Lecture 1

These lecture notes provide somewhat more detailed explanations than those given in the audio recordings. As such, they contain some sections which are entirely absent from the audio recordings. These sections are in italics, allowing readers already familiar with the material to easily bypass them.

1.1 Black Holes

A black hole can be simply thought of as an object so dense that even light cannot escape its gravitational field. Although a complicated mathematical framework is necessary to elucidate this concept in a fully relativistic context, we can demonstrate its basic features by examining two simple analogues.

1.1.1 Dark Stars

The basic idea of a black hole was first proposed by John Michell in 1784 and independently by Laplace in 1796, more than a century before Einstein formulated his theory of General Relativity. Newton's corpuscular theory suggested that even light should be subject to gravitational fields; based on this theory, Laplace and Michell inferred that sufficiently dense stars could have an escape velocity greater than the speed of light. These dense objects, called dark stars, could effectively become invisible to observers on Earth.

Since such an object must have a large density, there is an upper bound on its radius for a given mass. The escape velocity v_E of a body of mass m is defined as the minimum speed required at the surface of the star for the body to escape the star's gravitational field and reach infinity. This means that the body's kinetic energy $\frac{1}{2}mv_E^2$ must exactly cancel its potential energy $-mMG/R$ at the star's surface, so

that the body reaches infinity with zero kinetic energy: that is,

$$\frac{1}{2}mv_E^2 = \frac{mMG}{R}. \tag{1.1}$$

Here M is the mass of the dark star and R is its radius. Since the dark star must have an escape velocity greater than the speed of light c, the upper bound on its radius can be found by rearranging the above equation to find

$$R_s = \frac{2MG}{c^2}. \tag{1.2}$$

Remarkably, this formula remains valid even in General Relativity, where it defines the radius of the event horizon of a Schwarzschild black hole. For a sense of the density of a dark star, consider that $R_s \approx 3$km for a star with the mass of the Sun, while the actual radius of the Sun is approximately 7×10^5km.

We should note that the analogy between dark stars and black holes breaks down when we consider the behaviour of outgoing light. In the case of a dark star, light emitted from a distance $r < R_s$ can travel beyond R_s before reaching a turning point and returning to the star. But in the case of a black hole, light emitted at $r < R_s$ cannot even momentarily pass the event horizon.

However, dark stars do yield at least one more useful result, which is the strength of their tidal fields. A body of length Δx will experience a tidal force (*i.e.* a difference between the forces at each of its ends) over that length. From the expansion $F(r + \Delta x) \approx F(r) + \frac{\partial F}{\partial r}\big|_r \Delta x$, we see that the tidal force is $\frac{\partial F}{\partial r}\big|_r \Delta x$. In the case of the Newtonian force $F_g = -mMG/r^2$, this means that the tidal acceleration at $r = R_s$ is

$$\left.\frac{2GM}{r^3}\right|_{r=R_s} \Delta x = \frac{1}{R_s^2}\frac{2MG}{R_s}\Delta x = \frac{c^2}{R_s^2}\Delta x. \tag{1.3}$$

This result also holds for a radially-infalling object at the event horizon of a black hole. Note that the tidal force is larger for smaller black holes. For example, a person who is roughly 1m tall falling into a very small black hole with a Schwarzschild radius $R_s \sim 10^4$m would experience a tidal force of roughly one billion newtons. Obviously, the observer would be torn apart before even arriving at the event horizon. However, an observer would not even notice the tidal fields at the horizon of a supermassive black hole, such as the one with $R_s \sim 10^5$km which is thought to reside at the center of the Milky Way; unfortunately, the observer would inevitably arrive at the singularity roughly a second later.

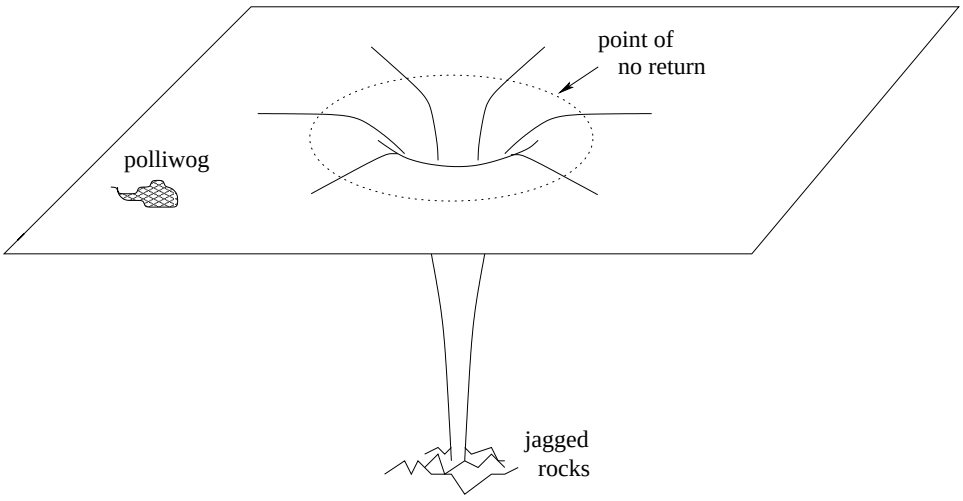

Figure 1.1: The tragedy of the polliwogs: a physical realization of a 'dumb hole'.

1.1.2 Dumb Holes

A more recent analogue for black holes was devised by Bill Unruh in 1981. Rather than a black hole from which light cannot escape, he considered a 'dumb hole' from which sound cannot escape. Imagine an infinitely large lake with a drain at its center. This drain causes all the water in the lake to flow radially inward. As water approaches the drain it speeds up, and at some radial distance from the center of the lake it surpasses the speed of sound. We call this distance the point of no return (PNR); it bears an obvious resemblance to the event horizon of a black hole.

Now imagine a community of polliwogs residing in this lake. These polliwogs are blind, so they can only communicate via sound, and they have strict laws (either physical or societal) that prevent them from traveling faster than the speed of sound. This means that if they swim past the PNR, they enter a region in which the water flows faster than they can swim or send any communication. Because they are only aware of their motion relative to the water, the polliwogs passing the PNR notice nothing strange (assuming that they are small enough to ignore tidal forces).

Consider a polliwog couple enjoying an evening out. The first polliwog, Bob, takes a seat that is fixed relative to the bottom of the lake, while the second, Alice, moves with the flowing water. Bob is called a fiducial observer (Fido); he has a fixed position. Alice is called a freely-falling observer (Frefo); she is not aware of the flow of the water, as she moves freely along with it.

Now, Alice is an accomplished singer, and she decides to enhance the mood by

serenading Bob. But as she does so, she moves ever closer to the PNR. Because of the Doppler effect, Bob hears Alice's tone lower, until it reaches zero frequency when Alice passes the PNR; this is precisely equivalent to the gravitational redshift of light traveling away from a black hole. Additionally, as Alice approaches the PNR, her voice takes an increasingly long time to reach Bob. At the PNR, her voice travels at the same speed as the water, meaning that it is motionless relative to Bob and will never reach him; at every later point, Alice's voice actually travels more slowly than the surrounding water and moves toward the drain. Thus, Bob hears Alice take an infinite time to reach the PNR, hears her singing (*i.e.* the rate of passing time for her) get ever slower, and will never hear anything that she sung after passing the PNR. But as far as Alice is concerned, nothing particularly strange has occurred. Even after she passes the PNR, she continues to hear Bob as usual. Only when she reaches the drain itself does her experience dramatically alter, as she ceases to experience anything at all.

The entirety of this description carries over to the case of a black hole, and it might be helpful to keep in mind in later sections.

1.2 The Schwarzschild Black Hole

In the remainder of this lecture we will examine the geometric structure of Schwarzschild black holes, introducing several useful coordinate systems in the process. Although Schwarzschild black holes are unlikely to exist in nature, they have all the features of more general black holes that are of interest to us. Specifically, they are perfectly suitable for studying information, entropy, Hawking radiation, and the Quantum Mechanics of black holes. A possible problem with studying Schwarzschild black holes is that they have a negative specific heat, heating up as they radiate and cooling down as they receive energy, meaning that they cannot exist in thermal equilibrium with a reservoir; but this should not overly concern us, since the Sun, and even atomic nuclei, have the same property.

Now, for convenience we will set $c = 1$ in what follows. We will also write the metric in terms of a proper time element rather than a spacetime line element, using the positive-time signature for the metric. In Schwarzschild coordinates the metric is described by

$$d\tau^2 = \left(1 - \frac{2MG}{r}\right) dt^2 - \left(1 - \frac{2MG}{r}\right)^{-1} dr^2 - r^2 d\Omega^2. \tag{1.4}$$

The coordinate t here refers to the time as measured by a Fido at infinity; for a given t, the surface defined by constant r is a sphere of area $4\pi r^2$; and $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$

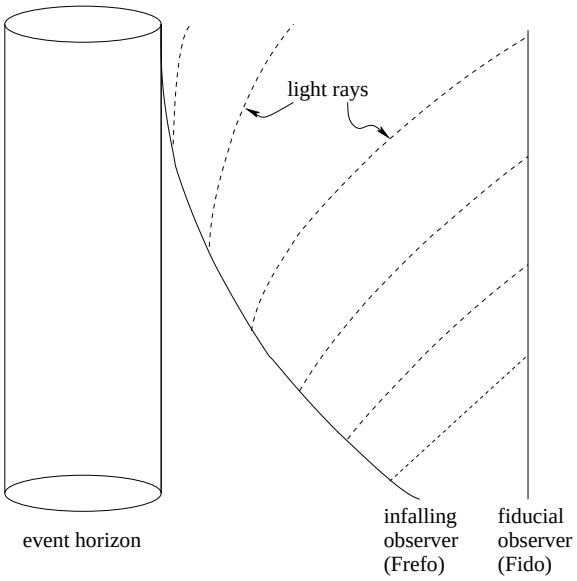

Figure 1.2: A freely-falling observer (a Frefo) emits light signals to a fiducial observer (a Fido) far from the black hole. As the Frefo approaches the event horizon, his signals are pulled toward the black hole and take ever longer to reach the Fido. (Schwarzschild time runs upward in this diagram.)

defines the metric on that sphere. Thus, at fixed times (as measured at infinity) the spacetime consists of concentric spheres of area $4\pi r^2$. However, we can note three more interesting features of this metric. The first is the factor in front of dr^2, which implies that the spatial distance between two concentric spheres with radial coordinates r_1 and r_2 is not $r_2 - r_1$ but instead $\int_{r_2}^{r_1} \sqrt{-g_{rr}} dr$; this is caused by spacetime curvature, and it means that π is not the ratio of circumference to diameter in this spacetime. The second feature is the divergence of the metric at both $r = 2MG = R_s$ and $r = 0$. And the third is the reversal in the signs of the dt^2 and dr^2 terms for $r < 2MG$. We will clarify the last two issues below.

If we consider a Frefo falling radially from rest at $r = R$, we can easily find the equation of motion $\left(\frac{dr}{d\tau}\right)^2 = \frac{R_s}{r} - \frac{R_s}{R}$. Integrating this, we find that the total proper time for the observer to reach $r = 0$ is

$$\tau = \frac{\pi}{2} R \sqrt{\frac{R}{R_s}}. \tag{1.5}$$

The time is perfectly finite for finite R, and the geodesic encounters no obstacles at

$r = R_s$. However, we see from Eq. (1.3) that the tidal forces on the observer diverge at $r = 0$. These facts indicate that the singularity at $r = R_s$ is merely a coordinate singularity, an artifact of our choice of coordinates, while $r = 0$ is a true geometric singularity. However, the singularity at the Schwarzschild radius $r = R_s$ also implies a vanishing coefficient of dt^2 in the metric. This means that as $r \to R_s$ there is an infinite time interval as measured at infinity for any finite proper time interval measured locally, so a Fido never sees the Frefo pass the coordinate singularity. More formally, for a light ray we have $d\tau = 0$, and for a radial path we have $d\Omega = 0$; thus, for a radial light ray we can find $\frac{dr}{dt}$ from Eq. (1.4). The result is

$$\frac{dr}{dt} = 1 - \frac{R_s}{r}, \tag{1.6}$$

which vanishes as $r \to R_s$, meaning that light at R_s has zero coordinate velocity and cannot escape the black hole. This suggests that $r = R_s$ marks the black hole's event horizon, as predicted in the dark star model.

Furthermore, the changes in sign in the metric for $r < R_s$ mean that intervals of r correspond to intervals of time and intervals of t correspond to intervals of space. This suggests that our Frefo passing the event horizon (who must always follow a time-like geodesic) will inevitably move toward decreasing values of r, supporting our supposition that $r = R_s$ is in fact the event horizon. (However, note that the Schwarzschild coordinates are only formally valid for $r > R_s$, and must be analytically continued within the event horizon. This analytical continuation does not preserve the geometric interpretation of the coordinates, meaning that we no longer have concentric spheres and times as measured at infinity. Most importantly, r cannot be interpreted as a radial coordinate behind the horizon.)

1.2.1 Specialized coordinates

We can clarify the above conclusions by analyzing the Schwarzschild geometry in several specialized coordinate systems.

1. Tortoise Coordinates

A useful class of coordinate systems are called *conformally flat* coordinates. A conformally flat spacetime is one with a metric that can be brought into the form $d\tau^2 = f(x^\mu)((dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2)$, where the function $f(x^\mu)$ is called the conformal factor and the coordinates x^μ are called conformally flat coordinates. Such coordinates maintain many useful features of flat spacetime: particularly, the

45^o angle of light rays in spacetime diagrams and the form of the 2D wave equation, in which the conformal factor conveniently disappears (as we shall see in the next lecture).

Now, all 2D geometries are conformally flat, so we can find conformally flat coordinates for the r-t plane in Schwarzschild spacetime. Taking the angular coordinates to be constant and defining $f(r) = 1 - \frac{2MG}{r}$, we can write the r-t part of the metric as

$$d\tau^2 = f(r)\left(dt^2 - \frac{1}{f(r)^2}dr^2\right) \equiv f(r)\left(dt^2 - (dr^*)^2\right). \tag{1.7}$$

The new coordinate r^* is defined by the differential equation $dr/f(r) = dr^*$, which has the solution

$$r^* = r + 2MG\ln\left(\frac{r - 2MG}{2MG}\right). \tag{1.8}$$

This radial coordinate satisfies

$$r^* \rightarrow r \text{ for } r \gg 2MG \tag{1.9}$$
$$r^* \rightarrow -\infty \text{ as } r \rightarrow 2MG \tag{1.10}$$
$$r^* = 4MG \text{ when } r = 4MG, \tag{1.11}$$

and the full metric is given by

$$d\tau^2 = \left(1 - \frac{2MG}{r}\right)\left(dt^2 - dr^{*2}\right) - r^2 d\Omega^2. \tag{1.12}$$

2. Near-horizon Coordinates (Rindler space)

We are often interested in examining the behavior of particles or waves near the horizon. To do this, we examine a locally flat region around an arbitrary point on the horizon. An interval of proper distance $d\rho$ is related to dr by $d\rho^2 = g_{rr}dr^2$, and the total proper distance from the horizon to a radial distance r is given by

$$\rho = \int_{2MG}^{r} \sqrt{-g_{rr}}dr \approx 2\sqrt{2MG(r - 2MG)}, \tag{1.13}$$

where we have used the fact that very near the horizon $g_{rr} \approx \left(\frac{2MG}{r-2MG}\right)$. In terms of ρ, the metric becomes

$$d\tau^2 = \rho^2 d\omega^2 - d\rho^2 - r^2(\rho)d\Omega^2, \tag{1.14}$$

where we have used $f(r) \approx \frac{\rho^2}{\rho^2+(4MG)^2} \approx \frac{\rho^2}{(4MG)^2}$ and defined the dimensionless time $\omega = \frac{t}{4MG}$.

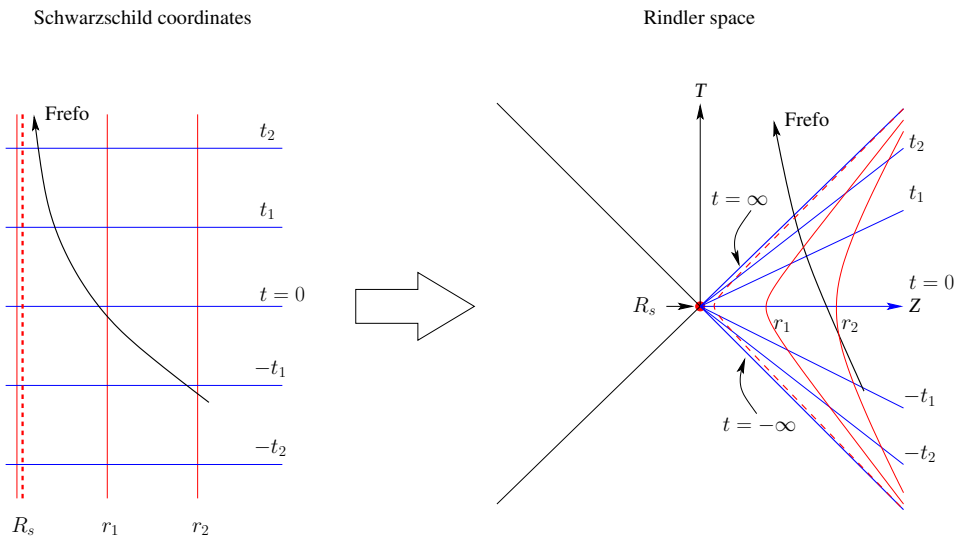

Figure 1.3: The approximately flat spacetime immediately outside the event horizon is mapped from Schwarzschild coordinates to Rindler space. Lines of constant radial distance are in red and lines of constant time are in blue. The dashed red line is at a constant radius just slightly outside the horizon. A Frefo's path starting at an arbitrary position is shown to cross the event horizon at infinite Schwarzschild time but obviously finite proper time.

Since the spacetime is spherically symmetric, we can pick our arbitrary point on the horizon at $\theta = 0$. *In this case the usual Cartesian spatial coordinates* x *and* y *reduce to* $x = 2MG\theta\cos\phi$ *and* $y = 2MG\theta\sin\phi$ *for small* ϕ. *Ignoring terms of order* $\rho^2\theta$ *from the expansion of* r^2, *our metric reduces to*

$$d\tau^2 = \rho^2 d\omega^2 - d\rho^2 - dx^2 - dy^2. \tag{1.15}$$

This is flat spacetime in cylindrical hyperbolic coordinates, the pseudo-Riemannian equivalent of cylindrical polar coordinates. It can be transformed into the Minkowskian form $d\tau^2 = dT^2 - dZ^2 - dx^2 - dy^2$ via the coordinate transformations $Z = \rho\cosh\omega$ and $T = \rho\sinh\omega$. The Rindler coordinates only cover a wedge (called the Rindler wedge) of this full Minkowski spacetime. In the context of a Schwarzschild black hole, this wedge corresponds to a local portion of spacetime lying on 'our' side of the horizon, and we call it *Rindler space*.

In the T-Z plane we have the event horizon at the origin. Lines of constant ω (or t) radiate radially outward, cutting across hyperbolae of constant ρ (or r) (see

Fig. 1.3), completely analogous to lines of constant ϕ and circles of constant r in the usual polar coordinates. Each curve of constant ρ corresponds to the local Lorentz frame of a "Rindler observer". These observers are Fidos sitting at a fixed radial distance from the horizon, maintaining a constant acceleration to counteract gravity; this is equivalent to the frame of an accelerated observer in flat space, as we would expect from the equivalence principle.

The lines $T = \pm Z$ correspond to $t = \pm\infty$, and lines of constant time accumulate densely around them. Since light rays travel at 45^o, and those originating from the event horizon stay on it, we can see that the lines $T = \pm Z$ also correspond to the future and past horizon. This makes clear that any timelike observer will pass through the horizon at $t = \infty$ but after finite proper time.

3. Kruskal-Szekeres Coordinates

We now seek global coordinates that reveal the entire structure of the spacetime. Keeping in mind the form of the Rindler coordinates, we propose a global metric

$$d\tau^2 = F(R)\left[R^2 d\omega^2 - dR^2\right] - r^2(R)d\Omega^2, \tag{1.16}$$

where $F(r)R^2 \to \rho^2$ as $r \to 2MG$. Comparing this with the metric in t-r coordinates, we find that

$$\frac{F(R)R^2}{(4MG)^2} = 1 - \frac{2MG}{r} \tag{1.17}$$

$$F(R)dR^2 = \left(1 - \frac{2MG}{r}\right)^{-1} dr^2. \tag{1.18}$$

From the first equation we find that $F(R) = \left(1 - \frac{2MG}{r}\right)\left(\frac{4MG}{R}\right)^2$. Substituting this into the second equation gives

$$\frac{4MG}{R}dR = \left(1 - \frac{2MG}{r}\right)^{-1} dr, \tag{1.19}$$

which we integrate to find $R = C\exp\left(\frac{r^*}{4MG}\right)$. The integration constant is chosen such that $|R| = MG$ at $r = 0$, so

$$R = MG\exp\left(\frac{r^*}{4MG}\right) \tag{1.20}$$

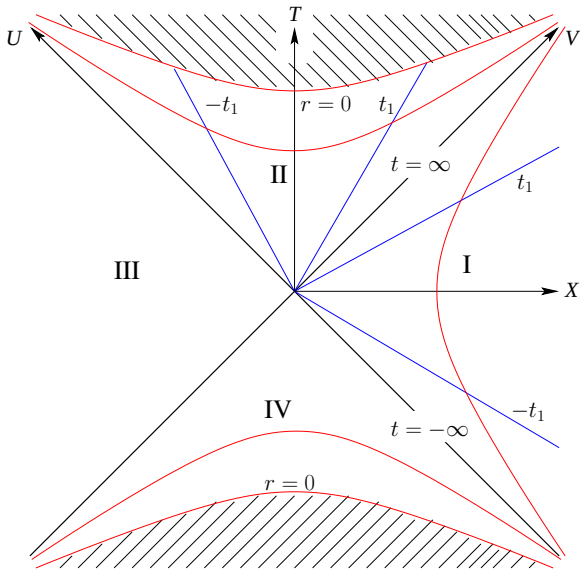

Figure 1.4: Schwarzschild spacetime in Kruskal coordinates, revealing four distinct regions. Curves of constant Schwarzschild time are again shown in blue and curves of constant r in red. The spacelike and timelike nature of these curves is evidently reversed behind the horizon. The shaded region lies 'beyond' the singularity and is not part of the spacetime.

and

$$\begin{aligned} F(R) &= 16\left(1-\frac{2MG}{r}\right)\exp\left(\frac{-r^*}{2MG}\right) \\ &= \frac{32MGe^{-r/2MG}}{r}. \end{aligned} \tag{1.21}$$

Near the horizon these results reduce to $R \approx \rho/4$ and $F(R) \approx 16$, so $F(R)R^2 \approx \rho^2$ as required. Furthermore, they are non-singular at the horizon. However, note that in terms of r we have $R = \sqrt{\frac{r}{2M}-1}MGe^{r/4MG}$, so R becomes imaginary and R^2 becomes negative for $r < 2MG$. Thus, we still have the flip in spatial and radial coordinates behind the horizon.

Now, we can introduce null Kruskal coordinates U and V, such that light travels along lines of constant U or V, via the transformations

$$U = -Re^{-\omega} \tag{1.22}$$

$$V = Re^{\omega}. \tag{1.23}$$

The metric in these coordinates takes the form

$$d\tau^2 = F(R)dUdV - r^2 d\Omega^2. \tag{1.24}$$

The axes $U = 0$ and $V = 0$ then correspond to $t = \infty$ and $t = -\infty$, with the origin still at $r = 2MG$. We can put the time-radial plane into conformally flat form by introducing the Kruskal-Szekeres coordinates $T = \frac{1}{2}(U + V)$ and $X = \frac{1}{2}(V - U)$, resulting in the metric

$$F(R)(dT^2 - dX^2) - r^2 d\Omega^2. \tag{1.25}$$

The Kruskal coordinates are well-behaved at the horizon, but give a singular metric at the true singularity $r = 0$. A plot in the T-X plane reveals the full structure of the Schwarzschild geometry, as shown in Fig. 1.4. This is called the maximal extension of Schwarzschild. Region I is 'our' space outside the horizon, and region II is the spacetime 'behind' the horizon. We also see an entirely distinct region of space (region III) that is causally disconnected from region I, and a time-symmetric reflection of region II called a 'white hole'(region IV). In the next lecture we will see that these last two regions are usually physically irrelevant. The diagram also shows clearly that the singularity at $r = 0$ is actually a spacelike curve, such that it marks the end of timelike curves instead of following a timelike curve itself. The exact curves describing the singularity are easily seen to be $T^2 - X^2 = (MG)^2$, or $R^2 = -1$, giving the two solutions pictured. This differs significantly from the popular-science image of the singularity as a point in space. The analytic continuation of the lines of constant t and r makes the source of this nonintuitive result clear, since the roles of the external space and time coordinates are reversed behind the horizon.

Lecture 2

2.1 Review: coordinate systems

	polar coordinates	conformal coordinates
near horizon	$d\tau^2 \simeq \rho^2 d\omega^2 - d\rho^2$	$d\tau^2 \simeq e^{2u}(d\omega^2 - du^2)$
	$\omega = \frac{t}{4MG}$, ρ is proper distance	$d\rho = \rho du$, $\rho = e^u$
global	$d\tau^2 = F(R)[R^2 d\omega^2 - dR^2]$	$d\tau^2 = \left(1 - \frac{2MG}{r}\right)(dt^2 - dr^{*2})$
	$R = MGe^{\frac{r^*}{4MG}}$	$r^* = r + 2MG \ln\left(\frac{r-2MG}{2MG}\right)$

Table 2.1: In all cases the angular part $r^2 d\Omega^2$ of the metric remains unchanged.

Schwarzschild coordinates (t, r) are useful to describe events far away from the event horizon. r is a radial coordinate such that spheres of constant r have surface area $4\pi r^2$. t is associated with the time translation symmetry of the spacetime and corresponds to proper time far away from the event horizon.

Polar (hyperbolic) coordinates use coordinates (ω, R) that can be thought of as hyperbolic angular and radial coordinates of a space which is conformal to $1+1$ dimensional Minkowski space. Near the horizon ω is rescaled Schwarzschild time t, and $\rho \approx R$ is a proper distance. The global coordinates are extended versions of the near horizon coordinates, where ω is still a hyperbolic angle.

Conformal coordinates are particularly well suited to describe the causal structure of spacetime since null geodesics are easy to describe in conformal coordinates. To go from polar to conformal coordinates we define a new radial

coordinate, such that the coefficient in front of it is identical to the one in front of the timelike displacement. The range of the conformal coordinates is $-\infty < r^* < \infty$, with $r^* = -\infty$ at the event horizon and $r^* \approx r$ far away from it. *The near horizon conformal coordinates ω, u were not yet introduced. They are defined analogously to the tortoise coordinates by defining u via $d\rho = \rho du$.*

2.2 Carter-Penrose diagrams

2.2.1 Minkowski spacetime

Idea: take all of spacetime and do a coordinate transformation that compactifies the coordinates, such that all of spacetime can be drawn on one diagram while preserving the light cone structure.

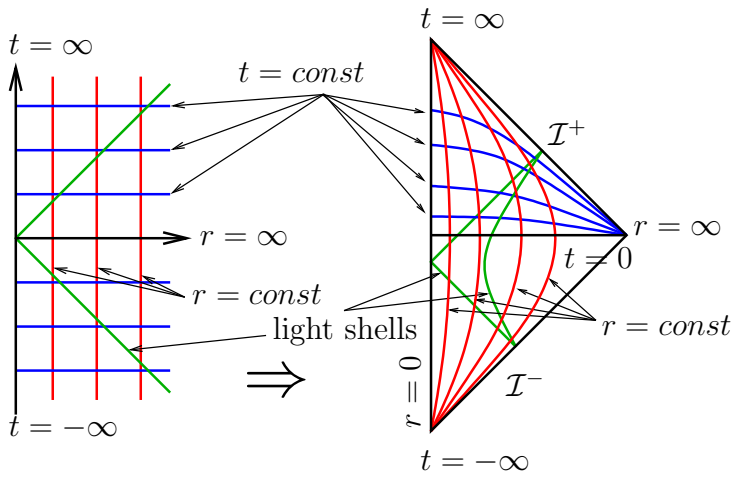

Figure 2.1: Penrose diagram of Minkowski space. red: lines of constant radius, blue: lines of constant time, green: light rays, $\mathcal{I}^-$: past null infinity, $\mathcal{I}^+$: future null infinity

We define new coordinates $Y^\pm = F(T^\pm)$, where $T^\pm = t \pm r$. These coordinates are based on the structure of the light cones $T^\pm = const$. By construction they preserve the light cone structure. We can use many different transformations to map the domain $0 < r < \infty$, $-\infty < t < \infty$ to a finite region of the plane. We choose

$$Y^\pm = \tanh(T^\pm). \tag{2.1}$$

2.2.2 Schwarzschild spacetime

The diagram in Fig. 2.2, which was already shown in Lecture 1, shows a Kruskal diagram of the maximally extended Schwarzschild spacetime. Regions I and III are asymptotically flat, infinite spaces (so there is "more" space than for flat spacetime). Observers in region I and III can send signals into region II, but they cannot communicate with each other. because of this, the spacetime is said to contain a non-traversable wormhole.

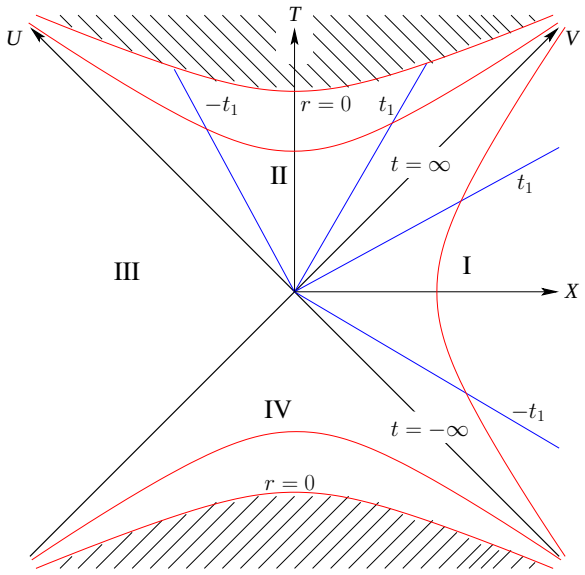

Figure 2.2: Schwarzschild spacetime in Kruskal coordinates with lines of constant t in blue and lines of constant r in red.

We recall the null Kruskal coordinates from Lecture 1

$$V = Re^{\omega} \qquad U = -Re^{-\omega}. \tag{2.2}$$

These coordinates cover all regions (I–IV) of the spacetime, but they are not yet compactified. Rather than using $\tanh(\cdot)$ to compactify the coordinates, we use $\arctan(\cdot)$ and define

$$Y^{+} = \arctan(V) \text{ and } Y^{-} = \arctan(U), \tag{2.3}$$

which maps the full Kruskal extension of Schwarzschild spacetime into the range $-\frac{\pi}{2} < Y^{\pm} < \frac{\pi}{2}$.

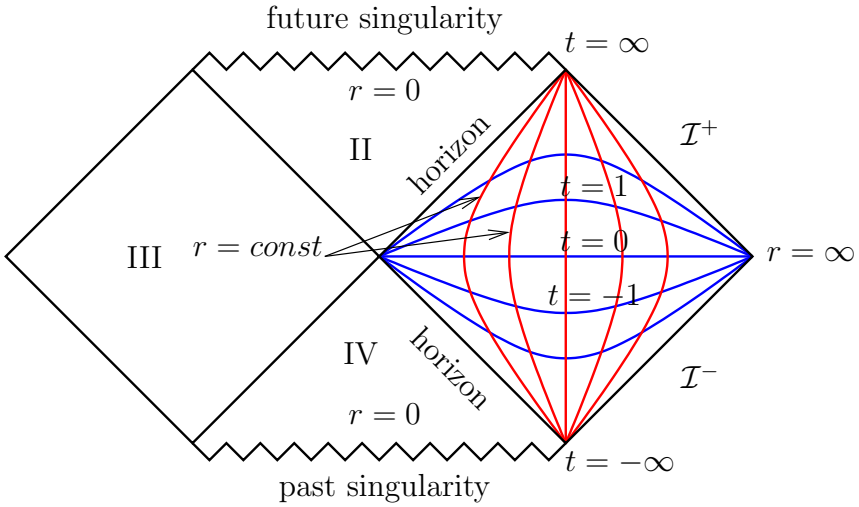

Figure 2.3: Penrose diagram for the Schwarzschild spacetime. The regions I – IV of Fig. 2.2 are shown. *Just as for the Kruskal diagram, an observer in region II is forced toward the singularity. Communication between regions I and III is impossible.*

(Radial) light rays move along straight lines of slope ± 1 in the Penrose diagram. A light ray sent radially outward by an observer in region I will escape the singularity and arrive at future null infinity $\mathcal{I}^+$. An inward travelling light ray on the other hand will cross the horizon and enter region II. Once there, it can never move back into either region I or III and will eventually encounter the future singularity. Similarly, light rays and any material observers entering region II from region III will also encounter the future singularity. Regions II and III are behind the future horizon of region I, while regions II and IV are behind the past horizon. Regions behind the past horizon are not relevant for black holes formed from collapsing matter.

2.3 Collapse of a spherical shell of light

In Newtonian gravity it is well known that the gravitational field inside a spherical shell of matter is zero, and seen from the outside the shell's gravitational field is that of a point mass at its centre.

Similarly, in General Relativity Birkhoff's theorem states that inside of a spherical shell of matter the spacetime is flat, while outside it is described by the Schwarzschild metric.

A Penrose diagram for a collapsing spherical shell of matter can thus be patched together using pieces of the Penrose diagrams for flat spacetime and the Schwarzschild

solution. We begin by drawing a shell of photons in the Penrose diagram of flat spacetime.

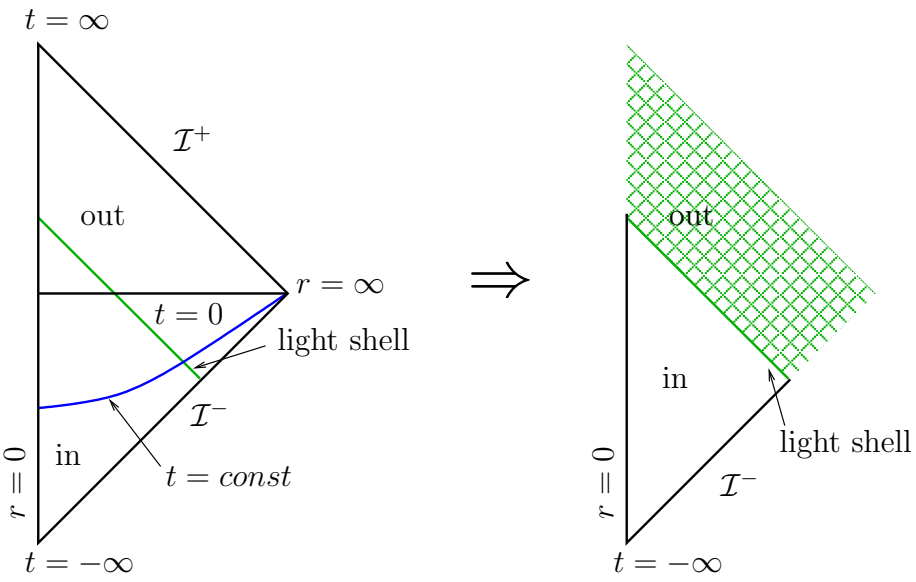

Figure 2.4: Penrose diagram of flat spacetime containing a spherical shell of photons. Birkhoff's theorem asserts that the metric in the region inside of the shell is still the Minkowski metric. Outside of the shell however, the metric is no longer Minkowski and we remove this part from the diagram.

Inside the shell of photons, Birkhoff's theorem asserts that the metric is flat, but outside the shell this is wrong. We cannot use the Penrose diagram there, and we have to remove this portion.

On the other hand, when drawing a spherical shell of matter in a Penrose diagram for a Schwarzschild black hole, the region outside of the shell of photons is correct, but the inside has to be removed

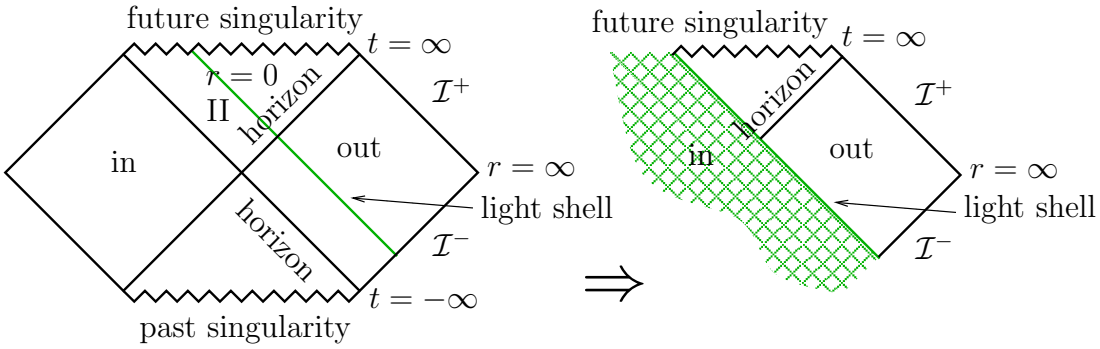

Figure 2.5: Penrose diagram of Schwarzschild spacetime containing a spherical shell of photons. Unlike the situation in Figure 2.4 Birkhoff's theorem asserts that the metric outside of the shell is correctly given by the Schwarzschild metric, while the inside piece has to be discarded.

Stitching the two Penrose diagrams together at the location of the shell of photons results in a Penrose diagram describing the complete spacetime of an infalling shell of matter (after a conformal transformation to match the radial coordinates across the boundary) .

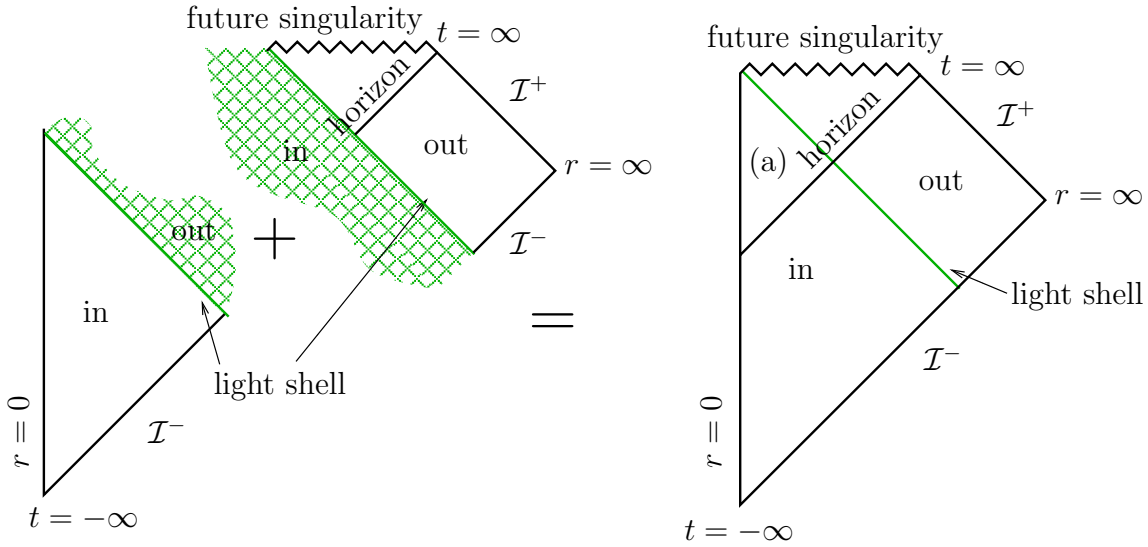

Figure 2.6: Full Penrose diagram of the spacetime generated by an infalling shell of photons.

The event horizon extends into the region of the diagram where spacetime is flat.

To an observer within the shell of light, nothing indicates that she is within the forming event horizon. The horizon is a global notion that depends on the future behaviour of the shell: It is simply the mathematical surface that separates the region in which escape to $r = \infty$ is possible from that where this is not possible (and where an observer will necessarily encounter the singularity). Thus an observer in region (a) will not be able to escape, even though the spacetime is perfectly flat there.

Different observers will describe the formation of the event horizon quite differently. Consider a free falling observer (Frefo) and a stationary, fiducial observer (Fido). The Fido never sees the Frefo cross the event horizon. Instead she sees the Frefo slow down as she approaches the horizon. The Fido can only see events within her past light cone, and as is clear from Fig 2.7, cannot see inside of the horizon. Hence she never sees her freely falling analogue cross the horizon.

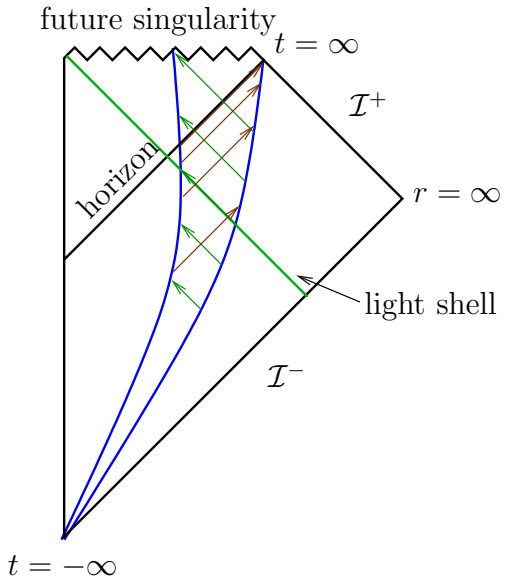

Figure 2.7: The formation of a black hole as seen by a Frefo and a Fido. Signals from the infalling observer arriving at the Fido at constant intervals have to be sent out with ever decreasing intervals by the Frefo.

This might suggest that the Frefo in turn sees the Fido speed up as she comes close the the horizon. This, however, is not true: to her, everything outside of the event horizon looks perfectly normal until she ends her journey at the singularity. No paradox is involved in this, as the Frefo simply does not get to see all of spacetime before encountering the singularity.

A different, but sometimes useful description of the formation of the event horizon is given in Fig. 2.8.

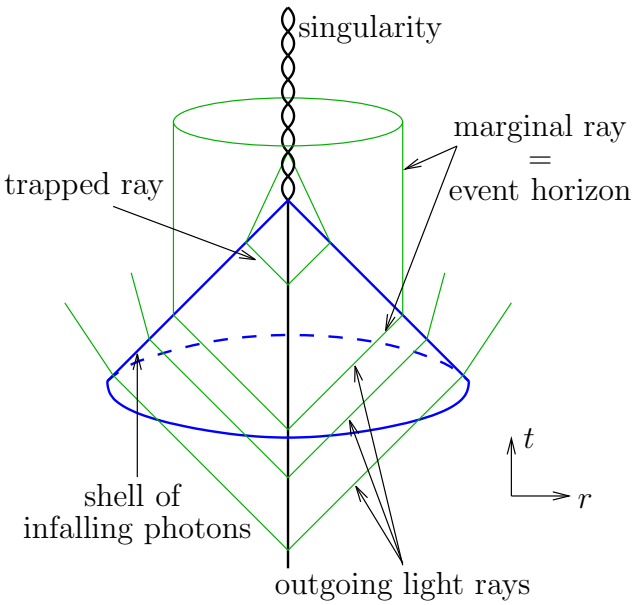

Figure 2.8: A different visualization of an infalling shell of photons. As the photons converge toward the origin an observer there sends out light signals. When a signal crosses the shell of photons, the light ray is deflected. At first the deflection is very small, but as the size of the shell approaches its Schwarzschild radius the signal is reflected back to the origin.

As the shell of photons collapses, radial light rays emanating from the centre of the shell are bent when they cross the shell. Initially the deflection angle is small, but it grows as the shell shrinks. At one point the outgoing light rays will be deflected straight upward, *i.e.* they no longer move outward. Even later outgoing light rays are deflected back to the centre—an event horizon has formed. The marginal light rays which are deflected vertically define the event horizon, which starts growing when these rays are emitted from the centre. This is actually before the size of the shell of photons has shrunk to the Schwarzschild radius. The diagram in Fig. 2.8 is misleading, however, in that it suggests that the singularity is timelike *i.e.* avoidable by an observer, when the correct description reveals the singularity to be spacelike *i.e.* made up of the endpoints of timelike curves.

2.4 Momentum increase near the horizon

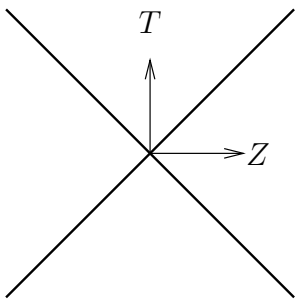

Figure 2.9: Local Cartesian coordinate system

A freely falling observer (or in fact any observer) erects a Cartesian flat coordinate system (T, x, y, Z) in his vicinity. A particle falling toward the black hole passing by the Frefo therefore has constant momentum and moves on a straight line in these coordinates. Since at any given moment the frames of a Fido and a Frefo at the same location are connected by a Lorentz boost, we have

$$Z = \rho \cosh \omega \text{ and } T = \rho \sinh \omega, \tag{2.4}$$

and $\partial_\rho = \cosh \omega \partial_Z + \sinh \omega \partial_T$. ∂_ρ is the momentum conjugate to ρ and since ρ measures proper distance away from the horizon, ∂_ρ is also the momentum measured in proper units by the Fido. A particle having momentum components P_Z and P_T in the frame of the Frefo has a momentum of

$$P_\rho = \cosh \omega P_Z + \sinh \omega P_T \tag{2.5}$$

as measured by the Fido. In the simplest case the particle moves along with the Frefo (not moving in Z) and $P_Z = 0$. For late times ($\omega \gg 1$)

$$P_\rho = \frac{1}{2} P_T e^{\omega} = \frac{1}{2} P_T e^{\frac{t}{4MG}}, \tag{2.6}$$

i.e. a Fido will observe the momentum of the particle increase exponentially as it approaches the horizon. Analogously, the proper distance from the horizon decreases exponentially with time:

$$\rho(t) \propto e^{-\frac{t}{4MG}}, \tag{2.7}$$

since $Z = Z_0 = const$ in the Frefo frame and $Z = \rho \cosh \omega \approx \frac{1}{2} \rho e^{\frac{t}{4MG}}$.

Thus, the particle hugs the event horizon more and more closely while gaining more and more momentum, which eventually drives its dynamics to trans-Planckian scales.

2.5 Classical field theory near the horizon

Before turning to field theory in Schwarzschild spacetime, consider field theory in Rindler coordinates. While really just Minkowski spacetime, the Rindler metric bears close resemblance to the Schwarzschild metric.

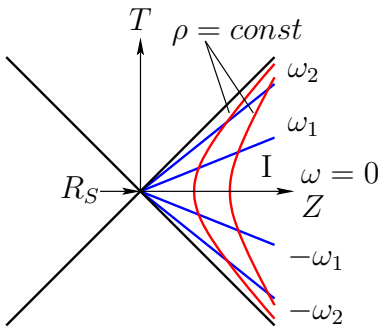

Figure 2.10: Surfaces of constant time and position in Rindler space

Signals entering region I from region IV have to pass through the horizon, so they appear as initial data. Therefore anything that happens within region I can be traced back to an initial condition at $\omega = -\infty$. This means that Rindler space is (classically) self-contained. Recalling the conformal coordinates (ω, u) near the horizon, we write the metric as

$$d\tau^2 = e^{2u}(d\omega^2 - du^2) - d\vec{x} \cdot d\vec{x} \tag{2.8}$$

where $d\vec{x} = (dx, dy)$ is the coordinate displacement tangential to the horizon. The action for a scalar field is given by the usual expression

$$I = \frac{1}{2}\int du\, d\omega\, dx\, dy \sqrt{-g} g^{\mu\nu} \partial_\mu \chi \partial_\nu \chi \tag{2.9}$$

$$= \frac{1}{2}\int du\, d\omega\, dx\, dy \left[\left(\frac{\partial \chi}{\partial \omega}\right)^2 - \left(\frac{\partial \chi}{\partial u}\right)^2 - e^{2u}(\partial_\perp \chi)^2 \right], \tag{2.10}$$

where $\partial_\perp = (\partial_x, \partial_y)$, $\sqrt{-g} = e^{2u}$ and $g^{\mu\nu} = \mathrm{diag}(e^{-2u}, -e^{-2u}, -1, -1)$. All of the interesting physics described by this action is contained in the derivatives along the horizon, as this is the only term having a non-constant coefficient. As the x and y directions are flat, we can use a Fourier analysis in the (x, y) subspace. Expanding χ as

$$\chi = \sum e^{i\vec{k}\cdot\vec{x}} \chi_{\vec{k}}(\omega, u), \tag{2.11}$$

the action decouples for individual $\vec{k}$ values and we find

$$I_k = \frac{1}{2}\int d\omega\, du \left[\left(\frac{\partial \chi_k}{\partial \omega}\right)^2 - \left(\frac{\partial \chi_k}{\partial u}\right)^2 - k^2 e^{2u} \chi_k^2 \right]. \tag{2.12}$$

For each value of k this describes a non-interacting scalar field with a potential (or mass) term of $k^2 e^{2u}$. The potential barrier grows exponentially as we move away from the horizon, with only the $k = 0$ mode experiencing no barrier. Thus the $k = 0$ mode alone can escape to future infinity.

The action Eq. (2.12) *is conformally invariant with a weight of zero. This is seen most easily by considering the equation of motion resulting from the action*

$$g^{\mu\nu} \nabla_\mu \nabla_\nu \chi = 0 \Leftrightarrow (\sqrt{-g})^{-1} \partial_\nu (\sqrt{-g} g^{\mu\nu} \partial_\mu \chi) = 0, \tag{2.13}$$

where $\mu, \nu \in \{\omega, u\}$. A conformal transformation $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$ transforms the metric determinant $\sqrt{-\tilde{g}} = \Omega^2\sqrt{-g}$ and the inverse metric $\tilde{g}^{\mu\nu} = \mathrm{diag}(\Omega^{-2}e^{-2u}, -\Omega^{-2}e^{-2u})$. The two effects cancel one another, so the conformal factor does not appear in the equation of motion.

From the action (2.12) we obtain the equation of motion

$$\frac{\partial^2\chi}{\partial\omega^2} - \frac{\partial^2\chi}{\partial u^2} + k^2 e^{2u}\chi = 0, \tag{2.14}$$

which for the Ansatz $\chi = e^{i\nu\omega}\psi$ yields the time independent "Schrödinger" equation

$$-\frac{\partial^2\psi}{\partial u^2} + k^2 e^{2u}\psi = \nu^2\psi \tag{2.15}$$

for a mode of Rindler frequency ν. ν^2 corresponds to the energy of the mode. High energy modes will be able climb the potential hill further before being reflected back to the horizon.

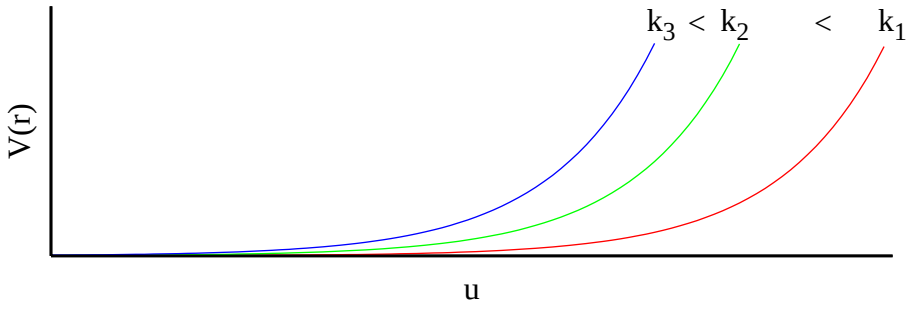

Figure 2.11: Potential barrier in Rindler space. Each mode mode with $k > 0$ experiences a potential barrier preventing it from escaping.

2.6 Field theory in Schwarzschild spacetime

We write down the metric in tortoise (t, r^*) coordinates

$$d\tau^2 = F(r^*)[dt^2 - dr^{*2}] - r^2 d\Omega^2 \tag{2.16}$$

and consider a minimally coupled scalar field χ, with the equation of motion

$$g^{\mu\nu}\nabla_\mu\nabla_\nu\chi = 0. \tag{2.17}$$

It is conventional to define a new field $\psi = r\chi$ to take out a measure of the surface of a sphere. As Schwarzschild spacetime is spherically symmetric, the appropriate decomposition is into spherical harmonic modes $\psi_{\ell m}$ using

$$\psi = \sum_{\ell m} \psi_{\ell m} Y_{\ell m}. \tag{2.18}$$

The connection between the spherical mode ℓ and the wave number k in the case of Rindler space is the usual one: Angular momentum $\vec{L}$ is $\vec{L} = \vec{r} \times \vec{p}$, so a wave with wave number k has angular momentum $\ell = kR_S = 2MGk$ (using $\hbar = 1$). Each mode possesses an associated action

$$I = \frac{1}{2} \int dt dr^* \Big[\left(\frac{\partial \psi}{\partial t} \right)^2 - \left(\frac{\partial \psi}{\partial r^*} \right)^2 - V(r) \psi^2 \Big], \tag{2.19}$$

where we have suppressed the index ℓm. $V(r)$ contains the centrifugal barrier and the Schwarzschild potential:

$$V(r) = \underbrace{\frac{r - 2MG}{r}}_{\substack{\text{at horizon} \\ \to 0}} \left(\underbrace{\frac{\ell(\ell+1)}{r^2}}_{\text{centrifugal}} + \underbrace{\frac{2MG}{r^3}}_{\text{Schwarzschild}} \right). \tag{2.20}$$

The ℓ-independent term can prevent even s-wave modes from leaving the vicinity of the horizon.

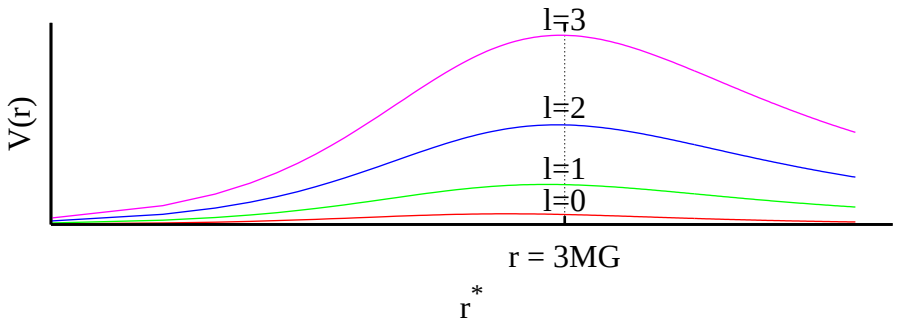

Figure 2.12: Close to the black hole, the barrier pushes an outward travelling wave packet back into the black hole. A wave packet energetic enough to clear the top of the potential hill at $r \approx 3MG$ can escape to infinity. Since higher angular momentum modes experience a higher potential (the smallest one being of order $1/(MG)^2$) only the s-waves escape at low energies.

Classically, in order to escape from the vicinity of the horizon a wave quantum must have a frequency ν, which is greater than the inverse light crossing time of the event horizon. Numerically,

$$\nu > \frac{0.16}{MG}. \tag{2.21}$$

Similarly, the centrifugal barrier can be overcome from the outside by sufficiently energetic particles, allowing matter to collapse to a point (for this reason Einstein did not like the Schwarzschild solution).

Quantum mechanically, s-waves also stand the best chance of getting out of the black holes, since their potential barrier is shallowest. This is one reason why black hole evaporation is so slow. The s-wave quanta have to be replenished by scattering from $\ell > 0$.

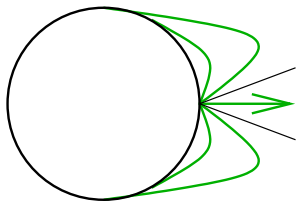

Figure 2.13: Of all particles with constant linear momentum, those with least angular momentum are most likely to leave the surface of a star.

Figuratively, for a given amount of total momentum, the more *angular* momentum a particle has, the shallower is the angle with the horizon. Such a particle is more likely to fall back toward the surface of the black hole than a particle heading straight outward.

Lecture 3

3.1 Review

(This section can be skipped if you remember everything from the first two lectures.) Recall our reason for concentrating on Schwarzschild black holes: we are interested in generic properties, and each black hole reduces to flat spacetime near its horizon, so we needn't go beyond Schwarzschild. Recently there has been much excitement amongst string theorists about extremal black holes, which don't reduce to flat spacetime near their horizon. These black holes might fill in some pieces of the picture that Schwarzschild can't provide, but they don't have the generic properties that we are interested in, and they are only simple to work with if one assumes supersymmetry.

3.1.1 Coordinates

Remind yourself of the following definitions:

$$t = \text{Schwarzschild time} = \text{Minkowski time far from the horizon} \tag{3.1}$$

$$\omega = \frac{t}{4MG} = \text{Rindler time, a hyperbolic angle} \tag{3.2}$$

$$\rho = \text{proper distance from the horizon} \tag{3.3}$$

$$u = \ln \rho \to -\infty \text{ as } \rho \to 0. \tag{3.4}$$

The coordinates u and ω are conformal, and the tortoise coordinates reduce to them near the horizon. As we shall see, the relationship between the temperature and mass of the black hole is related to that between t and ω.

Also, recall that the classical physics in Region I of Rindler space (see Fig. 3.1) is self-contained, in that the information from Region IV forms the initial data on the past horizon at $t = -\infty$ and no information can pass from Region III into Region I. This means that, classically, anything "on the other side" of the horizon doesn't

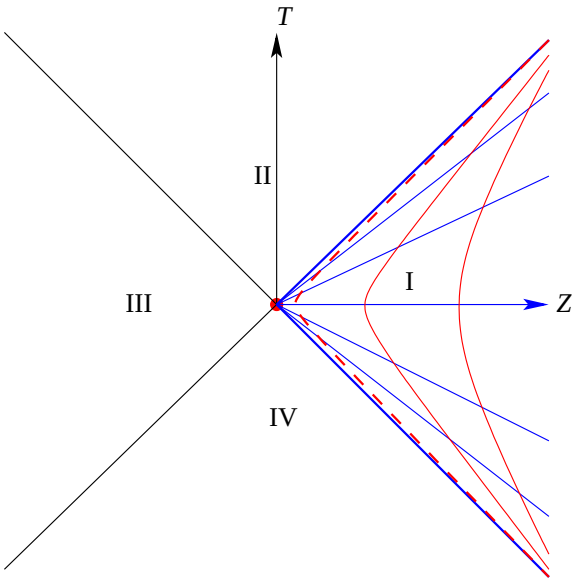

Figure 3.1: Rindler space, with lines of constant ρ in red and lines of constant ω in blue.

affect anything on "our side." This will change somewhat once we start examining quantum fields near the horizon.

3.1.2 Scalar fields

Last lecture we introduced the wave equation for the k-th mode of a massless scalar field in Rindler space:

$$\frac{\partial^2 \chi_k}{\partial \omega^2} - \frac{\partial^2 \chi_k}{\partial u^2} + k^2 e^{2u} \chi_k = 0. \tag{3.5}$$

This follows from an expansion of χ in plane waves e^{ikx}, where x is parallel to the horizon. Hence, for the $k = 0$ mode there is no momentum along the horizon, and all the field's momentum is radial. In this case there is no potential barrier. For nonzero k, the wave propagates a certain distance from the horizon until it is reflected by the exponential potential e^{2u}, which is shown in Fig. 3.2.

Note that Eq. (3.5) is obviously only valid for $r > 2MG$, and it breaks down around $r \sim 3MG$, so it is only valid very near the horizon.

When we move away from the horizon we must allow for angular momentum, since the horizon is spherical rather than flat. Expanding our wave in spherical

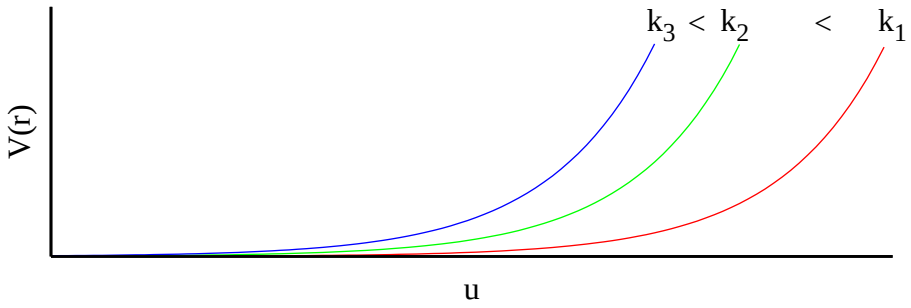

Figure 3.2: The radial potential of a scalar field very near the horizon for three values of k.

harmonics and using tortoise coordinates, we found

$$\frac{\partial^2 \psi_{lm}}{\partial t^2} - \frac{\partial^2 \psi_{lm}}{\partial r^{*2}} + V_{lm}(r^*)\psi_{lm} = 0, \tag{3.6}$$

where $\psi_{lm} \equiv r\chi_{lm}$ and

$$V_{lm}(r^*) = \left(1 - \frac{2MG}{r}\right)\left(\frac{l(l+1)}{r^2} + \frac{2MG}{r^3}\right). \tag{3.7}$$

We can intuitively relate these equations to those for plane waves in Rindler space as follows: since angular momentum is given by $\vec{L} = \vec{r} \times \vec{k}$, near the horizon we have $l \sim 2MGk$. The potential must reduce to $k^2 e^{2u}$ near the horizon, but differs from it far from the horizon; as we can see in Fig. 3.3, the potential eventually reaches a maximum and then decays to zero at large r. The potential peaks near $r \sim 3MG$ for all values of l, and peaks precisely at this point in the limit of large l.

Note the general behavior of the potential: it and its derivative vanish at the horizon, so the scalar field there is completely free; it contains a centrifugal barrier $\propto l(l+1)$; and it has a pure s-wave potential $\propto 2MG/r^3$ which dominates near the horizon, indicating the gravitational field that pulls matter fields back to the horizon. Obviously s-waves have the greatest chance of escaping the potential barrier. Consider an s-wave quantum with frequency ν (defined with respect to time at $r = \infty$). Recalling from the last lecture that the quantum has energy ν^2, we see that it will pass over the potential barrier if $\nu > \sqrt{V_{max}} = \frac{1}{\sqrt{2}}\left(\frac{3}{8}\right)^{3/2}\frac{1}{MG} \approx \frac{0.16}{MG}$. Of course, if the frequency is slightly lower than this then the wave still has a chance to tunnel through the barrier.

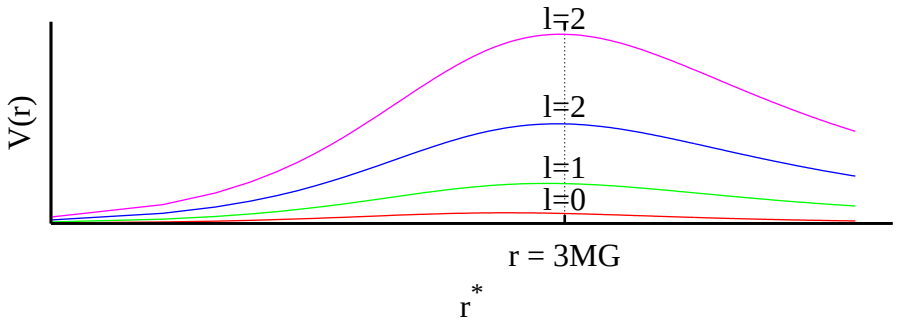

Figure 3.3: The radial potential in general for four values of l.

3.2 Black hole Quantum Field Theory

We now move into the full Quantum Field Theory regime. We are specifically interested in Hawking radiation, the entanglement of vacuum fluctuations on opposite sides of the horizon, and how this leads to a thermodynamic equilibrium state with a finite temperature. *We will uniformly neglect normalization in this section.*

3.2.1 The Rindler Hamiltonian

Consider Rindler space in the coordinates $Z = \rho \cosh \omega$ and $T = \rho \sinh \omega$ that we introduced in Lecture 1. The normal Hamiltonian that generates translations in T is then $H = \int T^{TT} d\rho d^2x$ (at $T = \omega = 0$, so $\rho = Z$, which is all we need to consider because T can always be translated to zero). Here $\vec{x}$ coordinates are again perpendicular to the T-ρ plane. Since an infinitesimal time interval dT is equal to $\rho d\omega$ (again at $T = 0$), the *Rindler Hamiltonian*, which generates translations in Rindler time ω, is given by

$$H_R = \int T^{TT} \rho d\rho d^2x. \tag{3.8}$$

For example, for a free massless scalar field we have

$$H_R = \frac{1}{2} \int \rho \left[\Pi^2 + (\partial_\rho \chi)^2 + (\partial_{\vec{x}} \chi)^2 \right] d\rho d^2x, \tag{3.9}$$

where $\Pi = \frac{1}{\rho} \frac{\partial \chi}{\partial \omega}$ is conjugate to χ. However, note that we don't want to restrict ourselves to free fields when we consider Hawking radiation and other relevant quantum effects.

We can derive the Rindler Hamiltonian more explicitly by looking at the generator of T-translations, $U = e^{-i\int H dT} = e^{-i\int T^{TT} dT dZ d^2x}$. Since $dTdZ = \rho d\rho d\omega$, we have $U = e^{-i\int T^{TT}\rho d\rho d\omega d^2x} = e^{-i\int H_R d\omega}$, where H_R is as given above.

3.2.2 The density matrix

Hawking radiation is created by vacuum fluctuations near the horizon; part of the field fluctuation (e.g. one particle) is lost behind the horizon, while another part (e.g. one antiparticle) can escape to infinity. However, the emitted particles are entangled: the correlation amplitude between fields ϕ_A in region A (region III in Fig. 3.1) and ϕ_B in region B (region I in Fig. 3.1) is given by

$$\langle \phi_A | \phi_B \rangle \sim \frac{1}{\Delta^2}, \tag{3.10}$$

where Δ is the distance between the points at which the fields are evaluated, and $|\phi\rangle$ is the eigenstate of the field operator ϕ. (*This result can be found by directly evaluating the usual scalar field propagator for small spacelike separations.*) Obviously, we need an account of the outgoing particles that includes the effects of the entanglement with the particles behind the horizon. However, note that there are no terms in H_R that couple ϕ_A with ϕ_B, so the fields do not actually interact—they are simply coupled.

If we can find the density matrix $\rho = |\Psi_{GS}\rangle\langle\Psi_{GS}|$ for the entire Minkowski vacuum at $T = 0$, then we can find the density matrix ρ_A for "our" side of the horizon by tracing out the fields in region B:

$$\rho_A(\phi_A, \phi_{A'}) \equiv \sum_{\phi_B} \Psi^*(\phi_A, \phi_B)\Psi(\phi_{A'}, \phi_B). \tag{3.11}$$

Then, for any operator $\hat{A}$ which acts on the space of states in region A, the expectation value for an observer in region A is $\langle \hat{A} \rangle = Tr(\rho_A \hat{A})$. Because the Rindler Hamiltonian is explicitly defined in terms of $\rho > 0$ (i.e. in region A), ρ_A evolves as $\frac{d\rho_A}{d\omega} = i[\rho_A, H_R]$. Note that a Rindler time translation is a Lorentz boost in flat space (which should be obvious based on the definition of lines of constant ω and t in the T-Z plane), and ρ_A should be Lorentz invariant, so we expect $\frac{d\rho_A}{d\omega} = \frac{d\rho_A}{dt} = 0$. However, a perturbation away from flat space, e.g. if we consider a larger region around the horizon, would make ρ_A time-dependent.

3.2.3 The vacuum state

Since we are interested in fluctuations in the vacuum, we must first find the vacuum state. We do this by Wick-rotating our time coordinates as $t \to -it$, effectively defining new time coordinates $\tilde{T} \equiv iT$ and $\theta \equiv i\omega$. This means that the generator of T-translations becomes $e^{-H\tilde{T}}$, and we can find the ground state by evolving any state by an infinite time $\tilde{T}$:

$$|\Psi_{GS}\rangle = \lim_{\tilde{T}\to\infty} e^{-H\tilde{T}}|\Psi_{any}\rangle. \tag{3.12}$$

To see this, assume we set the ground state energy to zero; then

$$\begin{aligned} \lim_{\tilde{T}\to\infty} e^{-H\tilde{T}}|\Psi_{any}\rangle &= \lim_{\tilde{T}\to\infty} \sum_E e^{-E\tilde{T}}|E\rangle\langle E|\Psi_{any}\rangle & (3.13)\\ &= |\Psi_{GS}\rangle\langle\Psi_{GS}|\Psi_{any}\rangle, & (3.14) \end{aligned}$$

since the exponential vanishes except for $E = 0$. This means that the ground state wave functional (neglecting normalization) is given by

$$\Psi_{GS}(\phi_A, \phi_B) = \lim_{\tilde{T}\to\infty} \langle\phi_A\phi_B|e^{-H\tilde{T}}|\Psi_{any}\rangle. \tag{3.15}$$

This can be written as a path integral using the Feynman-Kac formula:

$$\Psi_{GS}(\phi_A, \phi_B) = \int D[\chi] e^{-\int_0^\infty L_E(\phi) d\tilde{T}}, \tag{3.16}$$

where the integral runs over all fields in the upper half-plane which satisfy the boundary condition $\phi(Z, T = 0) = \phi_A(Z)$ for $Z > 0$ and $\phi(Z, T = 0) = \phi_B(Z)$ for $Z < 0$, and L_E is the Euclidean classical Lagrangian obtained by replacing T with $\tilde{T}$. Noting that the integral from $\tilde{T} = 0$ to $\tilde{T} = \infty$ can be rewritten as an integral from $\theta = 0$ to $\theta = \pi$, which covers the upper half-plane just as well, we find

$$\Psi_{GS}(\phi_A, \phi_B) = \langle\phi_B|e^{-H_R\pi}|\phi_A\rangle. \tag{3.17}$$

So the vacuum state is the propagator from the field configurations in region I of Rindler space to those in region III.

For readers unfamiliar with the functional representation and path integrals of Quantum Field Theory, note that $\Psi(\phi) = \langle\phi|\Psi\rangle$ is the probability amplitude of the field being found in a configuration $\phi(x)$, just as in regular quantum mechanics $\psi(x) = \langle x|\psi\rangle$ is the probability amplitude of a particle being found at position x.

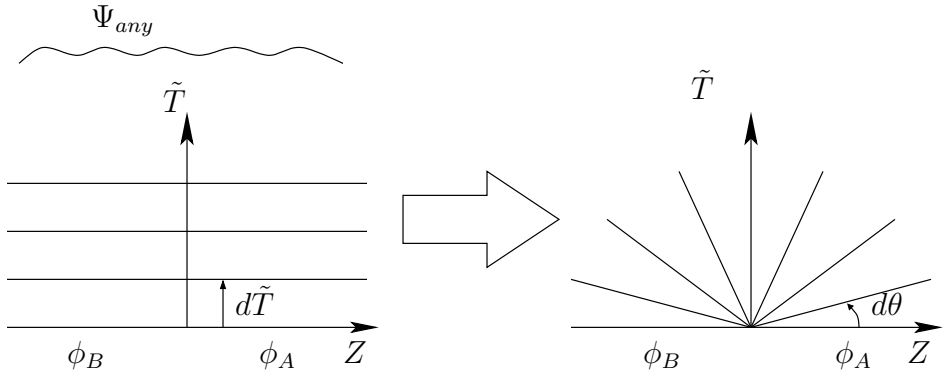

Figure 3.4: Left: the vacuum as the propagation of an arbitrary field at $\tilde{T} = \infty$ to ϕ_A and ϕ_B at $\tilde{T} = 0$. Right: the vacuum as the propagation of ϕ_A at $\theta = 0$ to ϕ_B at $\theta = \pi$. The two equivalent formulations both consist of an integral over all fields in the upper half-plane, subject to the boundary conditions imposed by ϕ_A and ϕ_B.

The probability amplitude of evolving from a configuration $\phi(x, t_0)$ at t_0 to another configuration $\phi(x, t_f)$ at t_f is given by $\langle\phi(t_f)|\phi(t_0)\rangle = \int D[\chi(x,t)] \exp\left(i \int_{t_0}^{t_f} L dt\right)$, where the integral is over all fields χ that satisfy $\chi(t_i) = \phi(t_i)$ $(i = 1, 2)$, and L is the classical Lagrangian; again, this is completely analogous to regular Quantum Mechanics, where $\langle x(t_f)|x(t_0)\rangle = \int D[x(t)] \exp\left(i \int_{t_0}^{t_f} L dt\right)$. Just as the functional integral in regular Quantum Mechanics is evaluated by dividing the time interval into N smaller intervals and letting $N \to \infty$, with the measure $D[x(t)] = dx(t_1)dx(t_2)...dx(t_{N-1})$, the functional integral in QFT is evaluated by dividing up both space and time, with the measure $D[\chi(x,t)] = \prod_i \prod_j d\chi(x_i, t_j)$. So, keeping this in mind, and letting $\phi = \phi_A \cup \phi_B$, we have

$$\begin{aligned}
\Psi_{GS}(\phi) &= \langle\phi| \lim_{iT\to\infty} e^{-iHT} |\Psi_{any}\rangle && (3.18)\\
&= \lim_{iT\to\infty} \langle\phi(0)|\Psi_{any}(T)\rangle && (3.19)\\
&= \lim_{iT\to\infty} \int_{\chi(0)=\phi(0)} D[\chi] e^{i\int_T^0 L dT} && (3.20)\\
&= \int_{\chi(0)=\phi(0)} D[\chi] e^{-\int_0^\infty L_E d\tilde{T}}. && (3.21)
\end{aligned}$$

Note that there is no boundary condition at $\tilde{T} = \infty$, since Ψ_{any} is arbitrary. We now

note that the Lagrangian L is the integral of the Lagrangian density $\mathcal{L}$, so

$$\Psi_{GS}(\phi) = \int_{\chi(0)=\phi(0)} D[\chi] e^{-\int\int_{-\infty}^{\infty}\int_{0}^{\infty}\mathcal{L}_E d\tilde{T} dZ d^2x} \tag{3.22}$$

$$= \int_{\substack{\chi(0)=\phi_A \\ \chi(\pi)=\phi_B}} D[\chi] e^{-\int\int_{0}^{\infty}\int_{0}^{\pi}\mathcal{L}_E \rho d\theta d\rho d^2x}. \tag{3.23}$$

Combining ρ with $\mathcal{L}_E$ and reversing all the above steps, we arrive at the result

$$\Psi_{GS}(\phi) = \langle\phi_B|e^{-H_R\pi}|\phi_A\rangle. \tag{3.24}$$

3.3 Black hole thermodynamics

Substituting our result for the vacuum wave functional into the density matrix, we find

$$\rho_A(\phi_A, \phi_{A'}) = \sum_{\phi_B} \Psi^*(\phi_A, \phi_B)\Psi(\phi_{A'}, \phi_B) \tag{3.25}$$

$$= \sum_{\phi_B} \langle\phi_A|e^{-H_R\pi}|\phi_B\rangle\langle\phi_B|e^{-H_R\pi}|\phi_{A'}\rangle \tag{3.26}$$

$$= \langle\phi_A|e^{-H_R\pi}|\phi_{A'}\rangle, \tag{3.27}$$

so

$$\rho_A = e^{-2\pi H_R}. \tag{3.28}$$

But this is a thermodynamic density matrix with a temperature

$$T_R = \frac{1}{2\pi}. \tag{3.29}$$

This means that the physics in the Rindler wedge is a thermal equilibrium at temperature T_R. Since our derivation has been in Rindler space, this result applies to both the Unruh effect and Hawking radiation. It also applies for any field theory, as long as the black hole is in equilibrium, since our derivation of the vacuum state did not depend on any particular Lagrangian.

Note, however, that this only applies to Rindler observers, since tracing out the fields in Region III is what transforms the Minkowski vacuum into a thermal bath at finite temperature. The physical reason for this is simple: vacuum fluctuations that encircle the origin in Minkowski space appear as normal fluctuations for free-falling observers, but they survive from $\omega = -\infty$ to $\omega = +\infty$, so they appear as persistent radiation for Rindler observers.

3.3.1 Proper temperature

However, T_R is a "ccordinate temperature" corresponding to Rindler coordinates. We need to convert it to a proper temperature to find what would actually be measured by an observer. Using $d\tau = \sqrt{g_{00}}d\omega$, we have the relationship $E_p = E_R/\sqrt{g_{00}}$ between proper energy and Rindler energy. Temperature is a measurement of energy, and $\sqrt{g_{00}}$ is equal to ρ near the horizon and $4MG$ at infinity, so

$$T_p = \frac{1}{2\pi\rho} \text{ near the horizon} \tag{3.30}$$

$$T_p = \frac{1}{8\pi MG} \text{ at infinity.} \tag{3.31}$$

This is a somewhat strange result. It indicates that proper temperature diverges at the horizon, regardless of the mass of the black hole, and yet it remains finite at infinity, decreasing with increasing mass. Restoring all the physical constants, the second formula gives the Hawking temperature

$$T_H = \frac{\hbar c^3}{8\pi MGk}. \tag{3.32}$$

This temperature is usually exceedingly small. For example, if $M \sim M_\odot$ then $T_H \sim 10^{-8}$K.

3.3.2 Entropy

As we shall see, the entropy of a black hole would be infinite were it not for quantum mechanical effects. We will show this using only the usual relationship $dE = TdS$ between energy and entropy (which, incidentally, can be used to define the temperature as the increase in energy of a system per unit information added). Substituting $E = Mc^2$ and the Hawking temperature into this equation yields

$$dM = \frac{\hbar c}{8\pi MG}dS \tag{3.33}$$

$$\Rightarrow \frac{4\pi G}{\hbar c}d(M^2) = dS \tag{3.34}$$

$$\Rightarrow S = \frac{4\pi M^2 G}{\hbar c}. \tag{3.35}$$

In the classical limit $\hbar \to 0$ and so $S \to \infty$.

Note that the area of the black hole horizon is $A = 4\pi R_s^2$, and $R_s = 2MG/c^2$, so

$$S = \frac{A}{4\hbar G}. \tag{3.36}$$

The quantity $\hbar G$ is the Planck area, so we have the terrifically simple result that $S = \frac{1}{4}A$ in Planck units.

3.3.3 Luminosity

The luminosity of a black hole is often caclculated using the Stefan-Boltzmann law for blackbody radiation. However, the luminosity in this case depends mostly on s-waves with wavelengths similar to the size of the black hole, which is not the case in the Stefan-Boltzmann law. Hence, we need a new calculation.

A typical thermal quantum in the ensemble behind the potential hill, in the "thermal atmosphere" of the black hole, has an energy $\sim \frac{1}{8\pi MG}$ (neglecting $\hbar$ and c). Refering to Eqn. (3.7) for the potential, we see that all the potential barriers are large compared to the average kinetic energy—except if $l = 0$. The s-wave quanta have roughly a 75% chance of tunneling past the barrier, while other quanta have a very low chance even for small black holes. (Note that a black hole capable of emitting visible light must be approximately the size of the light's wavelength, which is tiny, so we would never see it anyway.)

So, how many particles escape the potential barrier? The thermal wavelength is $\sim MG$, and the particles are separated by a distance on the scale of the black hole's size, and each of these will contribute an inverse factor to the luminosity, so

$$L \sim \frac{1}{(MG)^2}. \tag{3.37}$$

This general scaling can also be found using the Stefan-Boltzmann law $L = \sigma A T^4$, where A is area, but the numerical factors may differ in the correct result.

From this we see that $\frac{dM}{dt} \sim -1/(MG)^2$, so the time for a black hole to evaporate all its energy is $t_{evap} \sim M^3G^2$. For example, $t_{evap} \sim 10^{60}$ years if $M \sim M_\odot$. However, all the black holes in the universe are probably colder than the microwave background, so their absorption of energy is probably greater than their dissipation of it.

The main conclusion to draw from this section is a seeming paradox: if proper temperature means what it should mean, then a Fido near the horizon sees anything passing through the horizon get cooked, ionized, and so forth by the extreme temperature; but from the perspective of a Frefo, absolutely nothing happens.

Lecture 4

Aside: Thermal bath of particles around the black hole The Hawking temperature $T = \frac{1}{8\pi MG}$ corresponds to a thermal bath of particles surrounding the event horizon. Even though there is no potential barrier very close to the event horizon ($V(r)$ vanishes at the horizon), the particles in the thermal bath do not simply fall into the black hole. The thermal bath is only observed by Fidos. In the tortoise coordinates of the Fidos the horizon is located at $r^* = -\infty$, and even a relativistic particle that is seen to travel with unit coordinate speed is never observed to fall into the black hole. Rather, Fidos see the outside of the black hole surrounded by pancake shaped layers of matter.

4.1 Information and black holes

Black holes are incredibly dense reservoirs of information. In a sense they can be *defined* as objects possessing one bit of information per Plank area. Most if not all of this information is hidden as entropy in the thermal atmosphere of the black hole. Later on we will see that trying to derive this entropy using Quantum Field Theory on a static background will lead to divergences. This was first shown by 't Hooft. In the end a more careful calculation boils down to the Stefan-Boltzmann law. *The Stefan-Boltzmann law was originally derived for an object, like the Sun, that radiates quanta with wavelengths much much smaller than the size of the object and with random angular momenta. However, qualitatively the result holds even for black holes, which emit quanta with wavelengths comparable to the size of the event horizon and which, due to the high centrifugal barrier for high angular momentum particles, emits almost exclusively s-wave quanta. (Because of this last fact, the image of a black hole will be fuzzy, not sharp like the image of the Sun.)* The Stefan-Boltzmann law states that for every massless species of particle the three

dimensional entropy density is

$$\sigma = \gamma T^3, \tag{4.1}$$

where $\gamma = \gamma(T)$ is a measure of the number of species that make up the gas. For sufficiently high temperatures γ includes massive particles, as long as their mass is small compared to the thermal energy kT. *Equation* (4.1) *follows from the expression for energy density* $u \propto \sigma_{SB} T^4$ *of a photon gas and its relation* $s \propto u/T$ *to the entropy.* σ_{SB} *above is the Stefan-Boltzmann constant, and each of the constants of proportionality is of order* 1 *in units where* $c = 1$. It is widely believed that in Quantum Field Theory the number of degrees of freedom γ monotonically increases with T. As the temperature increases, previously frozen-out modes become available and/or more types of particles can be efficiently produced by thermal fluctuations.

The total entropy in the region near the event horizon is then

$$S = \int d^2x \, d\rho \gamma T^3 = \int d^2x \, d\rho \gamma \left(\frac{1}{2\pi\rho} \right)^3, \tag{4.2}$$

which diverges quadratically as $\rho \to 0$. It should be noted that this divergence persists in quantum field theory and is not the usual ultraviolet catastrophe of classical field theory. We will later see that G, rather than $\hbar$, is controlling the integral.

For now we deal with this by introducing a cutoff a short distance away from the horizon (a *stretched horizon*) at ρ_0:

$$S = \int dx^2 \int_{\rho_0}^{\infty} \frac{\gamma}{(2\pi\rho)^3} \, d\rho \sim \frac{\gamma A}{\rho_0^2} \tag{4.3}$$

where A is the surface area of the horizon. (The coordinates x here run over the whole surface, rather than just being limited to a small patch perpendicular to the T-Z plane.) In order to recover the correct Hawking temperature the cutoff has to be of order of the Planck length $\ell_P = \sqrt{\frac{\hbar G}{c^3}} \approx 10^{-33}\,\mathrm{cm}$.

Note that the quanta in the range $\rho < \rho_0$ are real and have to be cut off; they cannot be ignored. Many of the paradoxes involving black hole entropy are based on using naïve quantum field theory near the horizon.

4.1.1 Example: point charge falling into black hole

A high energy electron whizzes through the event horizon of a black hole, keeping its shape and momentum (in its rest frame). One might think that, as seen from the outside, the charge as it gets redshifted simply sticks to the black hole and creates

a charged spot. However this must be wrong; because of the No-Hair Theorem the particle has to eventually spread out uniformly across the horizon.

We can gain an intuitive understanding of what is going on by following a particle with charge e into the event horizon. To do this, we introduce a stretched horizon into Rindler space.

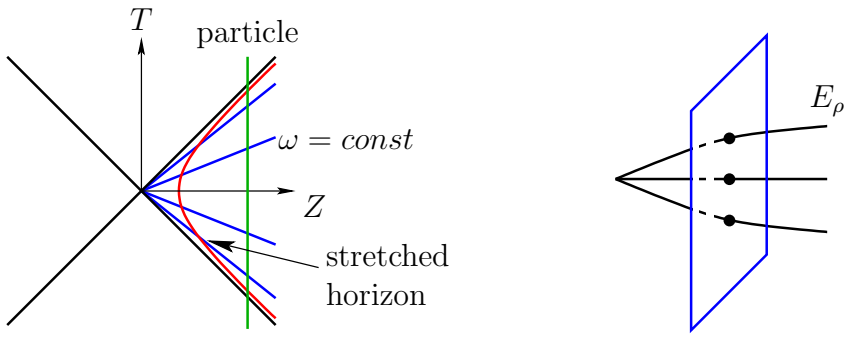

Figure 4.1: Left: Spacetime diagram showing the stretched horizon and an infalling particle. Since Rindler space is just boosted Minkowski spacetime, a particle at rest in the Minkowski frame appears to be accelerated in Rindler coordinates. Right: the field lines emanating from the event horizon after the charge has passed through it.

Note that the stretched horizon in Fig. 4.1 is timelike, whereas the real horizon is by definition lightlike. Eventually the charge pass through the stretched horizon. After that time, the particle's outgoing field lines will pierce the horizon and appear as charges to an external observer.

In order to calculate the charge density on the surface of the stretched horizon, we first need to calculate the normal component of the electric field near the horizon. Under a boost the component of the electric field parallel to the boost direction is unchanged. Therefore $E_\rho = E_Z$, and using Coulomb's law we find

$$E_Z = \frac{e(Z - Z_0)}{[(Z - Z_0)^2 + x_\perp^2]^{3/2}}, \tag{4.4}$$

where Z_0 is the constant Minkowski coordinate of the particle. This gives us the charge density

$$\sigma = \frac{E}{4\pi}. \tag{4.5}$$

Right at the stretched horizon, where $\rho = \rho_0 \approx \ell_P$, we have

$$Z = \rho_0 \cosh\omega \overset{\omega\to\infty}{\to} \frac{1}{2}\rho_0 e^\omega. \tag{4.6}$$

That is to say, at late times, long after the particle has passed through the horizon

$$\sigma \approx \frac{e}{4\pi} \frac{e^{\omega}}{(\rho_0^2 e^{2\omega} + x_\perp^2)^{3/2}} \tag{4.7}$$

as Z_0 becomes negligible. Defining a new coordinate $Y = e^{-\omega} x_\perp$ we express σ as

$$\sigma = \frac{e}{4\pi} \frac{e^{-2\omega}}{(\rho_0^2 + Y^2)^{3/2}}. \tag{4.8}$$

Thus the shape of the charge distribution is fixed, but the peak height decays exponentially while the distribution spreads out in the same way. Y is a comoving coordinate describing the shape of the charge distribution, but not its size. From this, we see that $\sigma(Y, \omega)$ flattens and broadens in such a way that the total charge remains constant. This is how a charge distribution in a conductor behaves. The speed with which it spreads is given by the conductivity of the conductor. *To see this, consider an Ohmic conductor for which $\vec{j} = \bar{\sigma}\vec{E}$, using the conductivity $\bar{\sigma}$. (Where$^-$ only serves to distinguish between the charge density σ and the conductivity $\bar{\sigma}$.) Taking the divergence yields*

$$\nabla \cdot \vec{j} = 4\pi\bar{\sigma}\sigma, \tag{4.9}$$

where we used Gauss' law $\nabla \cdot \vec{E} = 4\pi\sigma$. The continuity equation $\nabla \cdot \vec{j} + \dot{\sigma} = 0$ allows us to write

$$\dot{\sigma} = -4\pi\bar{\sigma}\sigma \Rightarrow \sigma = \sigma_0 e^{-4\pi\bar{\sigma}\omega}. \tag{4.10}$$

Knowing this, we can read off the conductivity of the stretched horizon as $\bar{\sigma} = \frac{1}{2\pi}$.

So, an outside observer sees a charge exponentially spreading over the fictitious surface of the event horizon. We can estimate the time it takes to spread over the horizon. To spread evenly across the horizon we must have $Y \lesssim \rho_0$, which renders the numerator of Eq. (4.8) nearly constant with respect to Y. Since $\rho_0 \approx \ell_P$, $X_\perp$ is on the order of R_S, and $Y = e^{-\omega} X_\perp$, this results in

$$\ell_P \approx e^{-\omega} R_S \Leftrightarrow \omega \approx 4MG \ln \frac{2MG}{\ell_P}. \tag{4.11}$$

This is the natural time scale for any process involving the horizon. An astronomer far away would observe the spreading out of the charge and attribute it to the effects of the hot thermal plasma around the horizon.

4.2 Entropy and information

The notion of information was vague for a very long time. In this section we will define what information is and how it evolves in a model system.

Before turning to information we first have to become familiar with its opposite, entropy. In classical Statistical Mechanics entropy is defined in terms of phase space. We usually possess only incomplete information of a system, so we cannot pinpoint its location in phase space precisely. Instead we only know that the system's state is in some region of volume V of phase space. Entropy is defined as the logarithm of this volume (normalized by some appropriate unit volume, *e.g.* some power of Planck's constant)

$$S = \ln V. \tag{4.12}$$

Liouville's theorem asserts that the volume V, and with it the entropy, is constant over the course of the evolution of the system. We will call entropy defined this way the *fine-grained entropy*. Thermodynamic or *coarse-grained entropy* is defined as the logarithm of the coarse-grained volume. By coarse-graining we mean the process of surrounding every point in a volume of phase space by a sphere of constant radius and taking the union of these spheres as the coarse-grained volume. As the original volume becomes more and more fractal, the fine grained entropy stays constant while the coarse-grained entropy increases, leading to the thermodynamic notion of entropy as an ever-increasing function. This process of coarse-graining lies at the centre of thermodynamics.

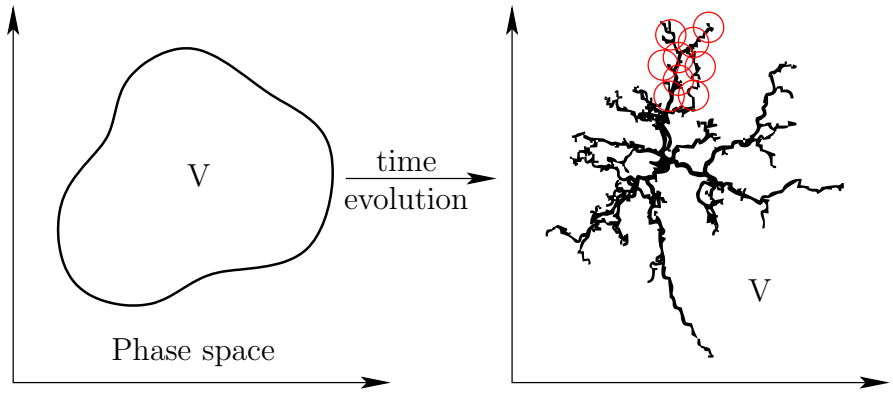

Figure 4.2: Time-evolution of the phase space volume of a system. The fine-grained volume of the region stays constant, but the coarse grained volume increases with time.

In some sense then, entropy is information hidden in objects too small or numerous to track.

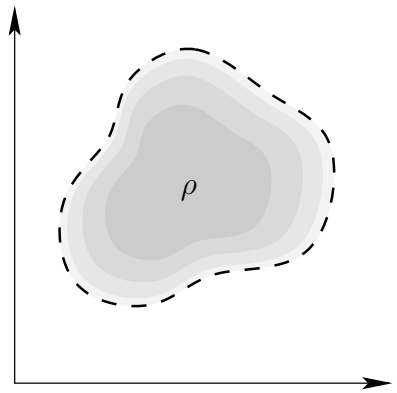

Figure 4.3: Probability distribution of finding the system in state $(\vec{x}, \vec{p})$ in phase space.

Instead of specifying a sharp volume in phase space, we can also specify a probability distribution ρ of finding the system in a certain state. This generalizes the previous notion of entropy in the following sense:

$$S = \ln V \qquad \rightarrow \qquad S = -\int dV\, \rho \ln \rho \tag{4.13}$$

If the system is equally likely to be at any given point within a certain volume V of phase space, then

$$\rho = \begin{cases} \frac{1}{V} & \text{within } V \\ 0 & \text{otherwise} \end{cases}. \tag{4.14}$$

With this definition

$$S = -\int dV\, \frac{1}{V} \ln \frac{1}{V} = +\frac{1}{V} \ln V \int dV = \ln V, \tag{4.15}$$

in agreement with the previous definition in Eq. (4.12).

If the collection of states of the system is discrete instead of continuous, the definition in Eq. (4.12) corresponds to some finite number n of equally probable states. The refined definition Eq. (4.13) allows for the inclusion of different probabilities P_i for the individual states. Contrasting again the "old" and "new" styles of defining entropy, *i.e.*

$$\text{old: } S = \ln n, \tag{4.16}$$

$$\text{new: } S = -\sum_i P_i \ln P_i, \tag{4.17}$$

we see that by setting $P_i = \frac{1}{n}$ we recover Eq. (4.16) from the probability density formalism. The entropy defined in this way in Eq. (4.13) is called von-Neumann entropy.

This refined definition of entropy using a probability distribution generalizes to Quantum Mechanics. There the probability distribution is replaced by the density matrix. It is a hermitian operator whose eigenvalues are the probabilities of finding the system in a certain state. All of the eigenvalues are positive and the trace of

the density matrix is unity, corresponding to the fact that the system has to be in some state. As the density matrix is symmetric, it can be diagonalized in some basis $\{|\psi_i\rangle; i \in \mathbb{N}\}$:

$$\rho = \begin{pmatrix} \rho_1 & & & 0 \\ & \rho_2 & & \\ & & \rho_3 & \\ 0 & & & \ddots \end{pmatrix}. \tag{4.18}$$

The states $|\psi_i\rangle$ which diagonalize the density matrix are called pure states, while states that are linear combinations of pure states are called entangled states. Using the density matrix, the entropy of a quantum mechanical system is defined as

$$S = -\operatorname{Tr} \rho \ln \rho. \tag{4.19}$$

If the system is in the pure state i, then $\rho_j = 0$ if $j \neq i$ and $\rho_j = 1$ for $j = i$. In this state the entropy vanishes. If on the other hand the system is maximally mixed, *i.e.* $\rho_i = \frac{1}{n}$ where n is the number of states possible, then

$$S = \ln n, \tag{4.20}$$

which is the maximum value the entropy can assume. Therefore S is a measure of how pure a state is.

4.3 Entanglement entropy

In classical mechanics the only source of entropy is ignorance about the precise state of the system. In quantum mechanics another source of entropy is present: entanglement. Note that this applies even at the level of *fine-grained* entropy; the thermodynamical coarse graining adds another source of entropy present in both classical and quantum mechanical systems.

Consider a composite system AB which can be split into subsystems A and B. Each of the subsystems is described by a complete set of commuting observables α (for subsystem A) and β (for subsystem B). The state of the complete system is then given by $\Psi(\alpha\beta)$. The subsystems can be described using density matrices

$$(\rho)_{\alpha\alpha'} = \sum_{\beta} \Psi^*(\alpha\beta)\Psi(\alpha'\beta) \text{ and} \tag{4.21}$$

$$(\rho)_{\beta\beta'} = \sum_{\alpha} \Psi^*(\alpha\beta)\Psi(\alpha\beta') \tag{4.22}$$

for the subsystems A and B respectively. The density matrices act as operators in their own subspaces and as scalars in the other subspace. *The expectation value of an operator X_A defined only in subsystem A is given by*

$$\langle X_A \rangle = \mathrm{Tr}_A \, \rho_{\alpha\alpha'} X_A, \tag{4.23}$$

where the trace is calculated over all states in subsystem A. $\rho_{\alpha\alpha'}$ contains the dependence on the states in subsystem B.

Density matrices allow us to define individual entropies for the two subsystems. This entanglement entropy is purely due to the hidden correlation between the two systems and is always equal for the two systems:

$$S_A = S_B. \tag{4.24}$$

If one considers a total system consisting of a very large subsystem A and a very small subsystem B then the statement that the entropies of these two different systems are equal seems surprising at first. Recall, however, that the entropy is due to the entanglement *between* the systems; it is a collective effect, not due to either system A or B alone. To show the equality of the two entropies, it is sufficient to show that the non-zero eigenvalues of the density matrices in subsystem A and B are equal.

Proof Let $\phi_A(\alpha)$ be an eigenvector of $\rho_{\alpha\alpha'}$ with non-zero eigenvalue λ. Then

$$\Psi^*(\alpha\beta)\Psi(\alpha'\beta)\phi_A(\alpha') = \lambda\phi_A(\alpha), \tag{4.25}$$

where a sum over the repeated index β is implied. We will show that $\chi_B(\beta') = \Psi^*(\alpha'\beta')\phi_A^*(\alpha')$ is an eigenvector of $\rho_{\beta\beta'}$ for the same eigenvalue. Substituting into the eigenvalue equation we find

$$\Psi^*(\alpha\beta)\Psi(\alpha\beta')\chi_B(\beta') = \Psi^*(\alpha\beta)\underbrace{\Psi(\alpha\beta')\Psi^*(\alpha''\beta')\phi_A^*(\alpha'')}_{\lambda\phi_A^*(\alpha)} \tag{4.26}$$

$$= \lambda\Psi^*(\alpha\beta)\phi_A^*(\alpha) \tag{4.27}$$

$$= \lambda\chi_B(\beta). \tag{4.28}$$

Therefore λ is also an eigenvalue of $\rho_{\beta\beta'}$.

The maximum entanglement that the two subsystems can have is determined by the smaller system. This maximum entropy is reached when the smaller subsystem is maximally mixed. In this configuration all eigenvalues ρ_i of the density matrix

are equal and $\rho_i = \frac{1}{n}$, where n is the dimension of the Hilbert space of the smaller system. As a consequence of this, the entanglement entropy of a system containing a giant subsystem and a very small subsystem cannot be larger than the maximal entropy of the of the small subsystem.

This entropy is *not* additive. Even a system in a pure state, possessing a vanishing total entropy, can allow for entanglement between its subsystems, thus $S_A = S_B \neq 0$ even though $S_{AB} = 0$. The entanglement entropy vanishes if and only if the wave function of the full system factorizes as

$$\Psi(\alpha\beta) = \psi(\alpha)\psi(\beta). \tag{4.29}$$

4.3.1 Entropy of small subsystems

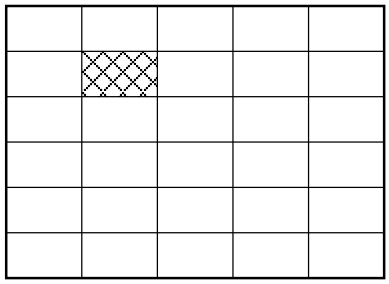

Figure 4.4: A large system subdivided into many identical smaller systems. The precise way of slicing up the larger system is not important as long as the small systems are less than half the size of the large system

Let's look at a big system composed of identical "smaller" subsystems (where small means less than half the size of the large system). We will assume that the total system is in a pure (vanishing entropy) state with energy E. Each small subsystem is entangled with all the other subsystems. Page [1] showed that independent of the state of the large system, each such small subsystem is almost always in a thermal state, *i.e.*

$$\rho = \frac{e^{-\beta H}}{Z}, \tag{4.30}$$

where $\beta = T^{-1}$, H is the subsystems Hamiltonian, and Z is the partition function of the subsystem (ensuring $\operatorname{Tr}\rho = 1$). Small parts of any system (even a system in an energy eigenstate) are thermal.

We use this to define the thermal entropy of a system as

$$S_{\text{thermal}} := \sum_i S_{\text{thermal}}(i), \tag{4.31}$$

where $S_{\text{thermal}}(i) = -\operatorname{Tr}\rho(i)\ln\rho(i)$ is the entropy of each subsystem. This entropy turns out to be independent of the precise nature of the slicing used as long as the subsystems are small and the coupling between subsystems is weak (see [1]). Any (arbitrarily large) subsystem Σ_1 of the full system Σ can then be described as the

union of some of the very small subsystems. In this context the thermal entropy is an upper bound to the (entanglement) entropy of Σ_1:

$$S(\Sigma_1) \leq S_{\text{thermal}}(\Sigma_1). \tag{4.32}$$

As a simple example consider a subsystem Σ_1 whose size changes until it includes the whole system Σ. While Σ_1 is smaller than $\Sigma - \Sigma_1$, its size determines the entanglement entropy of the constituent systems, so as Σ_1 grows both $S(\Sigma_1)$ and $S_{thermal}$ grow. If Σ is in a pure state then as Σ_1 grows to included all of Σ, it will approach a pure state and $S(\Sigma_1) \to 0$. On the other hand the thermal entropy is by definition additive, thus $S_{thermal}(\Sigma_1)$ continues to grow.

As mentioned earlier entropy is in some sense missing information, so it would be natural to *define* information as the difference between the maximum entropy a system can have and its actual entropy. Unfortunately there exist systems possessing no upper bound on their entropy. So instead we define the information to be the difference between the thermal entropy and the entanglement entropy of a system:

$$I = S_{\text{thermal}} - S. \tag{4.33}$$

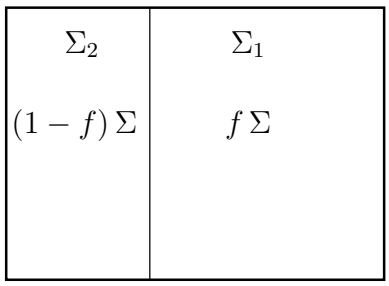

Figure 4.5: System Σ split into two subsystems Σ_1 and Σ_2.

Consider a large system Σ which is in a pure state and which is split into two subsystems Σ_1 and Σ_2. Assume the subsystem 1 contains a fraction f of the total number of degrees of freedom present in Σ. Since $\Sigma = \Sigma_1 + \Sigma_2$, the second system must contain a fraction $1-f$. If Σ_1 is small then its entropy will be thermal and since the thermal entropy is strictly additive, its fraction of the total entropy will be proportional to its size

$$S(\Sigma_1) = S_{\text{thermal}}(\Sigma_1) = f S_{\text{thermal}}(\Sigma). \tag{4.34}$$

Since the entanglement entropies of Σ_1 and Σ_2 are equal we find

$$S(\Sigma_2) = S(\Sigma_1) = f S_{\text{thermal}}(\Sigma). \tag{4.35}$$

Thus the information

$$I(\Sigma_1) = S_{\text{thermal}}(\Sigma_1) - S(\Sigma_1) \tag{4.36}$$

contained in Σ_1 vanishes.

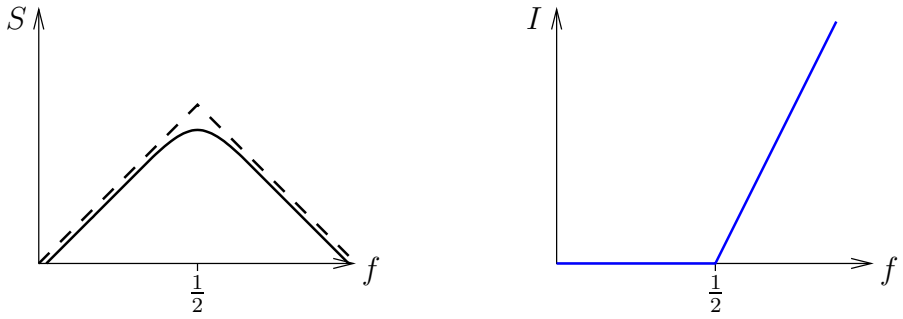

Figure 4.6: The left graph shows the fine grained entropy of a subsystem as it grows to include more and more degrees of freedom of the original system. The maximum value is reached when the system includes about $\frac{n}{2}$ of the n degrees of freedom. Until very close to this point, the entropy is given by the thermal entropy of the smaller system. Conversely the information contained in the subsystem is essentially zero until the size of the system approaches about half of the size of the total system.

Similarly if Σ_1 is large,

$$S(\Sigma_1) = S(\Sigma_2) = (1-f)S_{\text{thermal}}(\Sigma) \text{ and} \tag{4.37}$$
$$S_{\text{thermal}}(\Sigma_1) = fS_{\text{thermal}}(\Sigma). \tag{4.38}$$

In this case the information contained in Σ_1 is given by

$$I(\Sigma_1) = S_{\text{thermal}}(\Sigma_1) - S(\Sigma_1) \tag{4.39}$$
$$= fS_{\text{thermal}}(\Sigma) - (1-f)S_{\text{thermal}}(\Sigma) \tag{4.40}$$
$$= (2f-1)S_{\text{thermal}}(\Sigma). \tag{4.41}$$

The first result holds until Σ_1 is about half the size of Σ, *i.e.* the entropy in each subsystem closely follows the entropy of the *smaller* subsystem. Since the information content of Σ_1 is given by $S_{\text{thermal}}(\Sigma_1) - S(\Sigma_1)$, the information contained in Σ_1 is negligible as long as $f < 1/2$, *i.e.* as long as Σ_1 is small and $S(\Sigma_1) = S_{\text{thermal}}(\Sigma_1)$. Once Σ_1 is big, its information contents increases linearly as

$$I(\Sigma_1) = (2f-1)S_{\text{thermal}}(\Sigma) \tag{4.42}$$

until it reaches $I(\Sigma)$ once Σ_1 contains the whole system. Figure 4.6 illustrates the behaviour of both entropy and information in Σ_1 as Σ_1 grows.

4.4 An experiment

We will now apply the results of the last section to an important example.

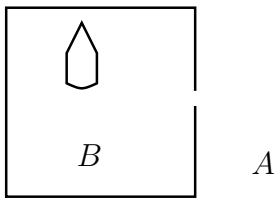

Figure 4.7: Combined system of bomb and environment before the bomb explodes.

Suppose we place a bomb in a box which is pierced by a very small hole. The whole arrangement is placed in vacuum. We will refer to the box as system B and to the outside as system A. Initially (before the bomb explodes) there are no photons anywhere in the system, and the bomb is in a pure state. Thus the thermal entropy and the entanglement entropy of both systems vanishes (we neglect the entanglement between vacuum states). Once the bomb explodes, the entropy inside the box rises and as photons slowly leak out of the box, so does the entropy in the outside environment. Eventually all the photons will have left the box, so that its entropy (thermal and entanglement) is zero again while the thermal entropy of the outside system has reached its maximum.

The evolution of entropy in the system would be as follows

Initial state Both systems are in a pure state. All entropies are zero.

Bomb explodes The thermal entropy in B suddenly rises to a non-zero value $S_{\text{thermal}}(B)$. However no photons have leaked out of the box yet and no entanglement exists between A and B, so $S(A) = 0$. And $S_{\text{thermal}}(A) = 0$ since A is still empty.

Photons leak out Entanglement arises between A and B. $S_{\text{thermal}}(B)$ decreases as photons escape the box, and $S_{\text{thermal}}(A)$ increases as photons appear outside of the box, S increases as A and B become entangled.

final state Everything has leaked out of the box. $S_{\text{thermal}}(B) = 0$ since no photons are in the box anymore. Thus $S(B) \leq S_{\text{thermal}}(B) = 0$ and since $S(A) = S(B)$ the entanglement entropy in A also vanishes. $S_{\text{thermal}}(A) \neq 0$ and is larger than $S_{\text{thermal}}(B)$ as the thermodynamical entropy always increases.

Thus the system evolves from a pure state into a mixed state and into a pure state again while photons move from the box to the outside. Substituting a black hole for the box and Hawking radiation for the exploding bomb, we see that it takes half the lifetime of the black hole before information appears in the Hawking radiation. Even

then local patches of the radiation will be very thermal. The whole system however will be in a pure state. Figure 4.8 illustrates this behaviour.

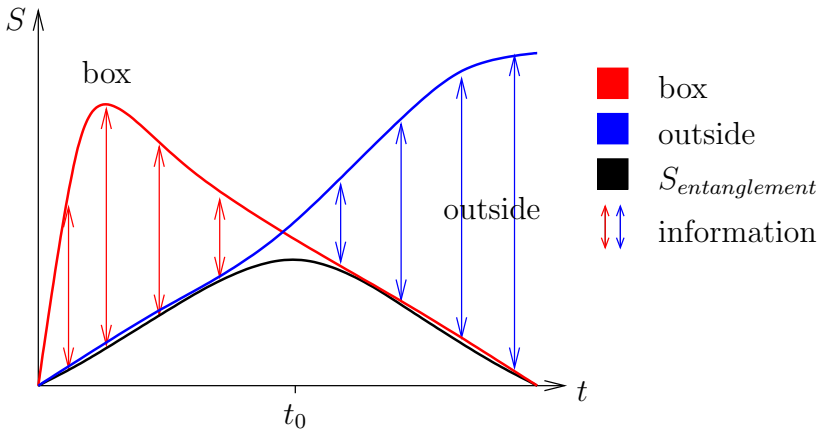

Figure 4.8: Before the crossover at t_0 there is essentially no information in the outside world. Information only starts to appear after half of the photons have leaked out after the information retention time t_0. By this time about half of the entropy and energy inside of the box have appeared outside.

Lecture 5

In this lecture we will explore the seeming contradictions that arise from the results of the previous lectures. This will lead us to suspect that Quantum Field Theory may have mislead us in its claim that information can be localized in space and time. We will then consider the situation in string theory, which offers a promising alternative perspective. (And although string theory may not describe *our* world, it does seem to consistently describe *some* world, and that world contains black holes.)

For interested readers, the lecture will close with a (very) brief introduction to string theory.

5.1 Three Clashing Principles

In the previous lectures we have explored basic principles of thermodynamics and General Relativity as they relate to black holes. As we have suggested, these principles seem to lead to contradictory conclusions. When we also consider a basic principle of Quantum Mechanics called the No-cloning theorem (described in Sec. 5.1.3), we see even more contradictions. We shall now briefly describe these principles and the contradictions that they imply.

5.1.1 Conservation of Quantum Information

The first principle we are concerned with is conservation of information. As we saw in the last lecture, this suggests that information passing through the event horizon will eventually be transmitted to the outside, roughly according to the behavior in Fig. 5.1. Essentially, the stretched horizon consists of a hot collection of thermal degrees of freedom, which absorb and re-radiate anything that hits the horizon. Although this behaviour hasn't been strictly proven for thermal systems, we will take it to be true.

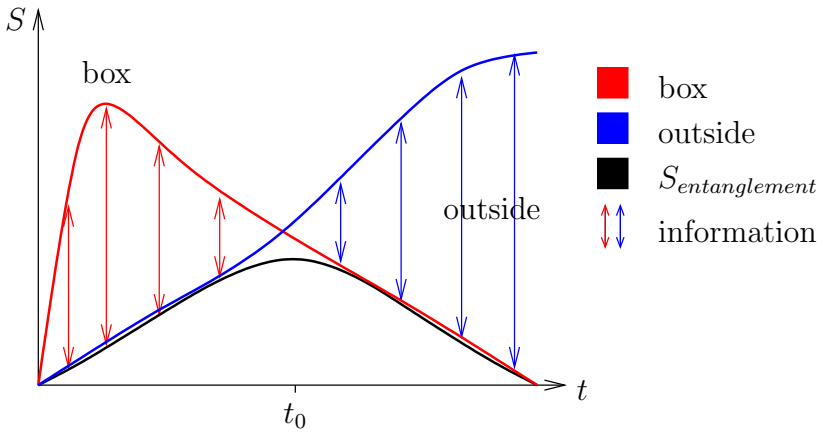

Figure 5.1: The behavior of entropy and information at the horizon of a black hole.

Note that our derivation in the last lecture ignored the "geometric entropy" resulting from entanglement of the ground states inside and outside the box. Also, as we mentioned last lecture, we have defined information slightly differently from how information theorists define it. They frequently call information "negative entropy," but we prefer the definition $I = S_{max} - S$, which makes the concept of entropy as "missing information" a bit more clear. Since the maximum entropy in QFT is typically infinite, for practical purposes we take S_{max} to be the entropy of a thermal system at fixed energy—i.e. the thermal entropy.

5.1.2 Equivalence Principle

The next principle is Einstein's strong equivalence principle, which states that spacetime is always locally flat. If we consider an extremely massive black hole of Schwarzschild radius $R_s \sim 10^9$ light years, then the proper curvature components at the horizon are $\sim 1/R_s^2 \sim 10^{-18}(\text{ly})^{-2}$. This means that an observer passing through the horizon observes only the vacuum of flat space.

We immediately see a seeming contradiction between the conclusion of the equivalence principle and that of quantum information theory. The temperature at the horizon is $T \sim \frac{1}{2\pi\rho}$ according to a Fido at a distance ρ from it. This implies that as an object nears the horizon, the Fido sees it evaporate and become part of the thermal radiation bath. According to information theory, any information contained in the object will be absorbed into the thermal bath and will eventually leak out via Hawking radiation. But according to a Frefo, the object passes through the horizon

unharmed, carrying its information with it.

5.1.3 No-cloning Theorem

The final principle is the No-cloning Theorem of Quantum Mechanics, which says that no quantum process can duplicate an arbitrary quantum state. Clearly, the first two principles seem to contradict this conclusion. According to the equivalence principle, the state of any particle will be unaffected by passing through the horizon; and according to information theory, the state will effectively be copied at the horizon. Hence, these two principles suggest that the event horizon acts as a surface of infinitesimal Xerox machines.

The proof of the no-cloning theorem is quite simple. Consider a process that perfectly duplicates quantum states. Since quantum mechanics is linear, this process can be represented as the action of a linear operator U, such that

$$U|\Psi\rangle = |\Psi\rangle \otimes |\Psi\rangle. \tag{5.1}$$

Now consider a state such as a spin oriented in the $+x$ direction: $|\rightarrow\rangle = \frac{1}{\sqrt{2}}|\uparrow\rangle + \frac{1}{\sqrt{2}}|\downarrow\rangle$. U must act on this state as

$$\begin{aligned} U|\rightarrow\rangle &= |\rightarrow\rangle \otimes |\rightarrow\rangle & (5.2)\\ &= \frac{1}{2}|\uparrow\rangle \otimes |\uparrow\rangle + \frac{1}{2}|\uparrow\rangle \otimes |\downarrow\rangle + \frac{1}{2}|\downarrow\rangle \otimes |\uparrow\rangle + \frac{1}{2}|\downarrow\rangle \otimes |\downarrow\rangle. & (5.3)\end{aligned}$$

But U is linear, so it also must act as

$$\begin{aligned} U|\rightarrow\rangle &= \frac{1}{\sqrt{2}}U|\uparrow\rangle + \frac{1}{\sqrt{2}}U|\downarrow\rangle & (5.4)\\ &= \frac{1}{\sqrt{2}}|\uparrow\rangle \otimes |\uparrow\rangle + \frac{1}{\sqrt{2}}|\downarrow\rangle \otimes |\downarrow\rangle. & (5.5)\end{aligned}$$

This contradicts the above result, so no such operator can exist.

5.2 Black Hole Complementarity

The black hole complementarity theorem claims that, although the above three principles appear to be contradictory, no such contradiction exists in individual experiments. The idea is just as in Bohr's complementarity principle. Complementarity principles such as these basically replace "and" with "or." Bohr's principle states that

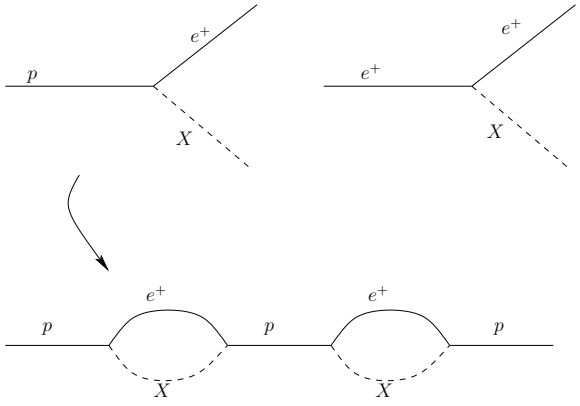

Figure 5.2: The interactions of a proton, positron, and X boson. The upper left diagram indicates the decay of a proton, the upper right diagram indicates the decay of a positron. These processes violate baryon number conservation. The lower diagram is the self-energy of the proton in this theory, suggesting that protons spend part of their time as a virtual pair of particles with the "wrong" baryon number.

we can treat a quantum state as either a particle or a wave in a given experiment, so no contradiction arises between the two descriptions; black hole complementarity states that in a given experiment we can only ever consider information to be either inside or outside the black hole, and it can never be shown to be both inside and outside.

Complementarity was famously debated in the early days of quantum mechanics, usually by way of various thought experiments. For example, consider Heisenberg's Microscope, which tries to track the position of an electron using a beam of photons. The photons give the electron a kick, altering its momentum even as they determine its position. And the more precisely one tracks the position, the higher the frequency of photons required, meaning the greater the "kick" to the electron. This argument showed that even if purely classical ideas are used, some form of complementarity arises (in this case, between knowledge of position and of momentum). The complementarity is explained on a deeper level by quantum mechanics. Analogously, thought experiments about black holes suggest the complementarity of information inside and outside the black hole, and hopefully a theory of quantum gravity can explain this result on a deeper level.

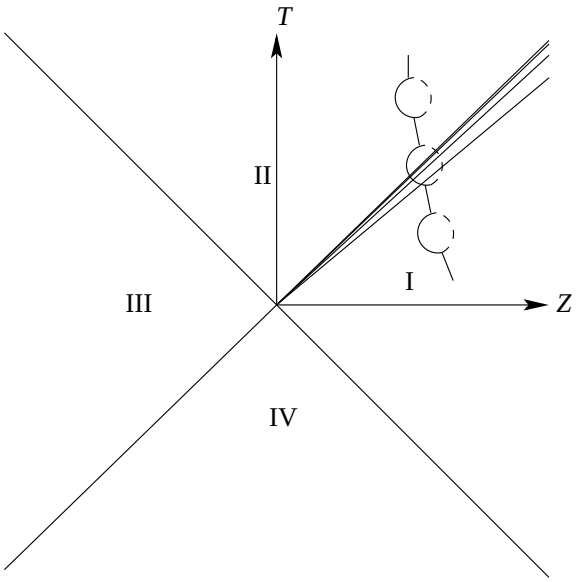

Figure 5.3: A proton passing through the event horizon. If it is caught in the "wrong" state near the horizon, the accumulating lines of constant time make it appear to an external observer to remain in that state indefinitely. Thus, a Fido sees the proton decay, violating baryon conservation, while a Frefo sees only an ordinary proton.

5.2.1 Gedanken Experiment 1

Consider an X boson with a large mass and which is coupled to fermions with a coupling constant $g \sim 1$. The coupling allows the processes $p \to e^+ + X$ and $e^+ \to e^+ + X$. If baryon number is conserved in this theory, then the first process implies that X has baryon number $+1$ and the second process implies that it has baryon number 0. Thus, baryon number cannot be conserved in these processes. But if the processes are possible, then a proton will spend part of its time as the virtual pair $e^+ + X$ (see Fig. 5.2). Although the heaviness of X prevents a decay into these particles (since such a process would violate energy conservation), the strong coupling between the fields means that the proton has a probability $\propto g^2 \sim 1$ of being in the virtual state with the "wrong" baryon number.

The "baryon number" that is actually measured is time-averaged over many oscillations between p and $e^+ + X$, and it is this number that is generally conserved. But consider a proton falling through an event horizon. As it approaches the horizon, a Fido sees its oscillations slow down. If it is caught in the $e^+ + X$ state very near

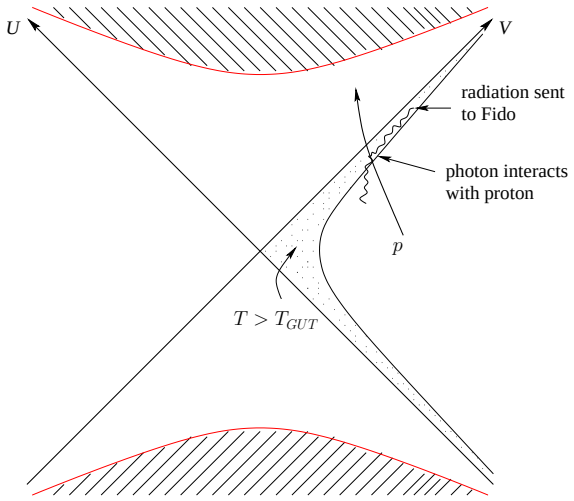

Figure 5.4: In order to determine whether a proton can decay near the horizon, a Fido must send measure radiation that interacted with the proton. But the window of time for the observation is very small, so any photon with a high enough frequency to make the measurement would have enough energy to itself cause the decay.

the horizon, then the Fido sees it stuck in that "wrong state" for all time (see Fig. 5.3). Thus, the Fido thinks that the proton has decayed. And he needn't wait for an infinite time to observe this; the proton arrives at the stretched horizon very quickly. But as far as a Frefo is concerned, this decay is impossible. These two observations seem contradictory.

However, we must consider how the Fido would make such a measurement in practice. Some radiation near the horizon would have to interact with the proton and then travel to the Fido. But clearly this interaction must occur in a very narrow window of time (see Fig. 5.4). Since the proton must be very close to the horizon for the interaction to be relevant for our measurement, we can consider the region within a distance $\sim 1/T_{GUT}$ of the horizon. (Here T_{GUT} is the grand unification temperature, above which the gauge symmetries of the electroweak and strong forces are unbroken according to a Grand Unified Theory). This means that the time in which a photon can interact with the proton in this region is $\tau \sim 1/T_{GUT}$. But in that case the photon must have a frequency (or energy, since $\hbar = 1$) $E \gtrsim T_{GUT}$, so its energy is sufficient to cause the proton to decay. Thus, the Fido would not be able to know if the decay would have occurred independent of his measurement.

5.2.2 Gedanken Experiment 2

As the next thought experiment, consider the following attempt to detect the duplication of quantum information at an event horizon (shown in Fig. 5.5). A Frefo O' and a Fido O begin outside the horizon. O' falls through the horizon carrying one bit of information; but according to O the bit is absorbed into the thermal bath at the horizon and can be determined by measurements of the Hawking radiation. Now, if O collects the information and then jumps through the horizon, O' can send him a signal containing the original bit, allowing O to verify the duplication of information.

The resolution of this paradox was suggested by John Preskill. From the discussion in the last lecture, we know that the information at the horizon is retained until the area of the horizon has decreased to half its initial value. Using this result, along with the result $t_{evap} \sim M^3G^2$, we see that the information retention time is $t \sim M^3G^2$. Thus, O must wait at least this long to determine the bit of information that O' carried. The elapsed Rindler time for O is then $\Delta\omega \sim M^2G$. Choosing $\omega = 0$ and $R \sim MG$ as O's initial coordinates, we find that O jumps through the event horizon at the Kruskal null coordinates $U = 0$ and $V \sim MGe^{M^2G}$.

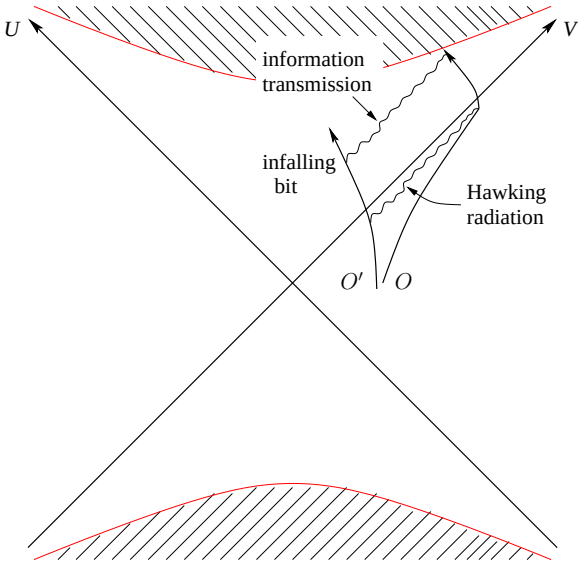

Figure 5.5: A Frefo O' carries a quantum bit of information through the horizon. A Fido O sees the information thermalized at the stretched horizon, and eventually reconstructs it from the Hawking radiation. O then jumps through the horizon and is sent a signal from O' containing the original information.

Since $UV = (MG)^2$ at the singularity, O arrives there after $\Delta U \sim MGe^{-M^2G}$. This means that if O' sends O a photon signal, he must send it within the interval $\Delta U \sim MGe^{-M^2G}$ after passing the horizon. But this corresponds to a proper time interval $\Delta\tau \sim MGe^{-M^2G}$ (since $d\tau^2 \sim MG/re^{-r/2MG}dUdV$), so the photon must have a trans-Planckian frequency $E \sim e^{M^2G}/(MG)$. Such a pulse would have sufficient energy to drastically alter the Schwarzschild spacetime, rendering the thought experiment inconsistent. Hence, O cannot observe quantum duplication.

Notice that the singularity plays an integral role in this scenario, since it limits the time available for transmission.

5.2.3 Gedanken Experiment 3

As our final thought experiment, consider an atom (of size $\sim 10^{-8}$cm) falling through the horizon ($\sim 10^9$ light years). According to a co-moving observer, the atom is completely unaffected by any tidal fields. But a Fido expects to see the atom ionize, denucleate, diffuse over the entire horizon, and come to thermal equilibrium.

Now suppose the Fido wants to determine if an atom actually maintains its integrity. He needs high-frequency (i.e. low-wavelength) photons to track its position, because of its large speed at the horizon. Since the atom's momentum approaches infinity as it approaches the horizon, the photons must eventually have energies $E \gg E_{Planck}$. Normally one would expect resolved distances to continually decrease with increasing energy, but two particles colliding at trans-Planckian energies will form a black hole. The larger the colliding particles' energies, the larger the radius of the black hole. And the black hole emits radiation with wavelength $\sim R_s$. Thus, as the Fido tries to measure the atom's position more and more precisely, he gets poorer and poorer resolution. This means that he actually sees the atom disperse, so he cannot determine whether it maintains its integrity as it passes through the horizon.

(The classical calculation for the formation of black holes by colliding particles, which is essential for our above conclusion, was performed by Payne and D'Eath).

5.3 Strings near the horizon

Let's now consider a string approaching the horizon, rather than a traditional particle. We shall see that we get a very different picture at the high time- and space-resolution provided by the horizon (both dimensions of resolution are needed, since localized particles are fast).

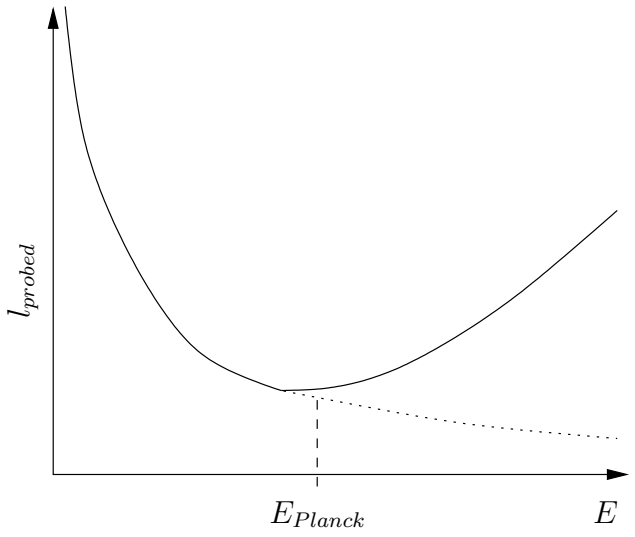

Figure 5.6: The usual QFT prediction for the distance probed by a particle of energy E (dotted line), and the prediction accounting for black hole formation at $E \gtrsim E_{Planck}$ (solid line).

In scale-invariant QFT we expect a Russian Doll pattern: for a conformal theory with localized particles, we will always find smaller particles (nucleons within nuclei, quarks within nucleons, etc.). As a particle nears the horizon, we can observe it with higher and higher "shutter speed," revealing smaller and smaller substructures without altering the particle's total size. (Of course, the third thought experiment outlined above shows that we cannot actually probe this behavior.)

However, in string theory we expect branching diffusion: for a free theory with no non-perturbative corrections, a particle repeatedly splits in two as our resolution improves. Any given path between splittings is a random walk of length $\sim \ell_s$ (the string length), so the total size grows as $R \sim \sqrt{n}$ for n divisions, just as in normal diffusion. The total number of particles grows much faster, as 2^n, so the density of particles keeps increasing.

How does this stochastic process result from string theory? A string is described by the standard string theory formula

$$X^i(\sigma = 0, \tau) = \ell_s \sum_{n \geq 0} \frac{a_n^i e^{-in\tau} + a_n^{i\,\dagger} e^{in\tau}}{\sqrt{n}}, \tag{5.6}$$

where X^i is a coordinate transverse to the string's direction of propagation, relative

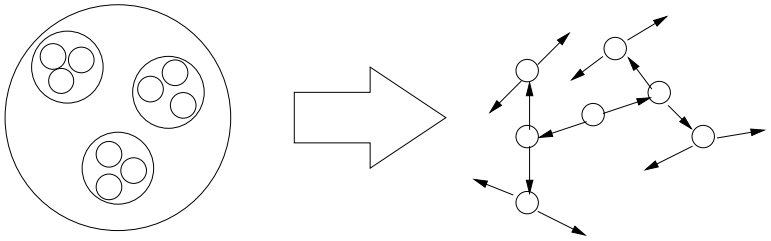

Figure 5.7: Left: a particle probed at greater and greater resolution in scale invariant QFT, revealing more and more substructure. Right: a particle probed at greater and greater resolution in free string theory, revealing a pattern of particles splitting and travelling on random paths.

to the string's center of mass. The variable σ and τ parametrize the string's world sheet: τ is light-cone time, equal to $U/(MG)$, and σ is a parameter running along the string's length. Since any measurement takes a finite time, we average this equation over a small time $\delta\tau$, giving

$$\bar{X}_i = \frac{\ell_s}{\delta\tau}\int_{\delta\tau}\sum_{n\geq 0}\frac{a_n^i e^{-in\tau} + a_n^{i\,\dagger} e^{in\tau}}{\sqrt{n}}d\tau \tag{5.7}$$

$$= \sum_{n=1}^{1/\delta\tau}\frac{a_n^i e^{-in\tau} + a_n^{i\,\dagger} e^{in\tau}}{\sqrt{n}}. \tag{5.8}$$

The second equality follows from the fact that any mode with a frequency $n > 1/\delta\tau$ will have multiple cycles within the interval $\delta\tau$, and will thus integrate to zero. Now, we should also take a quantum average of (the square of) this result:

$$\langle \bar{X}_i^2\rangle = \ell_s^2\langle 0|\sum_{n=1}^{1/\delta\tau}\frac{a_n^i e^{-in\tau} + a_n^{i\,\dagger} e^{in\tau}}{\sqrt{n}}\sum_{m=1}^{1/\delta\tau}\frac{a_m^i e^{-im\tau} + a_m^{i\,\dagger} e^{im\tau}}{\sqrt{m}}|0\rangle \tag{5.9}$$

$$= \ell_s^2\langle 0|\sum_{n=1}^{1/\delta\tau}\sum_{m=1}^{1/\delta\tau}\frac{a_n^i a_m^{i\,\dagger} e^{i(m-n)\tau}}{\sqrt{nm}}|0\rangle \tag{5.10}$$

$$= \ell_s^2\sum_{n=1}^{1/\delta\tau}\frac{1}{n} \tag{5.11}$$

For small $\delta\tau$, this sum diverges logarithmically (which we can see by converting the

sum to an integral), so

$$\langle \bar{X}_i^2 \rangle = \ell_s^2 \log \frac{1}{\delta\tau}. \tag{5.12}$$

Thus, as the resolution time decreases, revealing more and more modes of oscillation, the size of the string increases.

The exact relation between this phenomenon and the stochastic "splitting" process described above can be shown explicitly. In essence, each step forward in time for the string brings greater resolution for an observer, revealing more structure in the string. But it shall suffice to say that string theory provides a plausible explanation of the "spreading out" of particles over the horizon. It has also been shown that the information contained in the string becomes truly non-localized in this process (i.e. degrees of freedom evaluated at space-like separations do not commute).

5.3.1 The Basics of String Theory

String Theory is currently the most promising attempt to unify Quantum Mechanics with General Relativity. Its basic assumption is that the most fundamental type of object in the universe is a string, which can be open or closed, and that different types of particles can be described as different modes of oscillation in that string. Interactions consist of strings switching modes of vibration, splitting apart into more string, joining together into fewer strings, etc. The theory is unified in that a certain oscillation of a closed string corresponds to the graviton, which allows us to describe gravitational interactions in the same framework as the other three fundamental forces (keeping in mind that the graviton is a perturbation of a fixed background spacetime, not a description of the whole spacetime). Thus, string theory does away with the zoo of particles in particle physics, and largely reconciles Quantum Field Theory with General Relativity. However, in order for this to work, string theory requires many extra spatial dimensions, with the exact number fixed by demands such as Lorentz invariance and non-negative probabilities. The extra dimensions may be very small, such that we may never be able to detect them, or they may be large, with our three-dimensional space forming a hypersurface (called a D-brane) in the larger space. Typically, for these extra dimensions to be undetectable by low-energy experiments, they must be compactified by identification of points with one another (just as a line can be turned into a circle by identifying 0 with 2π).

Within this general picture of string theory there are two broad divisions: bosonic string theories and superstring theories. In bosonic string theories, strings exist in a 26 dimensional spacetime, and they can only describe bosons. In superstring theories, strings exist in a 10 dimensional spacetime with nine spatial dimensions and one time

dimension. Superstring theories assume supersymmetry, so fermions can be described as well as bosons. (Supersymmetry is basically the invariance of a Lagrangian under a transformation of fermions into bosons and vice versa.) Superstring theories are obviously the more realistic of the two types, since they includes fermions. Also, all superstring theories thus far (and an eleven-dimensional variation called M-theory) have been shown to be different limits of a single theory, suggesting a promising uniqueness. Thus, most research in string theory has been on the superstring side, but bosonic string theory can still be used in simple cases, since it has many of the same basic properties as superstring theory.

In either form of string theory, a string is described by coordinates X^μ, where $\mu = 0, ..., D-1$ in the D-dimensional spacetime, along with some internal degrees of freedom. A parametrization $X^\mu(\sigma, \tau)$ of the spacetime coordinates describes the two-dimensional worldsheet swept out by the string. Curves of constant τ represent the string in "space" (although these curves might not be spacelike), so τ is basically the "time" along the string's path, and σ parametrizes the string at each fixed τ. The starting point for describing the behavior of this string is then the Nambu-Goto action. For relativistic point particles, the action along a given worldline is proportional to the proper time along that worldine; analogously, for strings, the Nambu-Goto action over a worldsheet is proportional to the proper area over that worldsheet. Mathematically,

$$S \propto \int d\tau d\sigma \sqrt{\gamma}, \tag{5.13}$$

where $\gamma_{\alpha\beta}$ is the 2×2 metric on the worldsheet, induced by $g_{\alpha\beta}$.

Varying the Nambu-Goto action yields the equation of motion

$$\frac{\partial^2 X^\mu}{\partial \tau^2} - \frac{\partial^2 X^\mu}{\partial \sigma^2} = 0. \tag{5.14}$$

The solution to this equation can be written as a sum over normal modes, with the form

$$X^\mu = x_0^\mu + 2\ell_s^2 p^\mu \tau + \ell_s \sum_{n=1}^{\infty} \left(a_n^\mu e^{-in\tau} + a_n^{\mu\dagger} e^{in\tau} \right) \frac{\cos n\sigma}{\sqrt{n}}. \tag{5.15}$$

The first two terms represent the motion of the center of mass, while the sum represents oscillations. Using center of mass coordinates and setting $\sigma = 0$, we retrieve Eq. (5.6). The coefficients a_n^μ and $a_n^{\mu\dagger}$ are creation and annihilation operators if we quantize our strings, or ordinary complex constants if we do not.

Now, we have some freedom in our parametrization of the worldsheet. A particular parametrization is a gauge choice (in fact, Eqs. (5.14) and (5.15) only holds in a

certain class of gauge choices). The gauge we work in is called the "light cone gauge" in which we take $X^+ \propto \tau$, where $X^\pm = X^0 \pm X^1 = T \pm Z$. In this gauge we essentially take the string to be travelling along a light cone in the $+U$ direction. This condition leads to a constraint equation that allows us to find the longitudinal coordinates $X^\pm$ from the transverse coordinates X^i. We will discuss the longitudinal coordinates and the constraint equation in the final lecture.

Lecture 6

Information: Information is a good concept in that the distinction between states is eternal. Two systems that start out in different states will forever remain in different states, a concept otherwise known as unitarity of evolution. The concept of localized bits of information, however, is not a good concept in this sense.

An example of localized information would be an observer making measurements on an electron in a train who then communicates her information to me. After this I am able to predict the waveform of the electron using this information. This is the idea that will fail in trans-Planckian physics.

We have seen a few Gedanken experiments that demonstrate how this concept breaks down not over very short distances but over arbitrarily large distances, *i.e.* at the macroscopic scale. This will lead to a relativity of shutter speed. Short timescale (or alternatively high energy) physics is fundamentally different from the usual low energy quantum field theories. A shorter timescale will not always allow us to resolve more details of a process. This is very different from ordinary quantum field theory where a higher shutter speed will always allow one to get more information. So far we have always fallen into the trap of using quantum field theories where they are invalid.

6.1 An infalling string

An elementary point particle falling into a Schwarzschild black hole will get closer and closer to the horizon as Schwarzschild time increases. A string on the other hand will never come closer to the horizon than a string length (as seen from the outside).

To flesh out this idea, consider a string falling toward a black hole. We have already seen in the last lecture that it will expand laterally. Now we calculate how its length along the direction of motion changes.

First we introduce light cone coordinates X^+ and X^-. *Recall the conformal*

coordinates (ω, ρ) close to the event horizon introduced in lecture 2.

$$d\tau^2 \simeq \rho^2 d\omega^2 - d\rho^2 \tag{6.1}$$

We introduce Cartesian coordinates (T, Z)

$$T = \rho \sinh \omega \qquad\qquad Z = \rho \cosh \omega, \tag{6.2}$$

where we choose the negative sign for Z for the right hand Rindler wedge. In terms of these coordinates the metric reads

$$d\tau^2 = dT^2 - dZ^2. \tag{6.3}$$

Obviously light rays move along lines of $dT = \pm dZ$ and we define the light cone coordinates $X^\pm$ as

$$X^\pm = \frac{T \mp Z}{\sqrt{2}} = \mp \frac{\rho e^{\mp \omega}}{\sqrt{2}}. \tag{6.4}$$

The signs here are the opposite of those in the last lecture because we consider a string in region III rather than in region I.

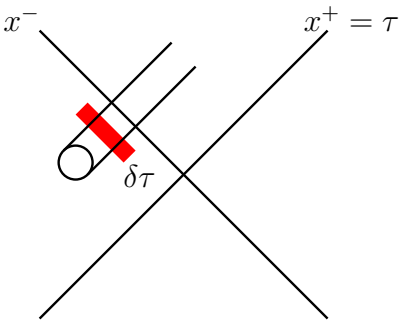

Figure 6.1: World tube of a string falling toward the horizon.

A string moving toward the horizon traces out a world tube along its trajectory. For a string moving inward at the speed of light, we have the equality $\tau = X^+$. Inspection of Eq. (6.4) reveals that for an ingoing string, which travels on a line $X^- = const$, we find $\rho = e^{-\omega}$ and $X^+ = -e^{-2\omega}$. X^+ continues through the horizon while ω accumulates. An interval ΔX^+ of light-cone time appears enormously long (by a factor of $e^{2\omega}$) to an observer far away from the horizon. This allows an outside observer to observe the string with a very high time resolution. In the light cone gauge the evolution equation of the transverse (*i.e.* along the horizon) components of the position of the string are given by ([6] p. 157)

$$\frac{\partial^2 X_i}{\partial \tau^2} - \frac{\partial^2 X_i}{\partial \sigma^2} = 0, \tag{6.5}$$

where X^i is a scalar field of mass dimension 0. The evolution of X^- is given entirely by the constraints

$$\dot{X} \cdot X' = 0 \qquad\qquad \dot{X}^2 + X'^2 = 0, \tag{6.6}$$

which represent the invariance of string theory under a reparametrization of σ. Explicitly spelled out the first constraint reads

$$\frac{\partial X^-}{\partial \sigma}\frac{\partial X^+}{\partial \tau} + \frac{\partial X^+}{\partial \sigma}\frac{\partial X^-}{\partial \tau} - \frac{\partial X^i}{\partial \sigma}\frac{\partial X^i}{\partial \tau} = 0 \tag{6.7}$$

or, using $X^+ \equiv \tau$,

$$\frac{\partial X^-}{\partial \sigma} = \frac{\partial X^i}{\partial \sigma}\frac{\partial X^i}{\partial \tau}, \tag{6.8}$$

which implies that X^- is a field of mass dimension $+1$. On dimensional grounds then, fluctuations in X^- of a string a distance $\delta\tau$ away from the horizon scale as

$$\sqrt{\langle(\bar{X}^-)^2\rangle} \propto \frac{\ell_S^2}{\delta\tau}. \tag{6.9}$$

Since $\tau = X^+$, this gives rise to an uncertainty relation

$$\Delta X^- \Delta X^+ = \ell_S^2 = \alpha' \tag{6.10}$$

between the light cone coordinates in free string theory. (α' here is defined to be the reciprocal of the string tension.)

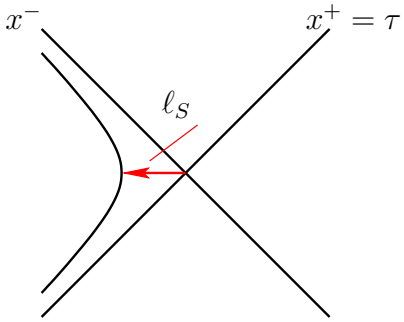

Figure 6.2: A string moving toward a black hole is spreading out as it is sampled with ever increasing time resolution. It never gets closer to the horizon than ℓ_S.

Geometrically this means that as the string gets closer and closer to the horizon the uncertainty in X^+ decreases which prompts a corresponding increase in ΔX^-. Seen from the outside, the string fills the region between the hyperbola $\Delta X^+ \Delta X^- = \alpha'$ and the $X^\pm = 0$ lines. Using Eq. (6.4) we see that the parabola $X^+X^- = \alpha'$ corresponds to

$$X^+X^- = \alpha' \Leftrightarrow \rho^2 = 2\ell_S^2. \tag{6.11}$$

The string is seen to hover a proper distance of ℓ_S above the horizon when probed from the outside, *i.e.* probed with ever increasing time resolution.

For an outside observer at least, this is very different from a point particle which is seen to move closer and closer to the horizon.

6.2 Black hole entropy in string theory

6.2.1 A black hole as a single string

Different from what is often done in the literature, we will calculate the entropy of a Schwarzschild black hole and not of an extremal black hole. We will however show the derivation for a D-dimensional Schwarzschild black hole. The metric of such a black hole is given by

$$d\tau^2 = \left(1 - \frac{R_S^{D-3}}{r^{D-3}}\right) dt^2 - \left(1 - \frac{R_S^{D-3}}{r^{D-3}}\right)^{-1} dr^2 - r^2 d\Omega_{D-2}^2, \tag{6.12}$$

where

$$R_S = \left(\frac{16\pi(D-3)GM}{\Omega_{D-2}(D-2)}\right)^{\frac{1}{D-3}} \tag{6.13}$$

is the radial coordinate of the event horizon, $d\Omega_{D-2}^2$ is an element of solid angle on the $(D-2)$-dimensional unit sphere, and r is the usual areal radius such that the area of spheres is given by

$$A = r^{D-2}\Omega_{D-2}, \tag{6.14}$$

where Ω_{D-2} is the area of the unit sphere. The Bekenstein-Hawking entropy of the horizon is just one quarter of its area

$$S = \frac{A}{4G} \propto (MG)^{\frac{D-2}{D-3}}. \tag{6.15}$$

Our task is to show that this is related to a counting of states in string theory. From now on we are going to use units in which $\alpha' = \ell_S^2 = 1$ and $\ell_P^{D-2} = G = g^2$, where g is the string coupling constant in D dimensions. In these units the energy of a string (in the light cone frame) is given by $E = \frac{m}{2}$. And also, since string theory in the light-cone frame is just the quantum field theory of $D-2$ scalar fields X^i (by Eq. (6.5)), we have $E = \pi T^2(D-2)$, where T is a fictitious temperature. Written in terms of the same temperature the entropy is $S = 2\pi T(D-2)$. Combining these three equation we find that

$$S = \sqrt{2(D-2)\pi}m, \tag{6.16}$$

or $S = \sqrt{2(D-2)\pi}m\ell_S$ after restoring the units. Thus a string's entropy is proportional to its mass and thus its length.

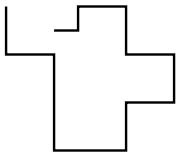

Figure 6.3: lattice string

This is not hard to understand if we consider, for example, a lattice string wandering around as shown in Fig. 6.3. Since the string has constant mass density its mass is proportional to its length. On the other hand the entropy is proportional to the number of vertices that the string occupies *i.e.* its length. Therefore the entropy is proportional to the string's mass.

The inclusion of subleading terms in the entropy yields

$$S = \sqrt{2(D-2)\pi} m\ell_S - c \ln m\ell_S, \tag{6.17}$$

where c is some positive constant. Equation (6.17) show that for a given large mass the most likely configuration is one long string. Consider the entropy of n smaller strings each of mass $\frac{m}{n}$:

$$S_n = nS(\frac{m}{n}) = \sqrt{2(D-2)\pi} m - cn \ln \frac{m}{n} \tag{6.18}$$

This prefers one long string over several shorter ones.

In order to understand this consider a long string and a collection of smaller strings whose total length equals the length of the large string. For every fluctuation of one of the short strings, an equivalent fluctuation of a segment of the long string exists. Thus the long string allows for all of the possible fluctuations that the short strings support. On top of this, however, the long string also possesses long wavelength fluctuations not possible for the smaller strings. Thus the long string supports more states and is entropically preferred.

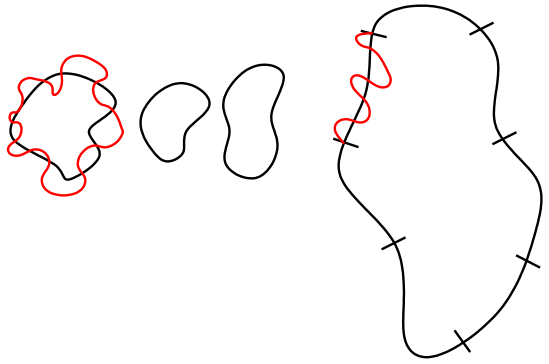

Figure 6.4: A long string supports at least as many configurations as a set of shorter strings formed by chopping up the long string.

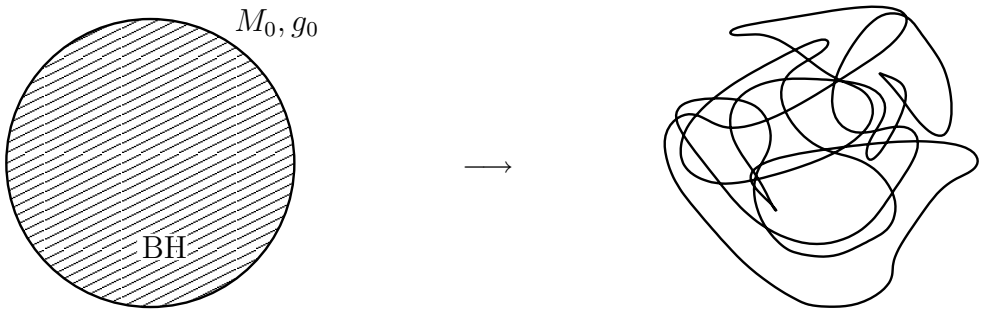

Figure 6.5: The left pane shows the initial state of the system, a Schwarzschild black hole. The right pane shows the final string into which the black hole evolves when g is dialed down to zero.

In string theory the coupling constant g is not fixed, but can be changed via boundary conditions. In the simplest string theories this is mediated by a dilaton field which describes the coupling strength. As g^2 is turned down, gravity vanishes (since Newton's constant is linked to g by $G = g^2$) and the mass of the black hole is converted into free strings (every other object becoming infinitely massive). Recalling the discussion above we expect the most likely configuration to be a single long string which is the object of largest entropy for a given mass, just as the black hole was for $G > 0$. At the end of the calculation we will recover the known expression for the Bekenstein-Hawking entropy of a black hole, further justifying this assumption. Ignoring possible problems caused by the singularity inside of the black hole, the transition from a black hole to a string is reversible and the system will oscillates back and forth between the two states. This oscillation will happen most easily if the entropies of the black hole and the string are equal, *i.e.* $S_{\text{black hole}} = S_{\text{string}}$. By changing g^2 slowly the evolution is adiabatic and the mass M will change over the evolution while the entropy remains constant.

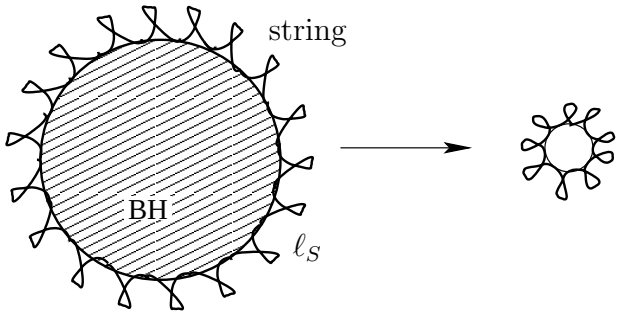

Figure 6.6: The black-hole—string transition occurs when the black hole radius approaches the string length

The transition from a black hole to a string occurs when the black hole radius is comparable to the string length. Figure 6.6 illustrates this. The black hole is covered with stringy matter up to a distance ℓ_S away from the horizon. As $g \to 0$ the black hole shrinks, while the size of the string loops stays constant. Once the horizon is approximately of the same size as the string scale the notion of a black hole becomes meaningless. A more detailed calculation of, for example, the force experienced by the strings confirms this.

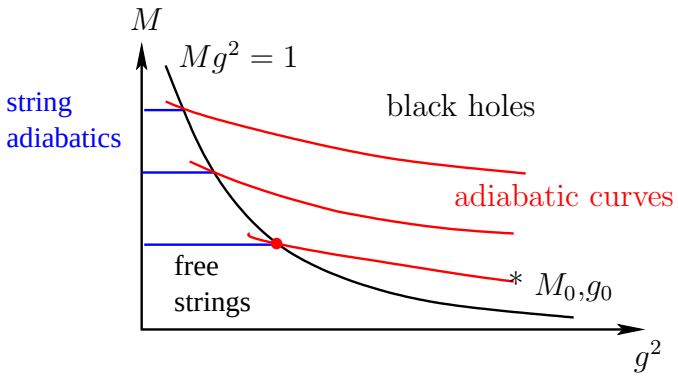

Figure 6.7: Adiabats of a black-hole-string system. The phase transition occurs when $Mg^2 = 1$ or in string units $R_S = 1$.

In four dimensions the transition seems to be first order, in higher dimensions it is smooth. The adiabatic curves in the black hole region are determined classically as G (or alternatively g^2) is varied (which corresponds to a change in the dilaton

background field) from

$$M g^{\frac{2}{D-2}} = M_0 g_0^{\frac{2}{D-2}}, \tag{6.19}$$

where M_0 and g_0 are the initial mass of the black hole and the initial string coupling constant, respectively. Once the crossing point is found we calculate the entropy of the string and hence of the black hole

$$S \propto (M_0 G_0)^{\frac{D-2}{D-3}} {G_0}^{-1}, \tag{6.20}$$

which (for $D = 4$) scales with the mass M_0 of the black hole as the Bekenstein-hawking entropy does. The numerical coefficient in front of it can also be calculated by this method. The difficulties encountered in doing so are not so much conceptual as they are due to our ignorance of where the transition occurs.

6.2.2 Direct state-counting

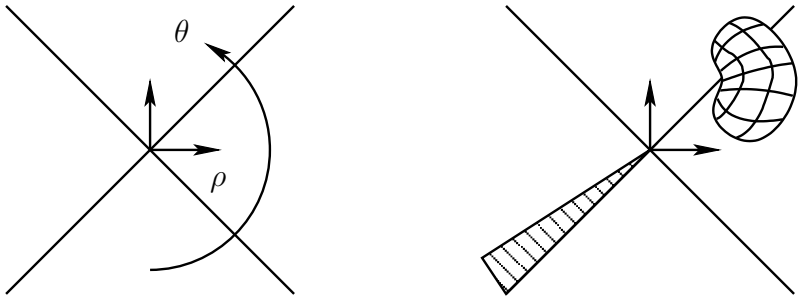

Figure 6.8: Left pane: Euclideanized Rindler space; θ is an angle and the horizon extends out of the page. Right pane: Conical deficit and a region whose Feynman diagrams do not contribute to the entropy

A different outlook on the entropy of a horizon is provided by the following. Consider Rindler space for which the Hamiltonian is given by $H_R = \partial_\theta$, where θ is the Euclideanized (imaginary) Rindler time. As seen in the previous lectures the (inverse) temperature is given by $\beta = 2\pi$. The standard thermodynamic definition for the free energy in terms of the partition function

$$F = -\frac{1}{\beta} \ln Z \tag{6.21}$$

yields

$$S = -\frac{\partial F}{\partial T}. \tag{6.22}$$

This presents a problem for us since we cannot vary F with respect to the constant $T = \frac{1}{2\pi}$. We have to add a conical deficit into the angle θ and make it periodic with period β instead of 2π. The deficit is then $2\pi - \beta$ ([3]). *Roughly speaking the leading gravitational term of the partition function*

$$Z(\beta) \propto \int_{\mathcal{F}} d[g] \int d[\phi] e^{-I[g,\phi]}, \tag{6.23}$$

where the integral is over all Euclidean metrics g and all matter configurations ϕ and $I[g, \phi]$ is the combined action, is given by

$$Z(\beta) \propto e^{-I_{EH}[g]}, \tag{6.24}$$

where I_{EH} is the Einstein Hilbert action. Thus combining Eqs. (6.21) *and* (6.22) *the entropy is*

$$S = -\frac{\partial}{\partial T}\left(\frac{1}{\beta} I_{EH}\right) = \frac{A}{8\pi G} \frac{\partial}{\partial T} \left(\frac{2\pi - \beta}{\beta}\right), \tag{6.25}$$

where we used the result from [4] that

$$I_{EH} = -\frac{1}{8\pi G} \int \epsilon_g R = -\frac{A}{8\pi G}(2\pi - \beta). \tag{6.26}$$

And finally evaluating the derivative we recover

$$S = \frac{A}{4G}, \tag{6.27}$$

which agrees with the Bekenstein-Hawking formula.

In string theory, on the other hand, the metric is no longer dynamical and the the entropy is due to strings surrounding the conical singularity. The lowest-order matter part of the vacuum contribution to the functional integral Eq. (6.23) *is described in terms of first quantized particle paths forming single loops. The (logarithm of the) functional integral represents the sum of the contribution of the loops. A single loop that does not intersect or encircle the origin is unaware of the existence of the deficit angle $2\pi - \beta$, since the spacetime is perfectly regular away from the origin. Thus each such loop contributes a constant amount to the path integral. Summing over all loops*

we have to integrate over the full range $0 \ldots \beta$ *of the Euclidean angle* θ *and thus pick up a factor* β *in front of the constant contribution from a single loop.* Thus the contribution to $\ln Z$ of all of these diagrams is linear in β and since

$$F = -\frac{1}{\beta} \ln Z, \tag{6.28}$$

we divide out one factor of β. Therefore these diagrams do not contribute to the entropy. The only ones that do contribute are then those that encircle the origin. In string theory the lowest-order contributions are from strings living on genus zero (singly connected) world sheets. On long scales the partition function of these is equal to the classical Einstein action (see [2]) and we recover the previous result

$$S = \frac{A}{4G}, \tag{6.29}$$

by picking up the calculation done before in Eq. (6.25). In [2] arguments are provided

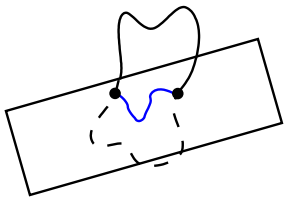

Figure 6.9: A closed string stradling the horizon appears to an external observer as an open string whose end points are fixed to the horizon.

for the fact that the remaining contributions from higher order terms of genus zero worldsheets and higher genus terms either vanish or appear as subleading terms in the entropy. Going beyond the results of the classical calculation, it is now possible to identify which states have to be counted. At a given surface of time the pierced world sheet of a string is seen by an observer located on the outside of the horizon as an open string whose end points are fixed to the horizon. These typically very excited strings account for the entropy.

Bibliography

[1] Don N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71(9):1291–1294, Aug 1993.

[2] Leonard Susskin and John Uglum. String physics and black holes. *Nuclear Physics B - Proceedings Supplements*, 45(2-3):115–134, 1996/2.

[3] L. Susskind and J. Uglum. Black hole entropy in canonical quantum gravity and superstring theory. *Physical Review D*, 50:2700, 1994. hep-th/9401070.

[4] Leonard Susskind. Some speculations about black hole entropy in string theory, 1993. hep-th/9309145.

[5] Leonard Susskind. *Introduction to black holes, information and the string theory revolution : the holographic universe*. World Scientific, Hackensack, NJ, 2005.

[6] Barton Zwiebach. *First course in string theory*. Cambridge University Press, New York, 2004.