S(\rho||\sigma_t) = \mathrm{Tr}\,(\rho \log \rho - \rho \log \sigma(t))\,.$$

Then $f(0) = f'(0) = 0$ and if $\rho \neq \sigma$ then $f''(t) > 0$ for $0 \leq t \leq 1$.

(Prove this using $\log \sigma(t) = \int_0^\infty \mathrm{d}s(1/(s+\sigma(t)) - 1/(s+1))$.)

So $S(\rho||\sigma) = f(1) > 0$ if $\rho \neq \sigma$.

Casini considered the vacuum state Ω of an arbitrary quantum field theory and any other state Ψ. The goal is to prove that the Bekenstein bound is valid for the state Ψ, with the correct interpretation of the terms. First, Casini divided space into the two halves $x > 0$ and $x < 0$ where x is one of the spatial coordinates. The corresponding division of spacetime looks like this in a two-dimensional picture where I show only x and the time t:



The left and right wedges are called Rindler wedges. Casini consider the density matrices $\rho = \rho_\Psi$ and $\sigma = \sigma_\Omega$ of the states Ψ and Ω reduced to region A – that means where one only considers measurements in region A. The relative entropy between them measures how well Ψ and Ω can be distinguished by someone who only makes measurements in region A. (For brevity I will use here the language of density matrices, though a rigorous formulation of what I am about to say uses von Neumann algebras instead.)

Let us compute this relative entropy:

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

I added and subtracted a term to write the relative entropy as the sum of two ultraviolet-finite expressions. One term is minus the difference of entanglement entropies between the state Ψ and the vacuum Ω:

$$(\mathrm{Tr}\, \rho \log \rho - \mathrm{Tr}\sigma \log \sigma) = S_\Omega - S_\Psi = -\Delta S.$$

The "area law" divergence cancels out when we take this difference.

To evaluate the other term

$$(\mathrm{Tr}\sigma \log \sigma - \mathrm{Tr}\rho \log \sigma)$$

we remember that if $\mathcal{O}$ is any operator, then $\langle \Psi|\mathcal{O}|\Psi\rangle = \mathrm{Tr}\,\sigma\mathcal{O}$, $\mathrm{Tr}\,\langle\chi|\mathcal{O}|\chi\rangle = \mathrm{Tr}\,\rho\mathcal{O}$. We use this with $\mathcal{O} = -\log \sigma$, and we remember that we showed before that $\sigma = \exp(-2\pi K)$ so

$$-\log \sigma = 2\pi K = 2\pi \int \mathrm{d}x\mathrm{d}x_\perp x T_{00}.$$

So the second term is

$$2\pi\langle\Psi|K|\Psi\rangle - 2\pi\langle\Omega|K|\Omega\rangle = 2\pi \int_0^\infty \mathrm{d}x\mathrm{d}x_\perp \, \langle\Psi|xT_{00}(x, x_\perp)|\Psi\rangle.$$

(With the usual regularization of T_{00}, the term with $\Psi \to \Omega$ vanishes.)

Let us define

$$\mathcal{E} = \int_0^\infty \mathrm{d}x \int_{-\infty}^\infty \mathrm{d}x_\perp \, \langle \Psi | x T_{00}(x, x_\perp) | \Psi \rangle.$$

Casini's insight is that if Ψ is a quantum state that describes a system that intuitively is localized almost entirely in the region $\mathrm{A} : x > 0$, and that has size $\mathcal{R}$ and energy E, then $\mathcal{E} \gtrsim E\mathcal{R}$. He proposed therefore to use the quantity $\mathcal{E}$, which is rigorously defined for all quantum states, as the stand-in for Bekenstein's heuristic $E\mathcal{R}$. He also proposed ΔS, the difference in entanglement entropies between the state Ψ and the vacuum state Ω, to stand in for the heuristic S in the Bekenstein proposal. The positivity of relative entropy says that

$$\mathcal{E} \geq \frac{\Delta S}{2\pi},$$

with equality only if $\Psi = \Omega$. This is Casini's rigorous version of the Bekenstein bound, valid for all quantum states.

In conclusion, I reviewed some of the early developments in black hole thermodynamics from the 1970's, and I hope I have given you at least a taste of how, in modern times, much more insight has come from understanding quantum entropy in a way that makes sense for all states, whether thermodynamics is valid or not. If I had had more time, in the first half of the talk, I would have wanted to describe the Euclidean approach to black hole entropy, by Gibbons and Hawking. In the second half of the talk, I would have wanted to explain A. Wall's proof of the generalized second law, the Ryu-Takayangi holographic formula for entanglement entropy, the notion of entanglement wedge reconstruction and the role of quantum error correction in holography. Alas, not for today!

I chose this topic because I think the modern developments involving the fine-grained entropy are genuinely very exciting and probably pointing toward something really new. I regret that I was really only able to explain the first result in that direction.