Black Holes from A to Z

Andrew Strominger

Center for the Fundamental Laws of Nature, Harvard University,
Cambridge, MA 02138, USA

Last updated: July 15, 2015

Abstract

These are the lecture notes from Professor Andrew Strominger's *Physics 211r: Black Holes from A to Z* course given in Spring 2015, at Harvard University. It is the first half of a survey of black holes focusing on the deep puzzles they present concerning the relations between general relativity, quantum mechanics and thermodynamics. Topics include: causal structure, event horizons, Penrose diagrams, the Kerr geometry, the laws of black hole thermodynamics, Hawking radiation, the Bekenstein-Hawking entropy/area law, the information puzzle, microstate counting and holography. Parallel issues for cosmological and de Sitter event horizons are also discussed. These notes are prepared by Yichen Shi, Prahar Mitra, and Alex Lupsasca, with all illustrations by Prahar Mitra.

Contents

1 Introduction

In the last decade, black holes have come into the forefront of modern science: in astronomy, mathematics, and physics.

The first indirect astronomical observation of black holes occurred in the 70s, but scientists were skeptical. Nowadays, new black holes are detected on a daily basis. It is now known that there is a supermassive black hole at the centre of our galaxy. The masses of black holes vary widely and can reach a million solar masses. Despite having tonnes of experimental evidence for the existence of black holes, there is still a lot about it that is unknown. The Event Horizon Telescope will, in the future, image the black hole at the centre of our galaxy to within its Schwarzschild radius. This observation is sure to produce more interesting data.

Black holes are solutions to Einstein's field equations. Mathematically, these equations are extremely interesting and complex differential equations. New properties of them are being discovered by mathematicians even today. Despite significant developments in our understanding of these equations, there are still several unproven conjectures surrounding them. Proving these conjectures is an important field in mathematics.

To physicists, black holes are enigmatic objects with lots of interesting physical properties. To me, the most interesting equation in modern physics is the *Bekenstein-Hawking area/entropy law*,

$$S_{\text{BH}} = \frac{Ac^3}{4G_N\hbar}, \tag{1.1}$$

where A is the surface area of the black hole. Note that the formula above is in its dimensionless form. Under the usual definition of entropy, the right hand side should be multiplied by Boltzmann's constant k_B. This equation is mostly derived from classical considerations, but our main understanding of it will come in the next semester, when we discuss its quantum aspects. Interesting enough, this relates all areas of modern physics: statistical mechanics, gravity, and quantum mechanics. And it applies to not only black holes horizons but also to other kinds of horizons. Note that the factor of 4 was computed by Hawking, while the rest of the formula was determined by Bekenstein using classical principles. It's deep, explicit, and not understood! The challenge to modern day physicists is to explain this formula. This will require a rethinking of physics.

Now, all the laws of physics for the last two thousand years have been built on the foundation of determinism, in the sense that there *are* laws. That is, if we know the present, there are laws that determine the future and can be used to retrodict the past. This is certainly true in classical physics. It is also true in quantum physics, where we determine the state, instead of classical variables, in the future or the past. We refer to this as the *unitarity* of quantum mechanical systems. Now, the causal structure in general relativity is so convoluted that we don't necessarily know what it means for the laws to be deterministic, since we don't even understand what the past and future are. This is related to Hawking's information puzzle. These are the motivations for this course and we'll learn how we can address these problems.

The prerequisite for this course is some knowledge of general relativity. It will be useful to have taken a course on quantum field theory, but it's not required.

Our course this semester is largely classical, leaving the quantum aspects to Spring 2016.

2 Causal structure, event horizons and Penrose diagrams

The most basic question we can ask about a spacetime is the causal relation between two points, x and x'. There are basically six possibilities as depicted in Figure 2.1. x' being in the past or future of x, x' being on the past or future lightcone of x, x' having a spacelike separation from x, or there may be no separation.

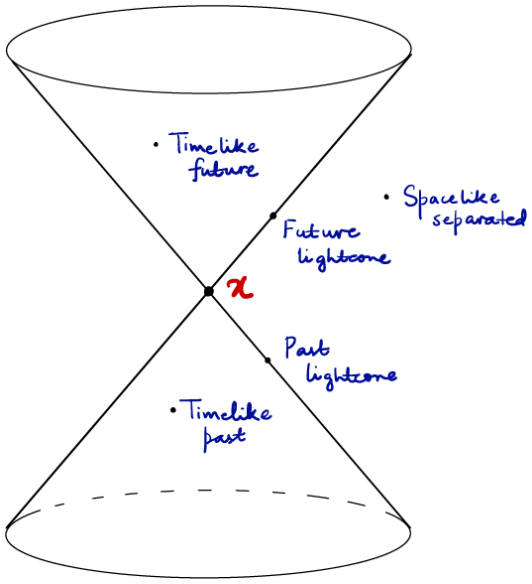


Figure 2.1: Causal relations in a Lorentzian manifold

Understanding the causal relationship between different points is subtle in the context of black holes, and *Penrose diagrams* are the tools we use.

2.1 Minkowski space

First note that it's a convention to refer to some spacetime as simply space. We start from the metric

$$ds^2 = -dt^2 + dr^2 + r^2 d\Omega_2^2, \tag{2.1}$$

where $d\Omega_2^2$ is the unit metric on the two-sphere. We move to null coordinates for simplicity and take

$$u \equiv t - r, \qquad v \equiv t + r, \tag{2.2}$$

such that $u - v > 0$. Then we have the metric

$$ds^2 = -dudv + \frac{1}{4}(u-v)^2 d\Omega_2^2. \tag{2.3}$$

Figure 2.2 portrays the situation, where

- $\mathscr{I}^{\pm}$ indicates where light rays come from or go to, i.e., the past or future infinity;
- $i^{\pm}$ indicates past or future timelike infinity;
- i^0 indicates spatial infinity.

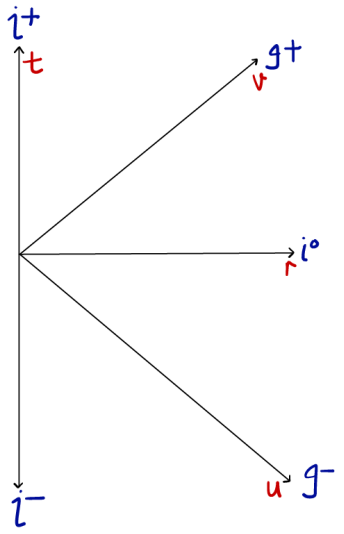


Figure 2.2: Boundaries of Minkowski space

The problem is that these interesting infinities can't be fit onto the blackboard, or any two-dimensional plane! So Penrose came up with ways to put these back to the blackboard. Consider new coordinates

$$v \equiv \tan \tilde{v}, \qquad u \equiv \tan \tilde{u}, \tag{2.4}$$

where now we have $\tilde{u}, \tilde{v} \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Since we don't need to know the derivatives explicitly, we simply have

$$ds^2 = \frac{du}{d\tilde{u}} \frac{dv}{d\tilde{v}} \left(-d\tilde{u} d\tilde{v} + \tilde{r}^2(\tilde{u}, \tilde{v}) d\Omega_2^2 \right). \tag{2.5}$$

Next, we define a new metric $d\tilde{s}^2$ such that

$$d\tilde{s}^2 \equiv \frac{d\tilde{u}}{du} \frac{d\tilde{v}}{dv} ds^2 = -d\tilde{u} d\tilde{v} + \tilde{r}^2 d\Omega_2^2. \tag{2.6}$$

We want to know how different points in spacetime are causally related. Suppose x' is in the causal past of x iff there is an everywhere past directed timelike line going from x to x', i.e., the norm of the tangent vector is everywhere negative along the curve. If this was the case when computed with s, it will also be the case when computed with $\tilde{s}$, because $\tilde{u}$ and $\tilde{v}$, being monotonically increasing functions, make $\frac{d\tilde{u}}{du} \frac{d\tilde{v}}{dv}$ positive. Since the causal structure of this metric depends only on the overall sign of ds^2, that of a spacetime defined by metric $d\tilde{s}^2 = \Omega^2 ds^2$ for any real factor Ω is the same as that of a spacetime defined by ds^2. This kind of rescaling is called a *Weyl transformation*. It is a local rescaling of the metric tensor that produces another metric in the same conformal class.

Now, the original range of u and v is bounded by $v = u$, $v = \infty$, and $u = -\infty$, which give

$$\tilde{v} = \tilde{u}, \qquad \tilde{v} = \frac{\pi}{2}, \qquad \tilde{u} = -\frac{\pi}{2} \tag{2.7}$$

respectively. So if we fix the light rays to be at 45°, we arrive at Figure 2.3.

The Penrose diagram for the Minkowski space explains all of its causal structure. First, if we go to the future timelike infinity, then all of the spacetime is in our causal past, i.e., we can see the whole universe if we wait long enough. Similarly, in the infinite past, all of the spacetime is in our future. Second, if we take a complete spacelike slice, then all of the future spacetime is in the *domain of dependence*, i.e., it can be determined by the evolution of the slice. Minkowski space is the only spacetime that has these properties.

Further, null infinities are larger than timelike infinities. If we have two light rays going at the same angle, we can ask which one reached the future null infinity first. But if we have two eternal objects, we cannot ask which reaches the future timelike infinity first. Hence, there are more structures to null infinities than timelike infinities.

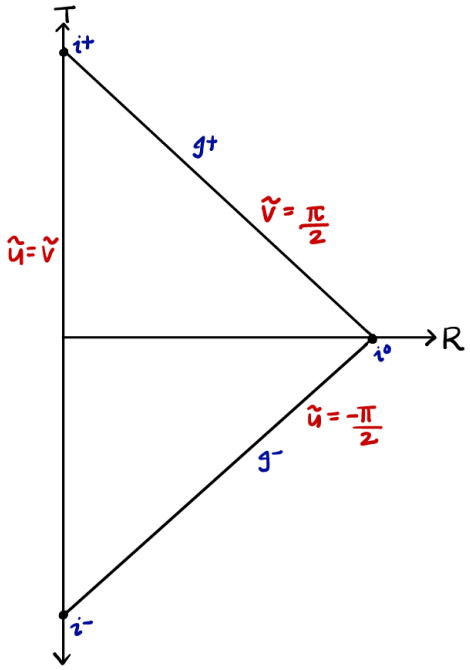


Figure 2.3: Penrose diagram of Minkowski space

Note that these transformations we made above do change length scales, and that we've suppressed the angular dependence so that every point on the interior is an S^2, except for the origin, which is a point.

2.2 de Sitter space

de Sitter space is a maximally symmetric solution to the Einstein equations

$$G_{\mu\nu} = \Lambda g_{\mu\nu} \tag{2.8}$$

with positive cosmological constant $\Lambda = 3/l^2$. To get de Sitter space, we start with a five-dimensional Minkowski space, and take a hyperboloid in that space, i.e., a four-dimensional sub-manifold. The four-dimensional sub-manifold is de Sitter space. Now, the line element for the five-dimensional Minkowski space is

$$ds_5^2 = -(dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2 + (dx^4)^2. \tag{2.9}$$

The Lorentz group of the 5-dimensional Minkowski space is $SO(4,1)$ and if we take a hyperboloid such that

$$-(X^0)^2 + (X^1)^2 + (X^2)^2 + (X^3)^2 + (X^4)^2 = l^2, \tag{2.10}$$

then the hyperboloid is mapped onto itself through $SO(4,1)$. Evident from the above constraint and Figure 2.4, as we move away from X^0, the radius of the sphere increases.

Now we consider this in other coordinates and look for a metric on the sub-manifold. We don't need five coordinates for a four-dimensional manifold. Setting $l \equiv 1$, take

$$X^0 = \sinh\tau, \tag{2.11}$$
$$X^1 = \cos\theta\cosh\tau, \tag{2.12}$$
$$X^2 = \sin\theta\cos\phi\cosh\tau, \tag{2.13}$$
$$X^3 = \sin\theta\sin\phi\cos\psi\cosh\tau, \tag{2.14}$$
$$X^4 = \sin\theta\sin\phi\sin\psi\cosh\tau. \tag{2.15}$$

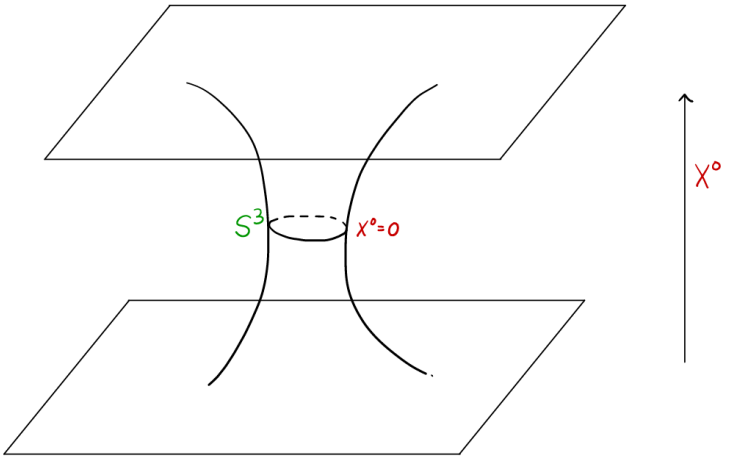


Figure 2.4: Varying sizes of S^3's in de Sitter space

Then we have

$$ds_4^2 = -d\tau^2 + \cosh^2\tau d\Omega_3^2, \tag{2.16}$$

where τ is some proper time along the hyperbola, and the $\cosh^2\tau$ tells us that the S^3 extends to the infinite past and future. Next we do a coordinate change, taking

$$\cosh\tau \equiv \frac{1}{\cos T}. \tag{2.17}$$

This gives

$$ds_4^2 = \frac{1}{\cos^2 T}(-dT^2 + d\Omega_3^2), \tag{2.18}$$

where $T \in (-\frac{\pi}{2}, \frac{\pi}{2})$. So we can define a new metric

$$d\tilde{s}_4^2 \equiv \cos^2 T\, ds_4^2 = -dT^2 + d\Omega^2. \tag{2.19}$$

Figure 2.5 shows the corresponding Penrose diagram. Every point inside the region is an S^2 parametrised by ϕ and ψ, except for the poles, where they are only points. With an observer on the South Pole, we refer to region I as the *causal diamond* while region IV is inaccessible.

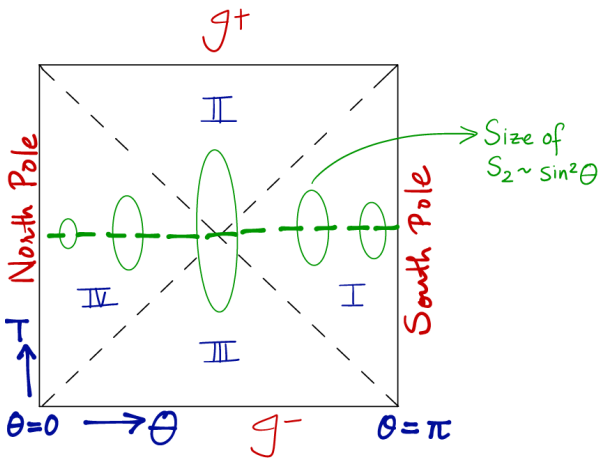


Figure 2.5: Penrose diagram of de Sitter space

As shown by the dashed lines, a light ray going from the beginning to the end of the universe just barely makes it from the north pole to the south pole. And unlike Minkowski space, if we wait long enough to get to the future

timelike infinity, our past is only half of the spacetime; and even if we go to the infinite past, only half of the universe is in our future. So causally, with respect to an observer, the universe divides into four causal regions. Note that in the Minkowski case, the whole spacetime is in the causal diamond, and that de Sitter space has no spatial infinity.

Another coordinate system used commonly for de Sitter space goes as follows.

$$X^0 \equiv \sinh t + \frac{1}{2}\vec{x}^2\, \mathrm{e}^t, \tag{2.20}$$

$$(X^1, X^2, X^3) \equiv \vec{x}\, \mathrm{e}^t, \tag{2.21}$$

$$X^4 \equiv \cosh t - \frac{1}{2}\vec{x}^2\, \mathrm{e}^t. \tag{2.22}$$

And the resulting metric is

$$ds^2 = -dt^2 + \mathrm{e}^{2t} d\vec{x}^2. \tag{2.23}$$

These coordinates only cover half of de Sitter space: the causal future of the observer at the South Pole. If we fix t, we have slices parametrised by $\vec{x}$. These are everywhere spacelike slices which are nevertheless asymptotic to future null and timelike infinity.

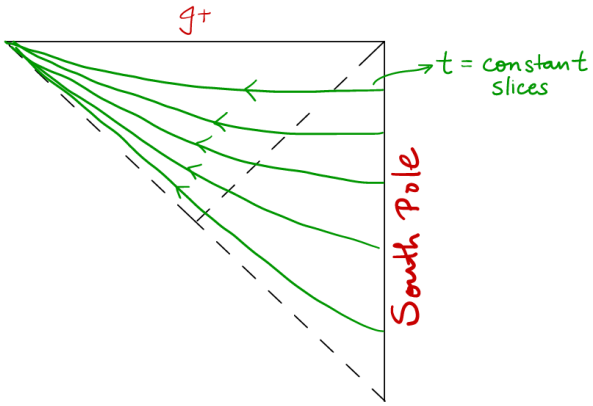


Figure 2.6: Penrose diagram in planar coordinates

As the universe evolves, baryonic matter becomes more and more dilute, but this doesn't happen to dark energy. So eventually, dark energy dominates and this metric needs to be modified. Because there is a singularity along the slice t_{BigBang}, we can ignore everything before it.

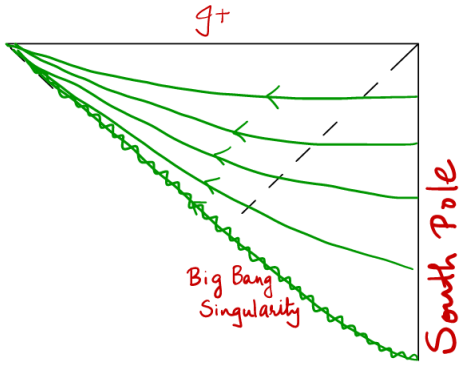


Figure 2.7: Penrose diagram in planar coordinates with a Big Bang singularity

2.3 Anti-de Sitter space

Anti-de Sitter space is a solution to the Einstein equation with negative cosmological constant $\Lambda = -\frac{3}{l^2}$. For simplicity, we tackle the two-dimensional case here. Take a metric with negative constant curvature,

$$ds^2 = \frac{l^2}{y^2}(-dt^2 + dy^2). \tag{2.24}$$

As before, set

$$u \equiv t - y, \qquad v \equiv t + y. \tag{2.25}$$

Then,

$$ds^2 = \frac{-dudv}{4(u-v)^2}. \tag{2.26}$$

Next, take

$$u \equiv \tan \tilde{u}, \qquad v \equiv \tan \tilde{v}. \tag{2.27}$$

And finally, make the change

$$\tilde{u} = \frac{1}{2}(\tau - \sigma), \qquad \tilde{v} \equiv \frac{1}{2}(\tau + \sigma). \tag{2.28}$$

We end up with

$$ds^2 = \frac{1}{\sin^2 \sigma}(-d\tau^2 + d\sigma^2), \tag{2.29}$$

which doesn't look good at $\sigma = 0, \pi$. Let's take some finite point, σ_0, and compute the proper distance from $\sigma = \varepsilon$ to $\sigma = \sigma_0$ in fixed τ. Then,

$$D(\varepsilon, \sigma_0) = \int_{\varepsilon}^{\sigma_0} \frac{1}{\sin \sigma} d\sigma \sim \ln \varepsilon, \tag{2.30}$$

which is infinitely far away. The $\sigma = \pi + \varepsilon$ case is similar.

Next, in Figure 2.8, suppose we start a light ray at $\sigma = \frac{\pi}{2}$, $\tau = 0$. It bounces at the wall $\sigma = 0$ and comes back to $\sigma = \frac{\pi}{2}$. Then the proper time $\tau_p = \pi$. So in anti-de Sitter space, we only need to wait for a proper time to find out things happening infinitely far away. This is an issue in anti-de Sitter space which is not present in Minkowski space. Thus, in anti-de Sitter space, the boundary conditions are extremely important. This also implies that there is no Cauchy surface, since boundary effects can modify the bulk and this prevents us from a deterministic evolution.

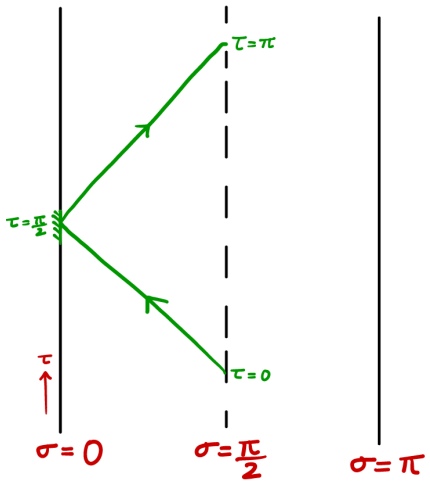


Figure 2.8: Anti-de Sitter space

One reason why we don't call Figure 2.8 the Penrose diagram of anti-de Sitter space is that it is not *geodesically complete*. The original coordinates (t, y) only cover part of the full spacetime known as the Poincaré patch. By

geodesic completion of the spacetime, we mean that we extend the spacetime so that all the null/timelike/spacelike geodesics either travel for infinite proper time or end in a singularity. Anti-de Sitter space, like Minkowski and de Sitter spaces, is geodesically complete.

Further, we note here the two rules for Penrose diagrams.

- Everything must be fit on blackboard;
- Null geodesics are at 45°.

Hence, in order for our anti-de Sitter space Penrose diagram to obey these rules, we have to make a Weyl transformation, which changes the whole structure and gives us Figure 2.9. Notice that the boundary is timelike.

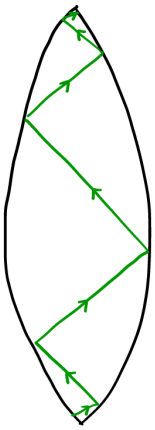


Figure 2.9: Penrose diagram for anti-de Sitter space

At the end of the day, we are not just interested in the spacetimes themselves. We would like to study physics on these spacetimes. An important and necessary step in this direction is to first figure out what set of questions one can sensibly ask in any spacetime.

The answer - or rather, the question - is well understood in Minkowski space. We throw some particles from i^- (massive particles) and $\mathscr{I}^-$ (massless particles) into the bulk of the spacetime. The particles interact and come out at i^+ and $\mathscr{I}^+$. The scattering problem in Minkowski spacetime is then: Given the stuff that goes in, what can come out? In principle, this works only in the absence of gravity. For instance, in the presence of gravity, throwing stuff into the bulk can create black holes, which changes the causal structure significantly.

In anti-de Sitter space, a similar question can be asked. However, in this case, the specification of boundary conditions at $\sigma = 0, \pi$ is extremely important. The choice of boundary conditions is guided by symmetries and typically, there aren't too many choices that respect the symmetries. The question we ask in anti-de Sitter space is if we perturb the boundary conditions at $\tau = 0$ by a bit, then how are the boundary conditions at later times affected?

On the other hand, de Sitter space is much more complicated. We don't know what the right question to ask is due to issues in this spacetime. In particular, due to the big bang singularity, we cannot talk about scattering particles from the past. Since space is contracted back then, one doesn't have a proper definition of particles. To alleviate the aforementioned issue, one might setup the problem at a time *after* the big bang and then ask what happens in the far future. However, since different points on i^+ are out of causal contact, different observers at i^+ cannot communicate with each other. This prevents us from even collecting data in the future. Note that this is strikingly different from Minkowski space since there, all the points i^0, $\mathscr{I}^+$ and i^+ are in causal contact and therefore, data can be collected. These issues concern us since de Sitter space is our universe.

3 Schwarzschild black holes

There is a classic set of black hole solutions in four dimensions.

- *Schwarzschild solution*: $Q = 0, J = 0$, i.e., not charged or rotating black holes,
- *Reissner-Nordström solution*: $Q \neq 0, J = 0$, i.e., charged black holes (which are not astrophysically relevant),
- *Kerr solution*: $Q = 0, J > 0$, i.e., rotating black holes (which are astrophysically relevant);
- *Kerr-Newmann solution*: $Q \neq 0, J > 0$, i.e., rotating and charged black holes.

The reason why charged black holes are irrelevant is because they would attract other charged particles and become neutralised. We will not be discussing black holes in higher dimensions, where they take a more enriching structure.

In this section, we start from the Schwarzschild metric which Schwarzschild wrote down in 1913,

$$ds^2 = -\left(1 - \frac{2m}{r}\right) dt^2 + \frac{dr^2}{1 - \frac{2m}{r}} + r^2 d\Omega_2^2. \tag{3.1}$$

There are two interesting regions.

- $r = 2m$, where the metric becomes
$$ds^2 = 0dt^2 + \frac{dr^2}{0} + 4m^2 d\Omega_2^2. \tag{3.2}$$
Although there seems to be a problem, here we have the Kretschmann scalar
$$K = R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} = \frac{48m^2}{r^6} = \text{finite}, \tag{3.3}$$
and we call this the *horizon*.
- $r = 0$, where the metric becomes
$$ds^2 = \frac{dt^2}{0} + 0dr^2 + 0d\Omega_2^2. \tag{3.4}$$
Here we have
$$K = R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} = \frac{48m^2}{r^6} = \infty. \tag{3.5}$$
We see that in this case the Einstein equations break down. We call this the *singularity*.

3.1 Near horizon limit

To tackle the region $r = 2m$, we take the *near horizon limit*. Introduce

$$\varepsilon \equiv r - 2m. \tag{3.6}$$

Then we have

$$1 - \frac{2m}{r} \simeq \frac{\varepsilon}{2m}. \tag{3.7}$$

Hence,

$$ds^2 = -\frac{\varepsilon}{2m} dt^2 + \frac{2m}{\varepsilon} d\varepsilon^2 + (\not{\varepsilon} + 2m)^2 d\Omega_2^2. \tag{3.8}$$

We see that near the horizon, the geometry is a product metric of a two-dimensional Lorentzian factor and an S^2 with radius $2m$. In fact, by a judicious choice of coordinates, we'll soon see that it is actually a flat two-dimensional metric. Let's take

$$\varepsilon \equiv -u^+ u^-, \tag{3.9}$$

and

$$e^{t/2m} = \frac{u^+}{u^-}, \tag{3.10}$$

where

$$t = 2m \ln u^+ - 2m \ln u^-. \tag{3.11}$$

Then,

$$dt = 2m\frac{du^+}{u^+} - 2m\frac{du^-}{u^-}, \tag{3.12}$$

$$d\varepsilon = -u^+ du^- - u^- du^+. \tag{3.13}$$

Hence,

$$ds^2 = \frac{u^+u^-}{2m} 4m^2 \left(\frac{du^+}{u^+} - \frac{du^-}{u^-}\right)^2 - \frac{2m}{u^+u^-}(u^+du^- + u^-du^+)^2 + 4m^2 d\Omega_2^2 \tag{3.14}$$

$$= -8m du^+ du^- + 4m^2 d\Omega_2^2, \tag{3.15}$$

which is a flat metric in null coordinates. So the singularity in the metric is just an artefact of the choice of coordinates. As we can see in the Figure 3.1, lines of constant ε, i.e., lines of constant r, are where u^+u^- is a constant.

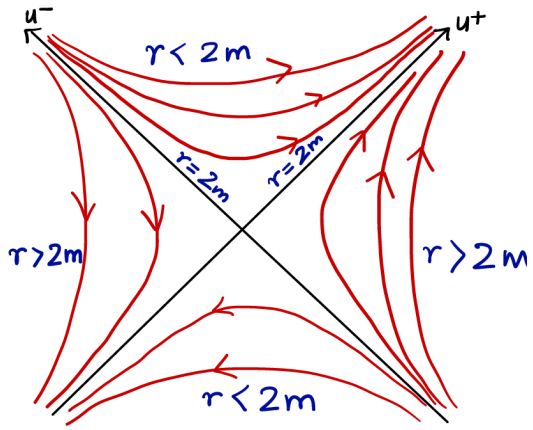


Figure 3.1: Near horizon geometry of the Schwarzschild solution

3.2 Causal structure

Now we move to a more complicated version of this calculation in order to understand the causal structure of the Schwarzschild geometry. Introduce tortoise coordinates $r^* = r^*(r)$ such that the metric can be put into the form

$$ds^2 = -\left(1 - \frac{2m}{r}\right)(dt^2 - dr^{*2}) + r^2 d\Omega_2^2. \tag{3.16}$$

To achieve this, we need

$$\left(1 - \frac{2m}{r}\right) dr^{*2} = \frac{dr^2}{1 - \frac{2m}{r}}, \tag{3.17}$$

which gives

$$r^* = \int \frac{dr}{1 - \frac{2m}{r}} = r + 2m \ln\left(\frac{r}{2m} - 1\right). \tag{3.18}$$

For large r, $r^* \sim r$, and for $r \to 2m$, $r^* \sim -\infty$. This is where the event horizon is located. We can fix this problem by going to Kruskal coordinates, where we take

$$u^* \equiv t - r^* \qquad v^* \equiv t + r^*, \tag{3.19}$$

followed by

$$u \equiv -4m\, \mathrm{e}^{-u^*/4m}, \qquad v \equiv 4m\, \mathrm{e}^{v^*/4m}. \tag{3.20}$$

Note that we will always write our null coordinates so that they increase towards the future. This way we arrive at the Kruskal-Szekeres metric

$$ds^2 = -\frac{2m}{r}\, \mathrm{e}^{-r/2m} du dv + r^2 d\Omega^2. \tag{3.21}$$

Clearly, at the horizon $r = 2m$, everything looks regular. This way, we have extended the near horizon construction to the entire spacetime. Now, the range of u and v is $u, v \in (-\infty, \infty)$, and we have

$$uv = -16m^2\, \mathrm{e}^{r/2m} \left(\frac{r}{2m} - 1 \right), \tag{3.22}$$

which vanishes at $r = 2m$. The singularities corresponding to $r = 0$ happen at $uv = 16m^2$. As we can see from Figure 3.2, the singularities are the hyperbolas, and we have two regions of $r = \infty$.

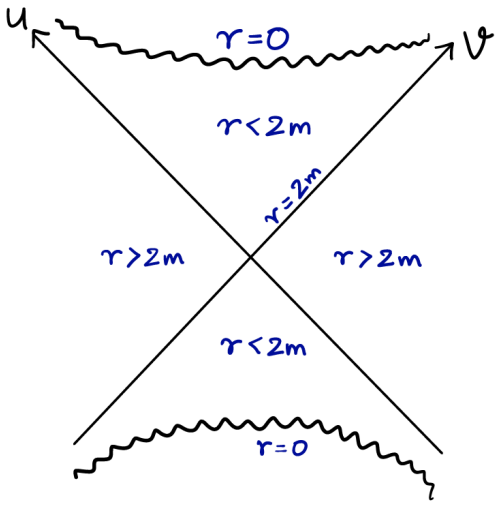


Figure 3.2: Global structure of the Schwarzschild solution

The last thing we want to do is to draw the Penrose diagram. Take

$$u \equiv \tan \tilde{u}, \qquad v \equiv \tan \tilde{v}, \tag{3.23}$$

such that we can conformally rescale the metric

$$ds^2 = \frac{d\tilde{u}}{du}\frac{d\tilde{v}}{dv} ds^2 = -\frac{2m}{r}\, \mathrm{e}^{-r/2m} d\tilde{u} d\tilde{v}. \tag{3.24}$$

Mapping the Schwarzschild geometry onto the plane, we arrive at the Penrose diagram as portrayed in Figure 3.3. Early works on the topic were done by Kruskal, Eddington and Finkelstein in the 50s and the 60s.

Notice that the top and the bottom of the diamond can be erased since they are beyond the singularities. Once an object moves past $r = 2M$, it runs into the singularity, since its trajectory must be timelike. In fact, even light rays that go beyond the horizon run into singularity. There, r becomes smaller and smaller. Right on the horizon, there is a marginally trapped light ray that stays on $r = 2m$.

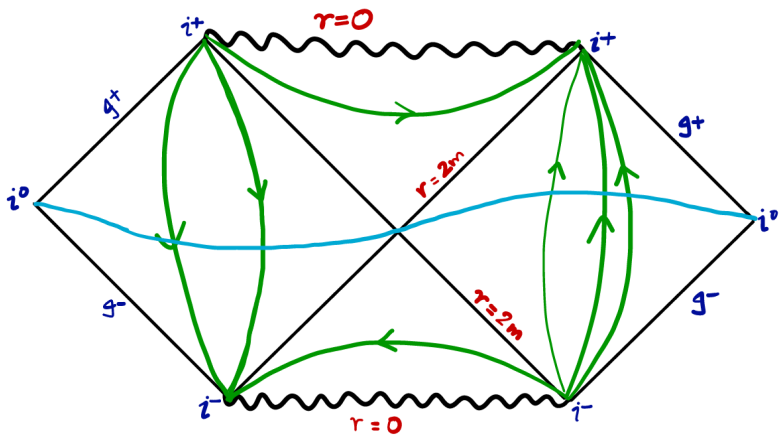


Figure 3.3: Penrose diagram of the Schwarzschild solution

Let's now consider a timelike geodesic $x^a(\tau)$ moving in the (r,t)-plane. It obeys

$$g_{ab}\dot{x}^a\dot{x}^b = -1. \tag{3.25}$$

Hence,

$$-\left(1-\frac{2m}{r}\right)\dot{t}^2 + \frac{\dot{r}^2}{1-\frac{2m}{r}} = -1, \tag{3.26}$$

and we have

$$\ddot{t} + 2\Gamma^t_{rt}\dot{t}\dot{r} = 0, \tag{3.27}$$

where

$$\Gamma^t_{rt} = -\frac{1}{2}\partial_r \ln\left(1-\frac{2m}{r}\right). \tag{3.28}$$

Dividing by $\dot{t}$, we get

$$\frac{\ddot{t}}{\dot{t}} - \partial_t \ln\left(1-\frac{2m}{r}\right) = 0. \tag{3.29}$$

Integrating, we have

$$\ln \dot{t} = \ln\left(1-\frac{2m}{r}\right) + \text{const.} \tag{3.30}$$

Proceeding accordingly, we arrive at

$$\frac{d\tau}{dr} = -\sqrt{\frac{r}{2m}}, \tag{3.31}$$

and the proper time for going from τ_1 to τ_2

$$\tau_2 - \tau_1 = \frac{1}{3}\sqrt{\frac{2}{m}}\left(-r_2^{3/2} - r_1^{3/2}\right). \tag{3.32}$$

So we can reach a black hole in a finite amount of proper time. Further, if we cross the horizon, the time to get to $r=0$, i.e., the singularity, is finite as well. An observer on the timelike infinity never sees an object reaching the black hole. Instead, signals are redshifted. This is where the notion of time becomes confusing! Notice that there are two regions where r becomes infinite.

If we take a spacelike surface as shown in Figure 3.3, going from one end to the other, r decreases to $2m$ then increases again. This surface has a wormhole geometry, as portrayed in Figure 3.4. This is what's called the *Einstein-Rosen bridge*. The regions are completely disconnected. The future horizon is where the last light ray can escape to infinity, and it's the surface behind which nothing can escape. This leads to the definition of a black hole: a region in space from which light rays cannot escape out to null infinity.

Finally, note that when $m<0$, the Penrose diagram is as in Figure 3.5.

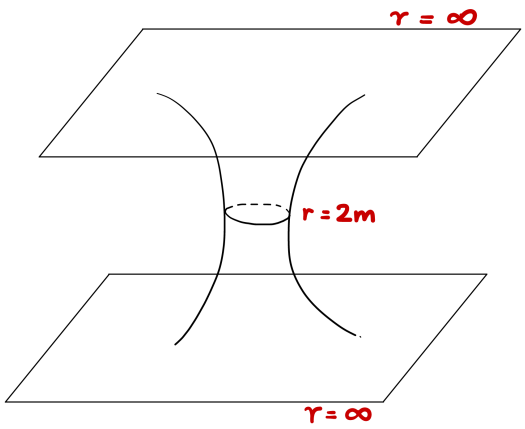


Figure 3.4: Wormhole geometry of Schwarzschild spacetime

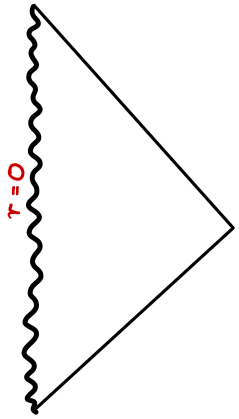


Figure 3.5: Penrose diagram for the $m < 0$ case of Schwarzschild solution

3.3 Vaidya metric

The Schwarzschild solution is an example of an eternal black hole. In our universe, black holes are formed by gravitational collapse, and they do not eternally exist as far as we know. We now discuss such black holes. To simplify our computation, we will assume spherically symmetric gravitational collapse. The most general form of a spherically symmetric metric can be written as

$$ds^2 = g_{ab}dx^a dx^b + \mathrm{e}^{-2\phi}d\Omega^2, \tag{3.33}$$

where $x^a = (r, t)$, $g_{ab} = g_{ab}(r, t)$ is a two-dimensional metric in the (r, t)-plane, and $\phi = \phi(r, t)$. And S^2 with fixed r and t have area $4\pi\,\mathrm{e}^{-2\phi}$. The components of the four-dimensional Einstein tensor turn out to be

$${}^4G_{ab} = 2\nabla_a\nabla_b\phi - 2\nabla_a\phi\nabla_b\phi + 3g_{ab}(\nabla\phi)^2 - 2g_{ab}\Box\phi - \mathrm{e}^{2\phi}g_{ab}, \tag{3.34}$$

$${}^4G_{\theta\theta} = \sin^2\theta\,{}^4G_{\phi\phi} = \left[(\nabla\phi)^2 - \Box\phi - \frac{1}{2}R\right]\mathrm{e}^{-2\phi}. \tag{3.35}$$

Here let's note that the Schwarzschild metric can be put into the form

$$ds_4^2 = -\left(1 - \frac{2m}{r}\right)dv^2 + 2drdv + r^2d\Omega^2, \tag{3.36}$$

which is in the form of the ansatz we now make (leaving out the angular part),

$$ds_2^2 = 2drdv - h(r, v)dv^2. \tag{3.37}$$

We want to describe a solution where we have matter collapsing in from infinity at the speed of light along null curves. This coordinate system adapts to this description. We also make an ansatz regarding matter. We assume that all components of $T_{\mu\nu}$ vanish except for

$$T_{vv} = \frac{\varepsilon(v)}{4\pi} \, \mathrm{e}^{-2\phi}. \tag{3.38}$$

With our form of metric, the only upper component of non-vanishing $T^{\mu\nu}$ is T^{rr}, and hence we have

$$\nabla_\mu T^{\mu\nu} = \frac{1}{\sqrt{g}} \partial_\mu \left(\sqrt{g} T^{\mu\nu} \right) = 0. \tag{3.39}$$

Now, to solve the Einstein equations

$$G_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{3.40}$$

let's first take

$$r \equiv \mathrm{e}^{-\phi}. \tag{3.41}$$

After some work, we find

$$\mathrm{e}^{2\phi} \, {}^4G_{\theta\theta} = \frac{1}{2r} \partial_r^2 (rh). \tag{3.42}$$

For large r, we want everything to be flat, i.e., $h = 1$. So we can take

$$h = 1 - \frac{2m(v)}{r}, \tag{3.43}$$

which correspond to what we had for the form of the Schwarzschild metric above. However, here we allow the mass to depend on v. Substituting the above h into G_{vv}, we also have that

$$G_{vv} = \frac{2}{r^2} \partial_v m. \tag{3.44}$$

Hence,

$$T_{vv} = \frac{\varepsilon(v)}{4\pi r^2} = \frac{1}{8\pi} G_{vv} = \frac{1}{8\pi} \frac{2}{r^2} \partial_v m, \tag{3.45}$$

and we arrive at

$$\partial_v m = \varepsilon(v). \tag{3.46}$$

Consider the simple case where

$$\varepsilon(v) = \delta(v - v_0)\mu. \tag{3.47}$$

The corresponding solution is known as the Vaidya shockwave. Then,

$$m = \theta(v - v_0)\mu. \tag{3.48}$$

- For $v < v_0$, we have a flat space with metric

$$ds^2 = 2dr dv - dv^2 + r^2 d\Omega^2. \tag{3.49}$$

- For $v > v_0$, we have the Schwarzschild metric

$$ds^2 = 2dr dv - \left(1 - \frac{2m}{r} \right) dv^2 + r^2 d\Omega^2. \tag{3.50}$$

Now, one of the biggest questions in modern physics that we want to address is "What's inside a black hole?" Or rather, "Is there an inside to a black hole?" What's funny about the region inside the horizon $r = 2M$ is that no information can be communicated to the outside, and we cannot do any experiment to learn about it. In fact, Einstein himself thought that it didn't make much sense to talk about it. This is related to the *Cosmic censorship conjecture* that we will discuss later.

Nevertheless, we can in fact gain insight into the inside of a black hole by looking at the Vaidya geometry. So what exactly is it? We glue the Minkowski Penrose diagram and the Schwarzschild Penrose diagram accordingly to obtain the Vaidya shockwave geometry as in Figure 3.6.

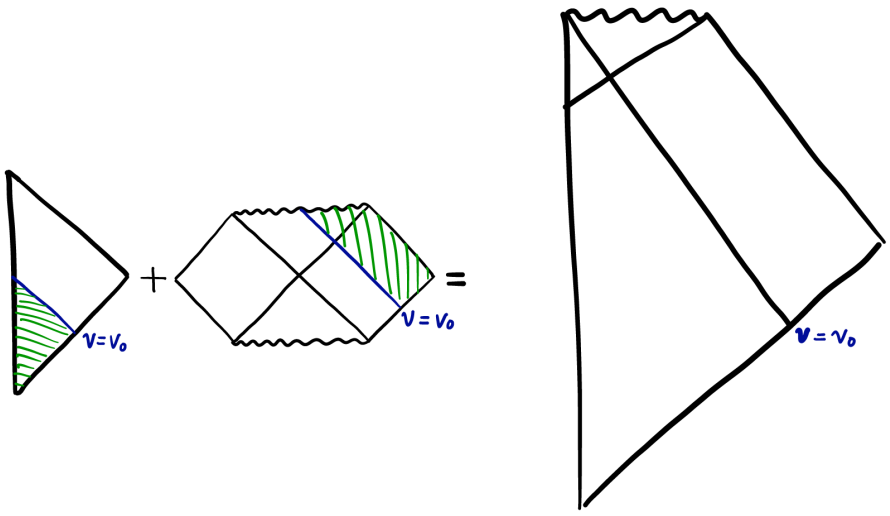


Figure 3.6: Penrose diagram of the Vaidya shockwave

In Figure 3.7, note that as the shockwave crosses P, it crosses the Schwarzschild radius. Once this happens, any outgoing lightray emitted from its surface (green) cannot reach infinity and instead reaches the singularity at $r = 0$ (blue). Note the interesting structure of lines of constant r. $r = 0$ transforms from a timelike to a spacelike line at O. The Vaidya structure is a good picture to represent the causal structure of black holes found in nature. Comparing to the Schwarzschild geometry, we have lost a lot of structure, in particular, the "second universe" including the white hole and the mirrored spacetime on the other side of the horizon.

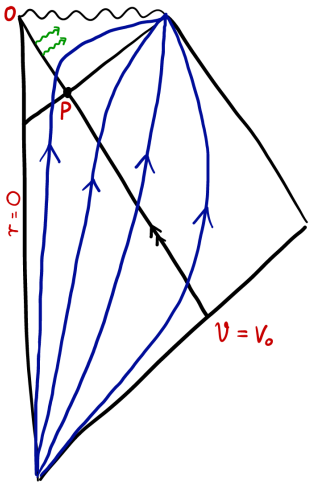


Figure 3.7: Penrose diagram of a Vaidya shockwave

Now, note that the Einstein equations are time-reversal invariant. This means that we can also have a solution that involves a white hole. The Schwarzschild solution contains both black and white holes. However, white holes don't seem to be around, unless we consider the Big Bang as a white hole. Given that we don't see white holes, physicists have decided to throw out the time-reversal solutions, i.e., we choose an arrow of time.

Finally, the shockwave doesn't have smooth initial data so we shouldn't be surprised that we end up with the construction of a black hole. In principle though, we can construct black holes with gravitational waves as well. However, there is no spherically symmetric gravitational wave by Birkhoff's theorem. Hence, we'll have to consider

more complicated solutions than the radially symmetric one considered above. We will not consider these solutions here, since the relatively simple example of the Vaidya spacetime already captures most of the qualitative features that we are interested in.

4 Reissner-Nordström black holes

Reissner-Nordström black holes are charged black holes and solutions to

$$G_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{4.1}$$

where

$$T_{\mu\nu} = \frac{1}{4\pi} g^{\alpha\beta} F_{\mu\alpha} F_{\nu\beta} - \frac{1}{16\pi} g_{\mu\nu} F_{\alpha\beta} F^{\alpha\beta}. \tag{4.2}$$

The most general static spherically symmetric solution to these equations has a form similar to the Schwarzschild solution, namely,

$$ds^2 = -\left(1 - \frac{2m}{r} + \frac{Q^2}{r^2}\right) dt^2 + \frac{dr^2}{1 - \frac{2m}{r} + \frac{Q^2}{r^2}} + r^2 d\Omega^2, \tag{4.3}$$

with the radial electric field

$$F_{rt} = \frac{Q}{2r^2}. \tag{4.4}$$

The equations of electromagnetism have a duality symmetry under which we can define a new field strength,

$$\tilde{F}_{\mu\nu} = \cos\theta F_{\mu\nu} + \frac{1}{2} \varepsilon_{\mu\nu\alpha\beta} F^{\alpha\beta} \sin\theta. \tag{4.5}$$

This transformation rotates electric fields and magnetic fields, and the stress energy tensor is invariant under this rotation. More over, if F is curl and divergence free, then so is $\tilde{F}$. So we can take our solution above that has purely electric charge, do the duality transformation, and obtain a solution with magnetic charge with the same geometry. Hence, here we only discuss a black hole with electric charge.

We can write

$$1 - \frac{2m}{r} + \frac{Q^2}{r^2} = \frac{1}{r^2}(r - r_+)(r - r_-), \tag{4.6}$$

where we have set the inner and outer horizons

$$r_\pm \equiv m \pm \sqrt{m^2 - Q^2}. \tag{4.7}$$

So we have three cases.

4.1 $m^2 < Q^2$: a naked singularity

In this case we have two imaginary roots, giving a naked singularity at $r = 0$ and no solution to the Einstein equations. Note that an electron has $m^2 \ll Q^2$. Does this mean that we are throwing out the electron? The answer is no since the Schwarzschild radius of the electron is far smaller than its Compton wavelength. So we cannot view the electron as a point particle when we go to short distances. Quantum mechanics becomes far more important than gravity there. So for classical objects, we throw out anything with $m^2 < Q^2$.

4.2 $m^2 = Q^2$: the extremal case

Here we have a degenerate solution, known as "extreme Reissner-Nordström", where the word *extreme* refers to the fact that this black hole is like a black body with zero temperature. This is the most interesting case here and has no analogue to the Schwarzschild case. We now discuss this case, and start by writing

$$ds^2 = -\left(1 - \frac{Q}{r}\right)^2 dt^2 + \frac{dr^2}{\left(1 - \frac{Q}{r}\right)^2} + r^2 d\Omega^2. \tag{4.8}$$

Recall that in the Schwarzschild case, as we pass the event horizon, time and space interchange because the coefficient in front of dt^2 changes sign. This is not the case here, since when r hits Q, the coefficient of dt^2 becomes zero but doesn't flip sign. This leads to interesting phenomena that we will discuss in detail. As in the case with Schwarzschild, let's do a near horizon analysis of the extreme Reissner-Nordström case. Define a new coordinate

$$\rho \equiv \frac{r - Q}{\varepsilon}, \qquad \tau \equiv \frac{\varepsilon t}{Q^2}, \tag{4.9}$$

where ε is a parameter of the coordinate transformation. The metric becomes

$$ds^2 = -\left(\frac{\varepsilon\rho}{\varepsilon\rho + Q}\right)^2 \left(\frac{Q^2 d\tau}{\varepsilon}\right)^2 + \left(\frac{\varepsilon\rho}{\varepsilon\rho + Q}\right)^{-2} (\varepsilon d\rho)^2 + (\varepsilon\rho + Q)^2 d\Omega_2^2. \tag{4.10}$$

And taking $\varepsilon \to 0$ forces r to be near Q. So zooming into the region $r = Q$, the metric takes the form

$$ds^2 = -Q^2\rho^2 d\tau^2 + Q^2\rho^2 d\rho^2 + Q^2 d\Omega_2^2. \tag{4.11}$$

Defining $\rho \equiv 1/y$ gives us

$$ds^2 = Q^2 \frac{-d\tau^2 + dy^2}{y^2} + Q^2 d\Omega_2^2. \tag{4.12}$$

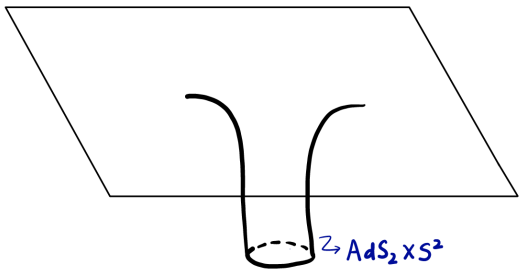


Figure 4.1: The $AdS_2 \times S^2$ structure of the Robinson-Bertotti universe

This is the geometry of $(\mathrm{AdS}_2)_{\ell=Q} \times (S^2)_{\text{radius}=Q}$, as in Figure 4.1. Thus, the near horizon geometry is $\mathrm{AdS}_2 \times S^2$. This is called the Robinson-Bertotti universe. It is a geodesically complete solution of the Einstein-Maxwell equations.

Recall our construction in the case of AdS_2. With metric taking the form

$$ds^2 = -\frac{dt^2 + d\sigma^2}{\sin^2\sigma}, \tag{4.13}$$

the Penrose diagram in the (t, σ) coordinates are portrayed in Figure 4.2. Note that since the range of y is $(0, \infty)$, $\rho = 1/y$ also has range $(0, \infty)$. Hence, this patch of the near horizon geometry covers the region outside the horizon. In the full two-dimensional anti-de Sitter space, the line $y = \infty$ has no special significance.

Let's now get back to the full solution to patch together the Penrose diagram. When $r > Q$, the region looks like Minkowski space as shown in Figure 4.3.

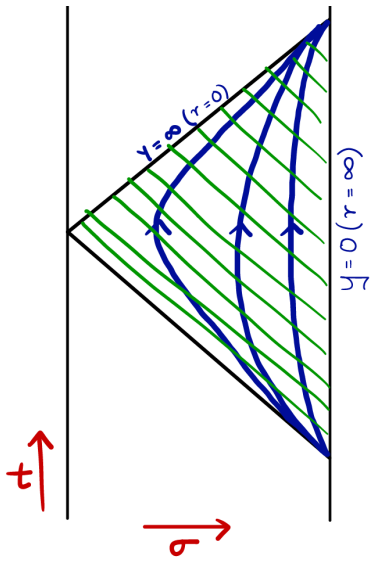


Figure 4.2: Penrose diagram of AdS_2

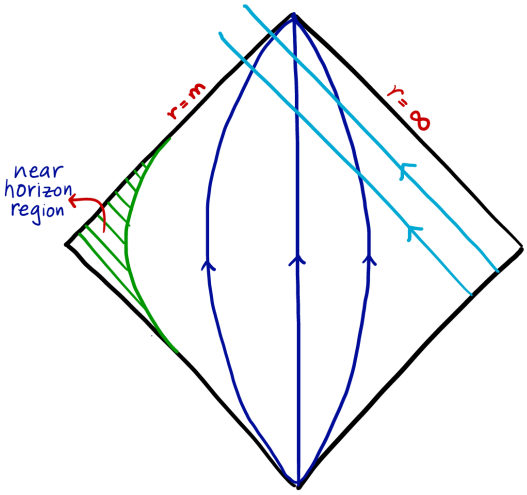


Figure 4.3: The $r > Q$ region of Reissner-Nordström black holes

Note that the reason why we are drawing $r = Q$ as a null line is that the coefficient of dt^2 vanishes. We cannot just use this region since it is not geodesically complete. The shaded region is the same as the near horizon region in Figure 4.3. As discussed previously, timelike and null directions do not take infinite time to reach the horizon, whereas spacelike curves do. Thus, we are compelled to talk about the region inside. To do so, we define new coordinates u, v such that

$$dv \equiv dt + \frac{dr}{(1 - \frac{Q}{r})^2}, \tag{4.14}$$

which leads to

$$ds^2 = -\left(1 - \frac{Q}{r}\right)^2 dv^2 + 2dvdr + r^2 d\Omega^2. \tag{4.15}$$

We see here that at $r = Q$, the metric is not degenerate, i.e., we can go beyond $r = Q$. The $r < Q$ case is shown in Figure 4.4. Attaching the two regions doesn't geodesically complete the spacetime: As shown in Figure 4.5, we need to add an infinite number of such patches to get a geodesically complete spacetime.

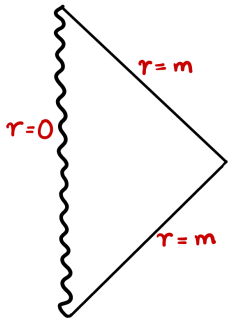
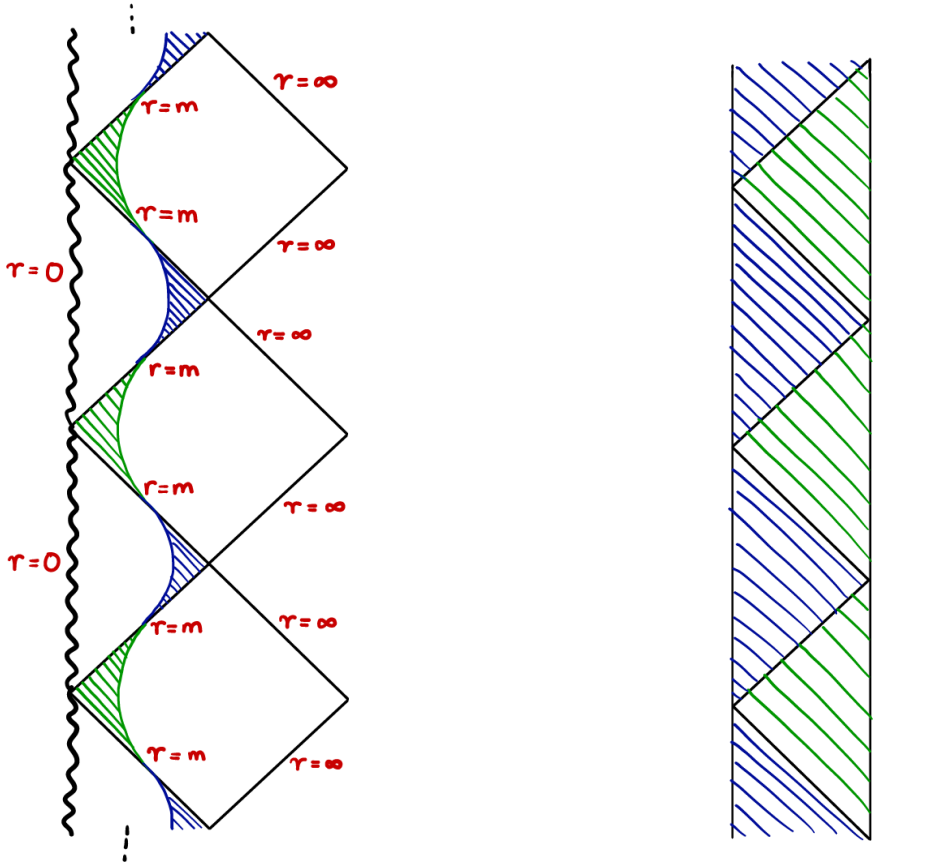


Figure 4.4: Penrose diagram of AdS_2



(a) Penrose diagram

(b) Near horizon geometry

Figure 4.5: $m^2 = Q^2$ Reissner-Nordström black holes

Amazingly, Carter found a single coordinate system that covers the entire geometry. Set

$$x^+ \equiv \tan^{-1} v, \quad x^- \equiv \cot^{-1} u. \tag{4.16}$$

Then,

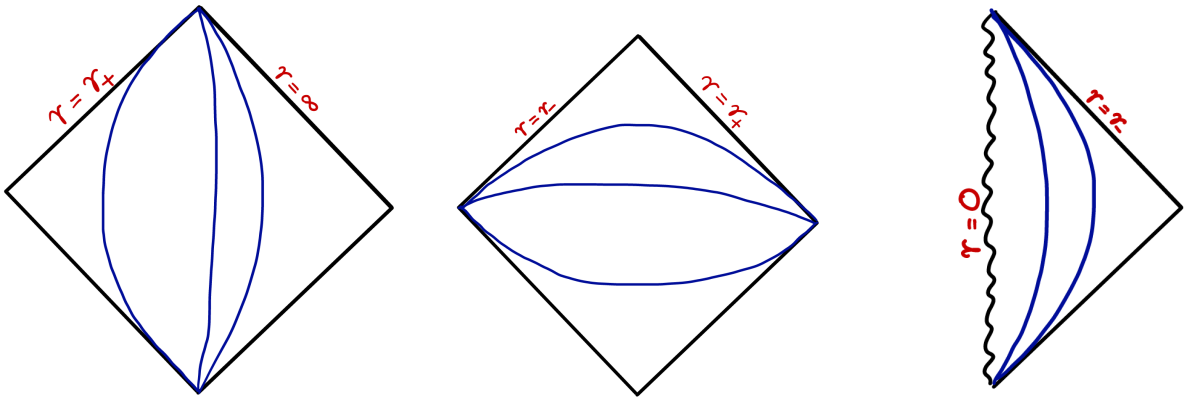
$$ds^2 = -\left(1 - \frac{Q}{r}\right)^2 \sec^2 x^+ \mathrm{cosec}^2 x^- dx^+ dx^- + r^2 d\Omega^2. \tag{4.17}$$

The infinite size of the Penrose diagram implies that the $r = Q$ hypersurface of an extremal Reissner-Nordström black hole is a Cauchy horizon. This is a horizon that lies at the lightlike boundary of the domain of validity of a Cauchy surface. The surface is an unstable horizon, in the sense that any tiny perturbation of the initial data on the Cauchy surface grows without bound at this horizon. This is because our entire infinite-sized universe lies in

the past of the horizon while any data flowing in reaches it in finite time. So from the point of view of an external observer, all the data that has existed throughout the infinite past of the universe reaches the horizon as the proper time of the observer approaches infinity. Therefore, there is a tendency for data to diverge at this surface.

4.3 $m^2 > Q^2$: a regular black hole with two horizons

Here we have two real roots, $r_+ > r_- > 0$. First notice that when $m^2 \gg Q^2$, Reissner-Nordström basically looks like Schwarzschild black hole with a little bit of electromagnetic field coming out of it. Now, we have three different kinds of region. Following the same procedure as in the extremal case, we can construct the Penrose diagram separately in $r > r_+$, $r_+ > r > r_-$ and $r_- > r > 0$ regions and sew them together. The fact that the regions are joined together smoothly can be shown by constructing coordinate systems that are well-defined across each horizon. The Penrose diagram in each region is portrayed in Figure 4.6.



(a) $r > r_+$ (b) $r_+ > r > r_-$ (c) $r_- > r > 0$

Figure 4.6: Pieces of $m^2 > Q^2$ Reissner-Nordström black hole Penrose diagrams

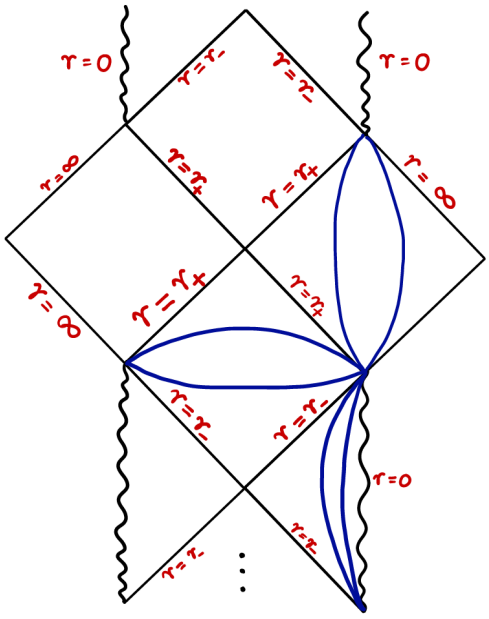


Figure 4.7: Penrose diagram of a $m^2 > Q^2$ Reissner-Nordström black hole.

Joining the regions together, we arrive at Figure 4.7. Each horizon is smooth. For instance, near $r = r_+$, we can setup coordinates

$$r_* \equiv r + \frac{r_+^2}{r_+ - r_-} \ln|r_+ - r| - \frac{r_-^2}{r_+ - r_-} \ln|r_- - r|. \tag{4.18}$$

Then, $e^{2\kappa r_*}$ is well-defined in the entire region of $r > r_-$ and therefore indicates that $r = r_+$ is smooth. Here,

$$\kappa = \frac{r_+ - r_-}{2r_+^2} \tag{4.19}$$

is the *surface gravity* of the black hole which we will discuss later. Other coordinates covering the other patches can be similarly constructed.

Finally, note that charged black holes have an electrostatic potential defined by

$$\Phi = -\chi^\mu A_\mu|_{r=r_+}, \tag{4.20}$$

where χ is the null vector that generates the event horizon and A_μ is a gauge field that vanishes at infinity. For the Reissner-Nordström black hole, $\chi = \partial_t$. Hence,

$$\Phi = \frac{Q}{r_+}. \tag{4.21}$$

5 Kerr and Kerr-Newman black holes

5.1 Kerr metric

Now we discuss the most important metric for the purposes of cosmology, the Kerr metric. It was discovered in 1963 by Roy Kerr and it was one the greatest discoveries of the 20th century. It's a shame that no Nobel prize was given out for this discovery. The Kerr metric is

$$ds^2 = -\frac{\Delta}{\rho^2}\left(dt - a\sin^2\theta d\phi\right)^2 + \frac{\sin^2\theta}{\rho^2}\left[\left(r^2 + a^2\right)d\phi - adt\right]^2 + \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2, \tag{5.1}$$

or,

$$ds^2 = -\left(1 - \frac{2Mr}{\rho^2}\right)dt^2 + \left[r^2 + a^2 + \frac{2a^2Mr}{\rho^2}\sin^2\theta\right]\sin^2\theta d\phi^2 - \frac{4aMr\sin^2\theta}{\rho^2}dtd\phi + \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2, \tag{5.2}$$

where

$$\Delta = r^2 - 2Mr + a^2, \qquad \rho^2 = r^2 + a^2\cos^2\theta. \tag{5.3}$$

This geometry has two obvious symmetries: time translations generated by ∂_t, and rotations around ϕ generated by ∂_ϕ. They have corresponding conserved quantities mass M and angular momentum J.

In addition to the Killing symmetries, the Kerr solution also has discrete symmetries. Schwarzschild and Reissner-Nordström black holes have two discrete symmetries $t \to -t$ and $\phi \to -\phi$. The Kerr solution on the other hand doesn't allow for the two independent symmetries. Rather, it is invariant under $(t, \phi) \to (-t, -\phi)$. This is consistent with the understanding that the Kerr solution represents a rotating black hole, so that time inversion also inverts the angular velocity.

This geometry also has an inner and an outer horizon, given by $g^{rr} = 0$, which implies $\Delta = 0$. The horizons are located at

$$r_\pm = M \pm \sqrt{M^2 - a^2}. \tag{5.4}$$

Naked singularities do not arise if $a \leq M$. This is known as the *cosmic bound*. Note that in the limit where $a \to 0$, this simplifies to the Schwarzschild black hole. In fact, this solution has angular momentum

$$J = aM, \tag{5.5}$$

so the parameter a is the angular momentum to mass ratio of the black hole. At the extremal value $a = M$, we have $r_+ = r_-$ and $J = M^2$. $r_\pm$ are smooth horizons at which the spacetime is smooth. This can be checked by constructing appropriate coordinates.

5.2 Singularity structure

To see the singular points of the spacetime, we compute the Kretschmann scalar

$$K = R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} = \frac{48M^2r^6}{\rho^2}\left[1 - 15\left(\frac{a\cos\theta}{r}\right)^2 + 15\left(\frac{a\cos\theta}{r}\right)^4 - \left(\frac{a\cos\theta}{r}\right)^6\right]. \tag{5.6}$$

This blows up at $\rho^2 = r^2 + a^2\cos^2\theta = 0$, i.e., at $r = 0,\ \theta = \pi/2$. To see that we obtain the geometric picture of a ring singularity, note that as $M \to 0$, we have the metric

$$ds^2 = -dt^2 + dr^2 + \left(r^2 + a^2\right)d\Omega_2^2 - a^2\sin^2\theta\left(d\theta^2 + \frac{dr^2}{r^2 + a^2}\right). \tag{5.7}$$

This is the metric of flat spacetime as can be checked by computing the Riemann tensor. However, (r, θ, ϕ) are not the standard spherical coordinates on this spacetime. Rather, the standard Cartesian coordinates are related to these via

$$\begin{aligned} x^1 &= \sqrt{r^2 + a^2}\sin\theta\cos\phi, \\ x^2 &= \sqrt{r^2 + a^2}\sin\theta\sin\phi, \\ x^3 &= r\cos\theta. \end{aligned} \tag{5.8}$$

In terms of the Cartesian coordinates, the singularity at $r = 0$, $\theta = \pi/2$ is located at

$$x^1 = a\cos\phi, \qquad x^2 = a\sin\phi, \qquad x^3 = 0. \tag{5.9}$$

Hence, the singularity is a ring of radius a on the (x^1, x^2) plane. It is therefore known as a ring singularity. The singularity can only be approached along $\theta = \pi/2$. Therefore, in a geodesically complete spacetime, we must extend r to $-\infty$.

To study this spacetime, we note that there are three regions of importance with constant (t, θ, ϕ) and varying r curves as follows.

- $r_+ < r < \infty$: spacelike curves;
- $r_- < r < r_+$: timelike curves;
- $-\infty < r < r_-$: spacelike curves.

The geodesically extended Kerr spacetime has *closed timelike curves* located at negative r. These are problematic and may violate causality. For instance, consider the curve $r = \varepsilon < 0$, $\theta = \pi/2$ and $t =$ constant. The metric in this region is

$$ds^2 = -\frac{a^2}{\varepsilon^2}\Delta d\phi^2 + \frac{(\varepsilon^2 + a^2)^2}{\varepsilon^2}d\phi^2 = \frac{2Ma^2}{\varepsilon}d\phi^2, \tag{5.10}$$

where ϕ is a compact coordinate. Therefore, at $r < 0$, we have closed timelike curves. However, in any black hole formed via gravitational collapse, the $r < 0$ region doesn't exist. In general, the formation of closed timelike curves from smooth initial configurations is not allowed by Hawking's *chronology protection conjecture*. The Penrose diagram of the Kerr solution is portrayed in Figure 5.1.

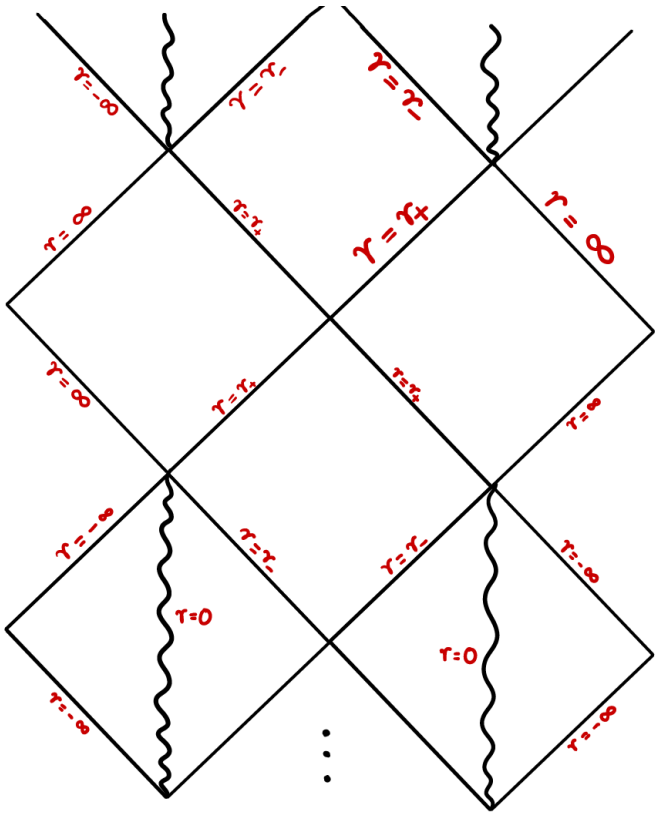


Figure 5.1: Penrose diagram of the Kerr solution

5.3 Ergosphere

The Kerr solution has a special region outside the horizon known as the *ergosphere*. This is a region in which a particle can have negative energy and it's what allows for extremely interesting phenomena in Kerr. To understand this region, we consider curves of constant r, θ, ϕ and varying t. With $\rho^2 = r^2 + a^2 \cos^2\theta$, the line element along this curve are

$$ds^2 = -\left(1 - \frac{2Mr}{\rho^2}\right) dt^2. \tag{5.11}$$

These curves are timelike if

$$1 - \frac{2Mr}{\rho^2} > 0, \tag{5.12}$$

i.e., if

$$r > r_s^+, \qquad r < r_s^-, \tag{5.13}$$

where

$$r_s^\pm = M \pm \sqrt{M^2 - a^2 \cos^2\theta}. \tag{5.14}$$

Recall that $r_\pm = M \pm \sqrt{M^2 - a^2}$. Since $r_s^- < r_+$, it is not accessible to an outside observer. However, $r_s^+ \geq r_+$ and it has the structure depicted in Figure 5.2.

The hypersurface of constant $r = r_s^+$ is known as the ergosurface or the stationary limit surface. The region $r_+ < r < r_s^+$ is known as the ergosphere. Inside the ergosphere, curves of constant r, θ, ϕ and varying t are spacelike. Further, all geodesics inside the ergosphere must co-rotate with the black hole. To see this, consider a geodesic at fixed r and θ. The line element along this geodesic is

$$ds^2 = -\frac{\Delta}{\rho^2}\left(dt - a\sin^2\theta d\phi\right)^2 + \frac{\sin^2\theta}{\rho^2}\left[\left(r^2 + a^2\right) d\phi - a dt\right]^2. \tag{5.15}$$

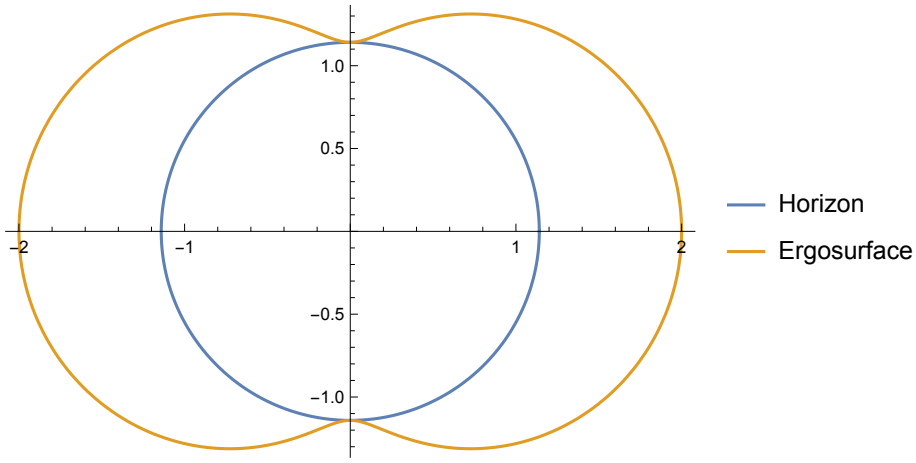


Figure 5.2: r_s^+/M and r_+/M at $a = 0.99M$.

We set $\theta \equiv \pi/2$ for simplicity. Then, the geodesic is null when

$$\frac{d\phi}{dt} = \frac{a \pm \sqrt{\Delta}}{(r^2 + a^2) \pm a\sqrt{\Delta}}. \tag{5.16}$$

Note the following four cases.

- At $r \gg r_s^+$, i.e., outside the ergosphere, null geodesics satisfy

$$\frac{d\phi}{dt} = \pm\frac{1}{r}, \tag{5.17}$$

  which is the expected result in Minkowski space, corresponding to lightlike trajectories orbiting the black hole clockwise or counter-clockwise.

- At $r = r_s^+ = 2M$, i.e., on the ergosurface, we have $\Delta = a^2$ and null geodesics are located at

$$\frac{d\phi}{dt} = 0 \quad \text{or} \quad \frac{d\phi}{dt} = \frac{2a}{(r_s^+)^2 + 2a^2}. \tag{5.18}$$

  So as one nears the ergosphere, one is dragged along with the rotation of the black hole. Here all timelike geodesics co-rotate with the black hole. Only null rays can remain stationary on the ergosurface.

- At $r_+ < r < r_s^+$, i.e., inside the ergosphere, it becomes impossible to maintain a fixed angle ϕ with respect to infinity. In other words, one is forced to co-rotate with the black hole. It is a law of physics in the ergosphere that you must spin along with the black hole. This is known as *frame-dragging*.

- At $r = r_+$, i.e., at $\Delta = 0$,

$$\frac{d\phi}{dt} = \frac{a}{r_+^2 + a^2}. \tag{5.19}$$

  The lightcones coincide and we find no rotating timelike geodesics but a single null geodesic. Thus, timelike geodesics cannot move at constant $r = r_+$ and all null geodesics must orbit at a fixed speed, independent of θ. This speed is known as the angular velocity of the horizon, Ω_H, and we have

$$\Omega_H = \frac{a}{r_+^2 + a^2} = \frac{J}{2M}\left[M^2 + \sqrt{M^4 - J^2}\right]^{-1}. \tag{5.20}$$

Another interesting property of the ergosphere is that particles inside may have negative energy, which is due to the fact that curves that are timelike (such as the r, θ, ϕ constant and t varying curves) outside the ergosphere can be spacelike inside. Consider a timelike Killing vector at infinity, K, such that

$$\nabla_\mu K_\nu + \nabla_\nu K_\mu = 0. \tag{5.21}$$

Energy in general relativity is defined with respect to K as

$$E = -K_\mu \partial_t x^\mu, \tag{5.22}$$

which is a conserved quantity. If K is globally timelike, then E is always positive, since the inner product of two timelike vectors is always negative. However, inside the ergosphere, energy can be negative since ∂_t is a spacelike vector.

Note that such a region exists even for the Schwarzschild and Reissner-Nordström black holes discussed earlier. However, in those cases, the region where ∂_t is spacelike is inside the horizon $r_- < r < r_+$ and is not accessible to an observer at infinity. The interesting fact about Kerr black holes is that the ergosphere is located outside the horizon. This fact can be used to extract energy from the black hole. We will describe this process in detail later.

5.4 Near horizon extremal Kerr

The cosmic bound is saturated by $a = M$. There, we have a black hole known as the extremal Kerr solution. The near horizon geometry of extremal Kerr is extremely interesting and was first discussed by Bardeen and Horowitz. To take the near horizon limit, we define coordinates

$$\hat{t} \equiv \frac{\lambda t}{2M}, \qquad \hat{r} \equiv \frac{r - M}{\lambda M}, \qquad \hat{\phi} \equiv \phi - \frac{t}{2M}. \tag{5.23}$$

In the limit $\lambda \to 0$, we reach the near horizon extremal Kerr region. It is described by the metric

$$ds^2 = 2M^2 \Gamma(\theta) \left[-\hat{r}^2 d\hat{t}^2 + \frac{d\hat{r}^2}{\hat{r}^2} + d\theta^2 + \Lambda^2(\theta) \left(d\hat{\phi} + \hat{r} d\hat{t} \right)^2 \right], \tag{5.24}$$

where

$$\Gamma(\theta) = \frac{1 + \cos^2\theta}{2}, \qquad \Lambda(\theta) = \frac{2 \sin\theta}{1 + \cos^2\theta}. \tag{5.25}$$

The (t, r) plane has the metric of AdS_2. Unlike the simpler case of the Reissner-Nordström black hole, the near horizon Kerr metric is not a simple product metric, but a fibration. The Kerr solution has a $U(1) \times U(1)$ symmetry, and in the extremal limit, it has a ribbon region in which we can take a scaling limit where we have the extra symmetries of $SO(2, 1)$, which is the conformal group. We show in Figure 5.3 how the near horizon extremal Kerr region fits in with the full Kerr geometry. In particular, we see how the near horizon geometry maps onto that of AdS_2.

5.5 Penrose process

The fact that energy of a particle can be negative inside the ergosphere can be used to extract energy from a rotating black hole, and this is known as the *Penrose process*. This was first described by Penrose and Christodoulou. Consider a process in which a rocket with Matthew Mcconaughey (MM) starts outside the ergosphere and travels along a geodesic into the ergosphere with initial four-momentum p_i^μ. The scenario is portrayed in Figure 5.4.

We have by the conservation of momentum,

$$p_i^\mu = p_f^\mu + p_{MM}^\mu. \tag{5.26}$$

The total energy of the spacetime, assuming there is nothing else, is conserved and is given by

$$E_{tot} = E_i + E_{BH}. \tag{5.27}$$

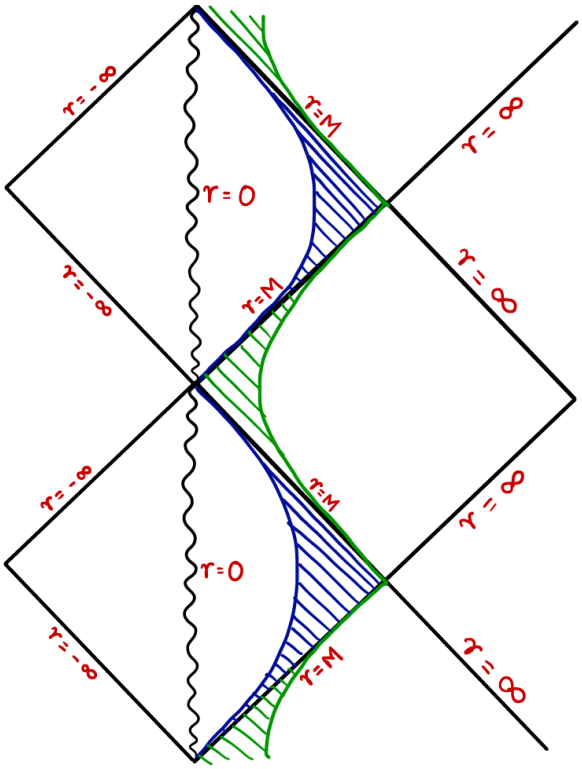


Figure 5.3: Penrose diagram of Extremal Kerr

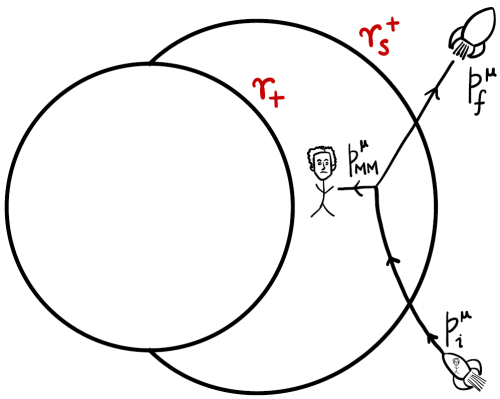


Figure 5.4: Travelling through the ergosphere

Once inside the ergosphere, MM ejects himself along a geodesic. At the instant of ejection, energy of the rocket+MM system is conserved and we have

$$E_i = E_f + E_{MM}, \tag{5.28}$$

where E_f is the final energy of the rocket. If MM chooses his geodesic appropriately, as he does in the movie Interstellar, then $E_{MM} < 0$. In this case, we have $E_f > E_i$ and the rocket has more energy than what it started with. Further, since $E_f > 0$, the rocket can escape to infinity. Since $E_{MM} < 0$, MM cannot exit the ergosphere, unless he accelerates and increases his energy to a positive number. He must either stay in the ergosphere forever or eventually fall into the black hole. If he falls into the black hole, the total change in the black hole energy is

$$\Delta E_{BH} = E_{MM}. \tag{5.29}$$

Of course, throughout this process, the total energy E_{tot} is conserved. In addition to energy, angular momentum is also conserved in this process and we have

$$\Delta J_{BH} = J_{MM}. \tag{5.30}$$

The fact that energy can be extracted from a rotating black hole raises further concerns. Naively, it seems that this process allows the violation of the cosmic bound. If we start with a black hole that satisfies $a < M$, we can extract energy to lower M so that eventually, $a > M$, thus obtaining a naked singularity. This is problematic for the cosmic censorship conjecture. However, as we will now show, this is impossible.

Consider the null generator (i.e., satisfying $\chi^2|_{r=r_+} = 0$) of the horizon,

$$\chi = \partial_t + \Omega_H \partial_\phi. \tag{5.31}$$

Its norm is

$$|\chi|^2 = g_{tt} + 2g_{t\phi}\Omega_H + g_{\phi\phi}\Omega_H^2. \tag{5.32}$$

When MM travels into the black hole, the condition that he travels forward in time is

$$\chi_\mu p^\mu_{MM} < 0. \tag{5.33}$$

This implies

$$-E_{MM} + \Omega_H J_{MM} < 0. \tag{5.34}$$

Since $E_{MM} < 0$, this bound is only satisfied if MM has angular momentum, thereby changing the angular momentum of the black hole and reducing the size of the ergosphere. Hence, we cannot get an infinite amount of energy. In particular, it must be that the black hole ends up with zero angular momentum *before* losing all its mass. This was first shown by Christodoulou. He introduced the irreducible mass of the black hole,

$$M^2_{\text{irr}} = \frac{1}{2}\left[M^2 + \sqrt{M^4 - J^2}\right]. \tag{5.35}$$

In terms of the mass and angular momentum of the black hole, the bound (5.34) is

$$\Delta M - \Omega_H \Delta J > 0. \tag{5.36}$$

And the change in the irreducible mass during the Penrose process is

$$\Delta M^2_{\text{irr}} = \frac{1}{2}\left[2M\Delta M + \frac{2M^3\Delta M - J\Delta J}{\sqrt{M^4 - J^2}}\right] \tag{5.37}$$

$$= \frac{J}{2\sqrt{M^4 - J^2}}\left[\frac{\Delta M}{\Omega_H} - \Delta J\right] \tag{5.38}$$

$$\geq 0. \tag{5.39}$$

So we find that in the Penrose process, the irreducible mass always increases. In fact, this quantity turns out to be proportional to the area of the event horizon of the black hole.

$$A = \int_{r=r_H} \sqrt{g_{\theta\theta} g_{\phi\phi}}\, d\theta d\phi \tag{5.40}$$

$$= 4\pi(r_+^2 + a^2) \tag{5.41}$$

$$= 16\pi M^2_{\text{irr}}. \tag{5.42}$$

Thus, what Christodoulou showed was that the area of the black hole always increases, i.e.,

$$\Delta A \geq 0. \tag{5.43}$$

In fact, Hawking proved that the area increases not just in energy extraction but in any process. We will review this proof later in the course. When this was discovered, it was called the second law of black hole mechanics due to its analogy to the second law of thermodynamics.

The change in the irreducible mass also provides us with a first law of black hole mechanics. In terms of area, we write

$$\Delta M = \frac{1}{8\pi}\frac{\sqrt{M^4 - J^2}}{2M\left(M^2 + \sqrt{M^4 - J^2}\right)}\Delta A + \Omega_H \Delta J. \tag{5.44}$$

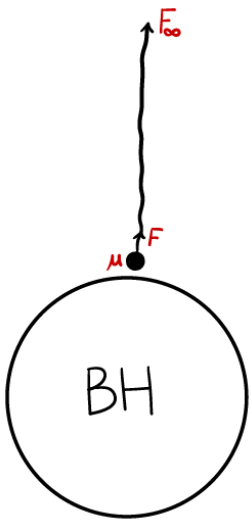


Figure 5.5: Particle suspended at black hole horizon

The coefficient of ΔA is proportional to the *surface gravity* of the black hole. It is the measure of the strength of the gravitational force at the horizon of the black hole. To understand this better, consider a rope with fixed proper length with one end at the horizon of the black hole and the other end fixed at infinity, as in Figure 5.5. Suppose a mass μ is hanging at the horizon. Let F_∞ be the force exerted at infinity required to keep it fixed at the horizon. Note that this is not possible in a Kerr black hole, since all objects must co-rotate with the black hole near the horizon. We therefore imagine this situation for a static black hole.

We then define *surface gravity*

$$\kappa \equiv \frac{F_\infty}{\mu}. \tag{5.45}$$

It is a measure of the gravitational pull just above the event horizon of the black hole. Let's now compute it. First, we note that a stationary black hole has a timelike vector ξ. We define $V^2 = -\xi^\alpha \xi_\alpha$. The energy of a particle moving with four-velocity u^α as measured at infinity is

$$E = -\mu u^\alpha \xi_\alpha. \tag{5.46}$$

If the particle is stationary, i.e., $u^\alpha = \frac{1}{V}\xi^\alpha$, then the energy of the particle as measured at infinity is

$$E = \mu V. \tag{5.47}$$

The force exerted on the particle at infinity is $F_\infty = \|\nabla E\| = \mu \|\nabla V\|$. The surface gravity is then

$$\kappa = \|\nabla V\|. \tag{5.48}$$

An alternative expression for the surface gravity is

$$\kappa^2 = -\frac{1}{2}\nabla_\alpha \xi_\beta \nabla^\alpha \xi^\beta. \tag{5.49}$$

The surface gravity for Schwarzschild black holes is

$$\kappa_{\text{Schwarzschild}} = \frac{1}{4M}, \tag{5.50}$$

and for Reissner-Nordström,

$$\kappa_{\text{RN}} = \frac{1}{2r_+^2}(r_+ - r_-) = \frac{\sqrt{M^2 - Q^2}}{2M^2 - Q^2 + 2M\sqrt{M^2 - Q^2}}. \tag{5.51}$$

In a Kerr black hole, we cannot keep a particle stationary at the horizon. In this case, the surface gravity does not have the same interpretation as above. However, we can continue to define κ as in (5.49) by replacing $\xi \to \chi$. The surface gravity of the Kerr black hole is then

$$\kappa_{\text{Kerr}} = \frac{r_+ - r_-}{2\left(r_+^2 + a^2\right)} = \frac{\sqrt{M^4 - J^2}}{2M\left(M^2 + \sqrt{M^4 - J^2}\right)}. \tag{5.52}$$

By (5.44), the first law for Kerr black holes then reads

$$\Delta M = \frac{\kappa}{8\pi}\Delta A + \Omega_H \Delta J. \tag{5.53}$$

For Reissner-Nordström black holes, the first law reads

$$\Delta M = \frac{\kappa}{8\pi}\Delta A + \Phi \Delta Q. \tag{5.54}$$

5.6 Kerr-Newman metric

The final family of black holes is the Kerr-Newman family of black holes which are charged and rotating. The metric is that of the Kerr solution, now with

$$\Delta = r^2 - 2Mr + a^2 + Q^2, \tag{5.55}$$

and electromagnetic potential

$$A = -\frac{Q}{\rho^2}\left(dt - a\sin^2\theta d\phi\right). \tag{5.56}$$

The Kerr-Newman solutions also satisfy a first law, namely,

$$\Delta M = \frac{\kappa}{8\pi}\Delta A + \Omega_H \Delta J + \Phi \delta Q. \tag{5.57}$$

All charged black holes allow for energy extraction using charged particles. The energy extraction satisfies the area theorem, $\Delta A \geq 0$.

5.7 No-hair theorem

So far, we have discussed Schwarzschild, Reissner-Nordström, Kerr and Kerr-Newman black holes. One might ask whether more general stationary black hole solutions exist. In fact, this is not true due to the *no-hair theorem*, which states that "*A stationary black hole in Einstein-Maxwell gravity is characterised up to diffeomorphisms by the mass M, the charge Q and the angular momentum J.*"

The caveat "up to diffeomorphisms" is important. In general relativity, we identify two metrics only if they are related by *small* diffeomorphisms, i.e., diffeomorphisms generated by a vector ξ that vanish at infinity. In general, there are additional diffeomorphisms that don't vanish at infinity. In that case, related solutions are generally physically different. For instance, a boosted Schwarzschild black hole is clearly different from a static Schwarzschild solution.

The no-hair theorem raises several new issues. For instance, two different initial data with the same total M, Q, J give rise to the same black hole final state. This already makes us uncomfortable since it seems like information is disappearing at infinity. This is known as the *information paradox* and we will come back to it in the end of this course.

6 The laws of black hole thermodynamics

Putting together our discussions so far, we have the *laws of black hole mechanics* as follows.

- Zeroth law:
$$\kappa = \text{const. on } H, \tag{6.1}$$
where κ is the surface gravity and H is the black hole horizon.

- First law:
$$\frac{\kappa}{8\pi}\delta A = \delta M + \Phi\delta Q - \Omega\delta J. \tag{6.2}$$
The first law has a general character because of the no-hair theorem, which says $A = A(Q, M, J)$.

- Second law:
$$\delta A \geq 0. \tag{6.3}$$
The second law is only derived in a special case. There's a general theorem that deals with not just infinitesimal perturbations of the black hole which we will prove later.

We now want to compare them with the *laws of thermodynamics*. The complete discussion will happen in the next semester, where we study quantum black holes.

- Zeroth law:
$$T = \text{const. in equilibrium.} \tag{6.4}$$

- First law:
$$T\delta S = \delta E - P\delta V. \tag{6.5}$$

- Second law:
$$\delta S \geq 0. \tag{6.6}$$

These look pretty similar. We see the correspondence under the exchange of

$$S \leftrightarrow \alpha A, \qquad T \leftrightarrow \frac{1}{\alpha}\frac{\kappa}{8\pi}, \tag{6.7}$$

where α is any constant. This similarity was noticed in the early 70's, but people thought it was a coincidence. When working on black hole radiation, Hawking too believed that this was just a confusing analogy with no hidden complication. But is that all?

In thermodynamics, we would consider systems in which we take a canonical ensemble and fix the energy and volume. We don't usually consider ensembles in which we fix the angular momentum, but we could. Then we would have some number of states of the system, which we denote by $\tilde{\Omega}(E, Q, J)$. And the entropy can be expressed as

$$S = \ln\tilde{\Omega}\,(E, J, Q)\,. \tag{6.8}$$

Note again that we're setting $k_B \equiv 1$. Factors of $\ln\hbar$ also appear in this formula, but these factors do not appear in either the first or the second law of standard thermodynamics. Therefore, as far as Boltzmann was concerned there is no need for a discussion of quantum mechanics. However, in the generalised laws, $\hbar$ appears as a power in $S_{\rm BH}$. Thus, to truly understand the statistical origin of the entropy of black holes, we need to understand the quantum theory of black holes. Continuing our present discussion, we have

$$\delta S_{\rm BH} = \partial_E \ln\tilde{\Omega}\,\delta E + \partial_Q \ln\tilde{\Omega}\,\delta Q + \partial_J \ln\tilde{\Omega}\,\delta J \tag{6.9}$$
$$\equiv \beta\delta E + \beta\Phi\delta Q - \beta\Omega\delta J. \tag{6.10}$$

And the analogy looks better. The derivation of the two sets of laws however, couldn't have been more different. For the black hole one, we derived the first and the second laws using techniques of gravity and differential geometry, but the first and the second laws of thermodynamics are derived using statistical reasoning.

The idea that this might be more than a mathematical analogy first came from Bekenstein and the *cosmic censorship conjecture*, which we will soon discuss in detail. The conjecture has the idea that the inside of a black hole doesn't exist, since we can't do any measurements of it. There is no point concerning us with what's beyond the event horizon (e.g., the naked singularity) of the black hole. With this mindset, it seems like we can have a violation of the second law of thermodynamics. For example, if we take a cup of coffee, which contains a lot of entropy, and throw it into the black hole, then the physical universe, i.e., the region outside of the black hole, has decreased entropy. However, when we do so, the black hole gets a little bigger, i.e., the area of the event horizon gets bigger.

Hence, Bekenstein thought that we should think of a black hole as having entropy proportional to its area. He proposed the *generalised* second law that for any process,

$$\delta S + \alpha \delta A \geq 0. \tag{6.11}$$

In order for this to be true, the universe must have some other characteristics such as a finite bound: the so-called *Bekenstein bound*. In fact,

$$\frac{S}{M} \leq C_M \tag{6.12}$$

for some constant C_M is necessarily the case since if we have a small object to which we add in an arbitrary amount of entropy while keeping the mass fixed, then we can violate the generalised second law. This is because we can decrease the entropy outside the black hole by an arbitrary amount while only increasing the area by a finite amount.

There also needs to be a bound on the amount of entropy that can be fit inside a region whose area is the size of a black hole, i.e.,

$$\frac{S}{A} \leq C_A \tag{6.13}$$

for some constant C_A. To see this, imagine a gas of matter within a volume V , surface area A and entropy S. Now, suppose the matter inside the volume coalesces into a black hole. Since the entropy can only increase in this process, we must have $S \leq S_{\text{BH}}$. However, if $S_{\text{BH}} \propto A$, then $S \leq C_A A$.

Notice that entropy accounts for the number of microstates. In ordinary statistical systems, it scales by the volume of the system instead of the surface area. If we have a three-dimensional box of side length L containing two-state systems, then the number of microstates is $2^N \sim 2^{L^3}$. Hence,

$$S_{\text{stat}} \sim L^3. \tag{6.14}$$

On the other hand,

$$S_{\text{BH}} \sim L^2. \tag{6.15}$$

So it can be thought of as a system with degrees of freedom living on the boundary. The Bekenstein bounds are the first forms of the *holographic principle*.

Now, we can think of the entropy of a system as the number of accessible microstates. But if we only know the energy, we can also think of the entropy as a measure of our ignorance of the system. So it's natural to associate the concept of entropy with a black hole.

Further, if we have a thermodynamic system, and we have n *species* of the entropy, then the entropy is linear in n. So an easy way to violate either of the above two bounds is to postulate more species of particles. If we have both oxygen and carbon dioxide in a box and we only know the energy, then the entropy would be bigger than the situation where we only have oxygen or carbon dioxide in it. Hence, keeping energy fixed, we can increase the entropy simply by adding the number of species. This thought experiment tells us that there has to be some fundamental bound on the number of different kinds of particles in nature.

Hawking managed to calculate α. The key idea is that black holes aren't completely black. Quantum mechanics allows tunnelling, and light can get out. In fact, taking quantum effects into consideration, we have that black holes radiate every kind of particle in nature with the Hawking temperature

$$T_H = \frac{\hbar \kappa}{2\pi}. \tag{6.16}$$

We will understand this formula in more detail when we introduce quantum mechanics. Comparing to the T we had above, we see that

$$\alpha = \frac{1}{4\hbar}. \tag{6.17}$$

Substituting into the above equation for S and putting back the Newton constant G_N and the speed of light c, we have, from the statistical origin of entropy,

$$S_{\text{BH}} = \frac{Ac^3}{4G_N\hbar}. \tag{6.18}$$

So instead of two sets of separate laws, we define the total entropy

$$S_{\text{tot.}} = S_{\text{matter}} + \frac{A}{4}, \tag{6.19}$$

where we have set the constants back to one. For the real world, we have statistical systems as well as gravity. So with the zeroth law remain unchanged, we write our generalised black hole laws as

- Generalised first law:
$$\delta S_{\text{tot.}} = \delta E + \Phi\delta Q - \Omega\delta J. \tag{6.20}$$
- Generalised second law:
$$\delta S_{\text{tot.}} \geq 0. \tag{6.21}$$

We can violate classical mechanics by throwing a cup of coffee into the black hole, and we can violate quantum mechanics since quantum black holes can evaporate. And by energy conservation, if a particle comes out, the black hole must get smaller. But even if the area decreases, it emits black body radiation, which carries entropy to the outside world. We can see that the total entropy increases in the process of black hole evaporation despite the fact that the area of the black hole decreases. Hence, the generalised laws do hold even quantum mechanically.

When we set $\hbar \equiv 0$, we use differential geometry to prove the area theorem. When there's no gravity, we make use of statistical properties of systems. So we need to have a unified viewpoint on this. We don't understand how to count $A/4$ as microstates of black holes. In fact, there's a belief that Einstein's equation is thermodynamics. A lot of the discussions will happen next semester.

7 Area theorem

Now we're at the point to discuss and prove the black hole area theorem,

$$\Delta A \geq 0. \tag{7.1}$$

This is not true for any random type of processes. In proving this, we will need to

- describe the allowed matter content. In particular, we will allow for matter satisfying the null energy condition, i.e., $T_{ab}u^a u^b \geq 0$ for all null geodesics with four-velocity u^a;
- describe what one means by a singularity. For instance, a type of singularity not considered here is the shell-crossing singularity;
- assume the *cosmic censorship conjecture*, which we now discuss.

7.1 Cosmic censorship

The area theorem follows from the *cosmic censorship* conjecture. The essential statement of the conjecture is that singularities are in some sense okay, because they always lie behind an event horizon, i.e., they can never affect the physics in regions outside the event horizon. When the conjecture first came out, it was a fascinating shift in perspective on how we should think of the laws of physics.

More formally, the conjecture states: "*In classical general relativity, all singularities which form from smooth initial data are behind event horizons and not visible from* $\mathscr{I}^+$." In other words, gravitational collapse always leads to a predictable black hole. This statement includes the event horizon and in the most naive sense, it applies to asymptotically flat spacetimes, although more general spacetimes have also been dealt with.

Now, one cannot just get rid of singularities, because they can be created from smooth initial data. Further, whether good data creates singularities or not requires evolving the data to its complete future. Thus, without obtaining information about the future, one cannot throw away such good data. This violates causality.

However, singularities shouldn't bother us since general relativity is not a complete theory. It will be corrected at small distances, and the corrections to the Einstein equations, which involve terms such as R^2 or $R_{\mu\nu\alpha\beta}R^{\mu\nu\alpha\beta}$, can become important near singularities. For instance, string theory predicts a correction of the form

$$R + \frac{1}{4}\ell_s^2 R_{\mu\nu\alpha\beta}R^{\mu\nu\alpha\beta} + \mathcal{O}(\ell_s^4) = 0, \tag{7.2}$$

where ℓ_s is the string scale. The Ricci scalar scales like ℓ^{-2} where ℓ is the radius of curvature of the spacetime. So the additional terms in the formula above are suppressed by $(\ell_s/\ell)^2$. Thus, the Einstein field equations are valid approximations of general relativity whenever $\ell \gg \ell_s$. However, at $\ell \sim \ell_s$, the additional corrections are important. In fact, it's possible that these corrections completely remove the singularities. Further, adding on quantum corrections to the field equations almost certainly introduces naked singularities. Hence, the motivation behind cosmic censorship is rather unclear.

The current status of cosmic censorship is that it's a major problem in mathematics. It's an interesting and non-trivial conjecture and a lot of theorems have been proved. However, one has never found a smooth initial data that evolves into a naked singularity.

7.2 Proof for the spherically symmetric case

Now we're ready to prove the area theorem. We first consider the spherically symmetric case, where many properties of black holes can be qualitatively understood. We start by drawing the Penrose diagram corresponding to gravitational collapse as in Figure 7.1, where the horizon H is the boundary of the causal past of $\mathscr{I}^+$.

We have in the (r,t)-plane, the spherical symmetric metric

$$ds^2 = g_{ab}dx^a dx^b + \mathrm{e}^{-2\phi}d\Omega^2. \tag{7.3}$$

In lightcone coordinates, with $\rho = \rho(x^+, x^-)$ and $\phi = \phi(x^+, x^-)$, it becomes

$$ds^2 = -\mathrm{e}^{2\rho}dx^+dx^- + \mathrm{e}^{-2\phi}d\Omega^2. \tag{7.4}$$

We want to study $\phi(x^+, x_H^-)$, where x_H^- is a null line at the horizon. Note that the area of the horizon of S^2 at (x^+, x_H^-)

$$A = 4\pi\,\mathrm{e}^{-2\phi(x^+, x_H^-)}. \tag{7.5}$$

As we have discussed earlier, for metrics of this form, the four-dimensional Einstein tensor can be written as

$${}^4G_{ab} = 2\nabla_a\nabla_b\phi - 2\nabla_a\phi\nabla_b\phi + 3g_{ab}(\nabla\phi)^2 - 2g_{ab}\Box\phi - \mathrm{e}^{2\phi}g_{ab}. \tag{7.6}$$

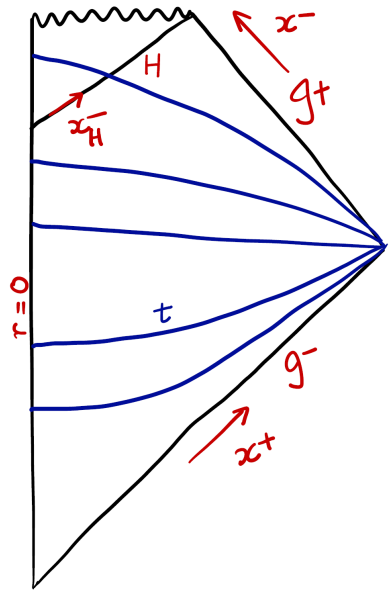


Figure 7.1: Gravitational collapse

In lightcone coordinates, since $g_{++} = 0$, we have

$$
\begin{aligned}
{}^4G_{++} &= 2\nabla_+\nabla_+\phi - 2(\nabla_+\phi)^2 && (7.7)\\
&= 2\partial_+\phi - 2\Gamma^+_{++}\partial_+\phi - 2(\partial_+\phi)^2. && (7.8)
\end{aligned}
$$

Note that we have $\Gamma^+_{++} = 2\partial_+\rho$ and the positivity condition $T_{++} \geq 0$ corresponding to the physical reality that matter is positive. (We will discuss various energy conditions in detail in the next subsection.) Hence, we have from the Einstein equation,

$$
\begin{aligned}
{}^4G_{++} &= 8\pi T_{++} && (7.9)\\
&= 2\partial_+^2\phi - 4\partial_+\rho\partial_+\phi - 2(\partial_+\phi)^2. && (7.10)
\end{aligned}
$$

Next, we write the equation with an affine parameter $s \equiv s(x^+)$ on the horizon, which is of course a geodesic. We have by the geodesic equation,

$$
\frac{d^2x^+}{ds^2} + 2\partial_+\rho\left(\frac{dx^+}{ds}\right)^2 = 0. \tag{7.11}
$$

Clearly,

$$
\frac{dx^+}{ds} = \mathrm{e}^{-2\rho}, \qquad \partial_s = \mathrm{e}^{-2\rho}\partial_+ \tag{7.12}
$$

satisfy the geodesic equation. We also have the term in the metric

$$
\begin{aligned}
\mathrm{e}^{2\rho}dx^+dx^- &= \mathrm{e}^{2\rho}\frac{dx^+}{ds}dsdx^- && (7.13)\\
&= \mathrm{e}^{2\rho(x^+,x^-)-2\rho(x^+,x^-_H)}dsdx^- && (7.14)\\
&= dsdx^- && (7.15)
\end{aligned}
$$

on the horizon H. So we're essentially setting the connection to zero. Hence,

$$
\begin{aligned}
\nabla_s^2\,\mathrm{e}^{-\phi} &= \partial_s^2\,\mathrm{e}^{-\phi} && (7.16)\\
&= -4\pi T_{ss}\,\mathrm{e}^{-\phi} && (7.17)\\
&\leq 0. && (7.18)
\end{aligned}
$$

Notice that it's proportional to $\sqrt{A}$. So since $T_{ss} > 0$, we have that

$$
\partial_s^2\sqrt{A} \leq 0. \tag{7.19}
$$

Suppose that $\partial_s \sqrt{A} < 0$ at some s_0 on H. Then for any $s > s_0$, the area must head towards zero linearly or faster in a finite affine parameter time. The cosmic censorship says that there is no singularity outside or along the horizon, so if the area starts to decrease, we are doomed to have a violation of cosmic censorship.

This is the proof of the area theorem in the spherically symmetric case. We will present a general proof of the theorem, but before we do so, we must first introduce some machinery.

7.3 Energy conditions

An important input in the proof of the area theorem is the choice of energy condition, which is a condition on the stress tensor $T_{\mu\nu}$ of our theory. Basically, these are the rigorous coordinate invariant generalisations of the simple statement that *energy is positive*.

For this purpose, we first start with the basic standard form of the stress tensor in Minkowski spacetime. In the standard $(t, \vec{x})$ coordinates of Minkowski spacetime, $T_{00} = \rho$ is the energy density as measured by an observer that measures time t, i.e., the one whose four-velocity is $u^\mu = (1, 0, 0, 0)$. T_{0i} is the energy flux travelling in the ith direction and T_{ij} corresponds to the flux of the ith component of the momentum along the jth direction.

- Weak energy condition

The statement that the energy density of a system is always positive is

$$\rho \geq 0. \tag{7.20}$$

Note however that this notion of energy positivity is not coordinate invariant. To generalise, we can write

$$\rho = T_{\alpha\beta} u^\alpha u^\beta. \tag{7.21}$$

where u^α is the four-velocity of the observer. An appropriate diffeomorphism invariant notion of energy positivity is then

$$T_{\alpha\beta} u^\alpha u^\beta \geq 0. \tag{7.22}$$

where u is any timelike four-vector. This condition is known as the *weak energy condition* (WEC).

Apart from the WEC, there also exist other energy conditions that one might wish to impose on the stress tensor. Not all of them are as intuitive as the WEC. Therefore, to get a feel for these other energy conditions, we move to a local frame in which the stress tensor is diagonalised. In this frame, it takes the form

$$T_{\mu\nu} = \mathrm{diag}\left(\rho, p_1, p_2, p_3\right), \tag{7.23}$$

where p_i is the stress - sometimes called pressure - in the ith direction. Let's now understand the full condition that the WEC imposes on ρ, p_i. A general timelike four-vector takes the form

$$u^\alpha = u^0 \left(1, a, b, c\right), \qquad u^0 \neq 0, \quad a^2 + b^2 + c^2 < 1. \tag{7.24}$$

Then the WEC reads

$$T_{\alpha\beta} u^\alpha u^\beta = (u^0)^2 \left[\rho + p_1 a^2 + p_2 b^2 + p_3 c^2\right] \geq 0. \tag{7.25}$$

Since this is true for any a, b satisfying the conditions above, we must have

$$\rho \geq 0, \qquad \rho + p_i \geq 0, \qquad i = 1, 2, 3. \tag{7.26}$$

- Null energy condition

An energy condition that corresponds to the positivity of energy with respect to a null observer is

$$T_{\alpha\beta}k^\alpha k^\beta \geq 0, \qquad k^2 = 0. \tag{7.27}$$

This is known as the *null energy condition* (NEC). To understand this, we note that any null vector k can be written as

$$k^\alpha = k^0\,(1, a, b, c)\,, \qquad k^0 \neq 0, \quad a^2 + b^2 + c^2 = 1. \tag{7.28}$$

This implies

$$T_{\alpha\beta}k^\alpha k^\beta = (k^0)^2\left[\rho + p_1 a^2 + p_2 b^2 + p_3 c^2\right] \geq 0. \tag{7.29}$$

Now, writing $c^2 = 1 - a^2 - b^2$, we find

$$\rho + p_3 + (p_1 - p_3)a^2 + (p_2 - p_3)b^2 \geq 0, \tag{7.30}$$

which gives

$$\rho + p_3 \geq 0. \tag{7.31}$$

Similarly, we find $\rho + p_i \geq 0$ for $i = 1, 2, 3$. Thus, the conditions here are

$$\rho + p_i \geq 0, \qquad i = 1, 2, 3. \tag{7.32}$$

- Strong energy condition

A third energy condition is the *strong energy condition* (SEC)

$$\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right)u^\alpha u^\beta \geq 0. \tag{7.33}$$

This is the most important energy condition for our purposes as it shows up in the proof of the area theorem. The physical interpretation of this theorem is less transparent. As we will see, it will turn out to imply that gravity is an attractive force. For now, we simply note here that in terms of ρ, p_i, this condition reads

$$\rho + \sum_{i=1}^{3} p_i \geq 0, \qquad \rho + p_i \geq 0, \qquad i = 1, 2, 3. \tag{7.34}$$

In particular, this condition is not satisfied by dark energy, for which $p = p_1 = p_2 = p_3$ and $\rho = -p$. Thus, the area theorem proven here does not hold in the presence of dark energy.

- Dominant energy condition

For the sake of completeness, we also note a final energy condition, the *dominant energy condition* (DEC) which states that matter should only travel along timelike or null geodesics, i.e.,

$$\ell^\mu = -T^\mu{}_\nu u^\nu \tag{7.35}$$

is future directed and either timelike ($\ell^2 > 0$) or null ($\ell^2 = 0$). This implies

$$\rho \geq 0, \qquad \rho \geq |p_i|. \tag{7.36}$$

7.4 Hypersurfaces

A *hypersurface* Σ is a codimension-one surface in the spacetime. It is specified by the equation

$$\Phi(x^\mu) = 0. \tag{7.37}$$

The hypersurface is classified as spacelike, timelike or null depending on whether $\partial_\mu\Phi$ is timelike, spacelike or null on the hypersurface respectively.

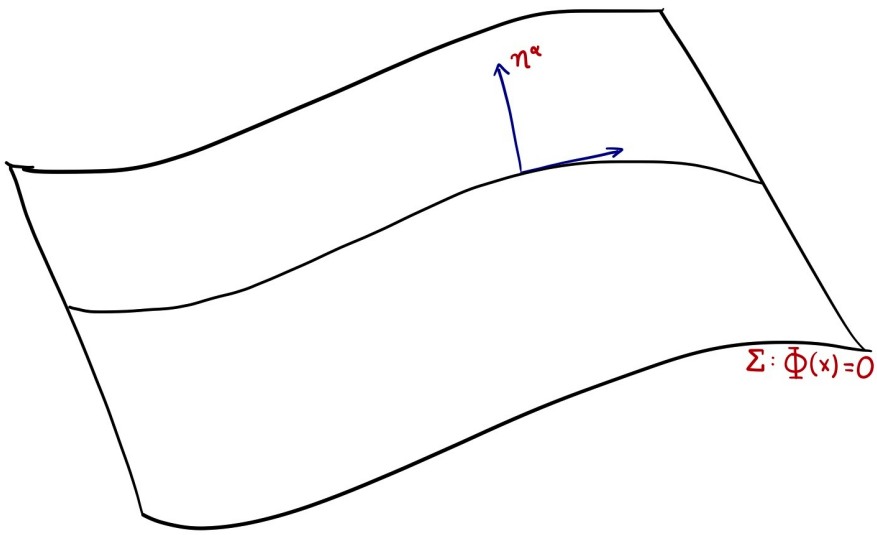


Figure 7.2: Hypersurface

When the hypersurface is not null, we can define the unit normal vector

$$n_\alpha \equiv \frac{\varepsilon \partial_\alpha \Phi}{\left[\varepsilon g^{\alpha\beta} \partial_\alpha \Phi \partial_\beta \Phi\right]^{1/2}}, \tag{7.38}$$

where $\varepsilon = -1$ $(+1)$ for spacelike (timelike) hypersurfaces, since $n^2 = \varepsilon$. If the hypersurface is null, then we define the normal vector as

$$k_\alpha = -\partial_\alpha \Phi, \qquad k^2 = 0, \tag{7.39}$$

where the minus sign simply picks out the future-directed direction if Φ increases towards the future. Null hypersurfaces have several weird properties. For instance, the normal to the hypersurface is also tangent to it. To see this, we note that since Φ is constant along the hypersurface, the vector $\partial_\alpha \Phi$ must be normal to it. However,

$$k^\alpha \partial_\alpha \Phi = -k^2 = 0, \tag{7.40}$$

i.e., k is normal to the normal vector, implying that it is a tangent vector as well. Additionally, every null curve on a null hypersurface is a geodesic. To prove this, let $x^\mu(\tau)$ be a curve such that $\Phi(x(\tau)) = 0$ for all τ, i.e., it lies entirely on the null hypersurface. Then

$$0 = \partial_\tau \Phi(x(\tau)) = u^\alpha \partial_\alpha \Phi(x(\tau)) = -u^\alpha k_\alpha, \tag{7.41}$$

where $u^\alpha = dx^\alpha/d\tau$. However, if the curve is null, then $u^2 = 0$, which then implies that $u^\alpha \propto k^\alpha$. Thus, we have finally proven that every null curve is tangent to k_α. Finally, to prove that the curve is a geodesic, we must show that $k^\alpha \nabla_\alpha k^\beta \propto k^\beta$. We see that

$$k^\beta \nabla_\beta k_\alpha = k^\beta \nabla_\beta \nabla_\alpha \Phi = k^\beta \nabla_\alpha \nabla_\beta \Phi = \nabla^\beta \Phi \nabla_\alpha \nabla_\beta \Phi = \frac{1}{2} \nabla_\alpha k^2. \tag{7.42}$$

But note that k^2 is constant (in fact, zero) on the hypersurface. Thus, $\nabla_\alpha k^2$ can only have non-zero components in the normal direction, i.e., $\nabla_\alpha k^2 \propto k_\alpha$, which is what we set out to prove.

7.5 Geodesic congruences

A congruence is any continuous collection of curves. Here, we study *geodesic congruences*, which are continuous collections of null or timelike geodesics. We could also study congruences of spacelike geodesics but that is not relevant for the present purpose. In fact, studying null geodesics is subtle primarily due to the subtleties of null hypersurfaces described previously. Therefore, for simplicity, we study first timelike geodesics in some detail. We will then briefly describe how those results generalise to null congruences. On a timelike geodesic, we can choose an affine parameter so that

$$u^\mu = \frac{dx^\mu}{d\tau}, \qquad u^\mu u_\mu = -1. \tag{7.43}$$

Under the affine parameterization, we also have

$$u^\mu \nabla_\mu u^\nu = 0. \tag{7.44}$$

For a congruence of geodesics, we consider families of geodesics $x^\mu(\tau, \alpha^i)$ where α^i are spacelike coordinates that span the congruence. We can now define

$$u^\alpha = \frac{\partial x^\mu}{\partial \tau}, \qquad \xi^\mu_i = \frac{\partial x^\mu}{\partial \alpha^i}. \tag{7.45}$$

The deviation vector ξ measures the displacement to the nearest geodesic and satisfies

$$u_\mu \xi^\mu_i = 0. \tag{7.46}$$

Finally, we have

$$\frac{\partial^2 x^\mu}{\partial \alpha^i \partial \tau} = \frac{\partial^2 x^\mu}{\partial \tau \partial \alpha^i}, \tag{7.47}$$

which implies

$$\partial_{\alpha^i} u^\mu = \partial_\tau \xi^\mu_i, \tag{7.48}$$

i.e.,

$$\xi^\nu_i \partial_\nu u^\mu = u^\nu \partial_\nu \xi^\mu_i, \tag{7.49}$$

or, using covariant notations,

$$[\xi, u]^\mu = \xi^\nu_i \nabla_\nu u^\mu - u^\nu \nabla_\nu \xi^\mu_i = 0. \tag{7.50}$$

Next we ask how the deviation vector evolves with τ. This is

$$u^\nu \nabla_\nu \xi^\mu_i = \xi^\nu_i \nabla_\nu u^\mu = B^\mu{}_\nu \xi^\nu_i, \tag{7.51}$$

where we have defined

$$B^\mu{}_\nu \equiv \nabla_\nu u^\mu. \tag{7.52}$$

Let's now study the properties of $B_{\mu\nu}$. First, note that

$$u^\mu B_{\mu\nu} = u^\mu \nabla_\nu u_\mu = \frac{1}{2} \nabla_\nu (u^\mu u_\mu) = 0, \tag{7.53}$$

$$u^\nu B_{\mu\nu} = u^\nu \nabla_\nu u_\mu = 0. \tag{7.54}$$

Thus, $B_{\mu\nu}$ as a matrix has a single null eigenvector and is effectively three dimensional. To describe this, we define a metric

$$h_{\alpha\beta} = g_{\alpha\beta} + u_\alpha u_\beta, \tag{7.55}$$

which is transverse to u since $u^\alpha h_{\alpha\beta} = u_\beta + u^2 u_\beta = 0$, and is effectively three dimensional, i.e., $h^\alpha{}_\alpha = g^{\alpha\beta} h_{\alpha\beta} = 3$. We can now use this transverse metric to decompose $B_{\alpha\beta}$, which is also transverse, into its symmetric and antisymmetric parts as

$$B_{\alpha\beta} = \frac{1}{3} \theta h_{\alpha\beta} + \sigma_{(\alpha\beta)} + \omega_{[\alpha\beta]}, \tag{7.56}$$

where

$$\theta = B^{\mu\nu} h_{\mu\nu}, \qquad \sigma_{\mu\nu} = B_{(\mu\nu)} - \frac{1}{3} \theta h_{\mu\nu}, \qquad \omega_{\mu\nu} = B_{[\mu\nu]}. \tag{7.57}$$

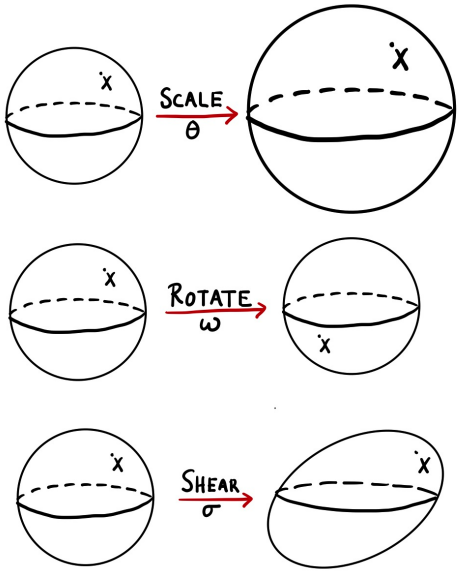


Figure 7.3: Expansion, twist, and shear

The physical interpretation of θ, ω and σ is that of *expansion*, *twist* and *shear* of the geodesic congruence as depicted in Figure 7.3.

The interpretation of θ is quite obvious. Setting $\sigma = \omega = 0$, we find the evolution equation for the deviation vector

$$\frac{\partial \xi_i^\alpha}{\partial \tau} = \frac{1}{3}\theta \xi_i^\alpha, \tag{7.58}$$

i.e.,

$$\xi_i^\alpha(\tau) = \xi_i^\alpha(0) e^{\theta\tau/3}. \tag{7.59}$$

This shows that θ essentially captures the change in the size of the congruence. $\theta > 0$ indicates that congruences diverge in the future and $\theta < 0$ indicates that they converge.

To determine how B changes with time, we compute

$$\begin{aligned} u^\alpha \nabla_\alpha B_{\mu\nu} &= u^\alpha \nabla_\alpha \nabla_\nu u_\mu = u^\alpha \nabla_\nu \nabla_\alpha u_\mu + u^\alpha R_{\alpha\nu\mu\beta} u^\beta \\ &= \nabla_\nu \left(u^\alpha \nabla_\alpha u_\mu\right) - \nabla_\nu u^\alpha \nabla_\alpha u_\mu + u^\alpha R_{\alpha\nu\mu\beta} u^\beta \\ &= -B^\alpha{}_\nu B_{\mu\alpha} + R_{\alpha\nu\mu\beta} u^\alpha u^\beta. \end{aligned} \tag{7.60}$$

Tracing over $g^{\mu\nu}$, we find,

$$\frac{d\theta}{d\tau} = -\frac{1}{3}\theta^2 - \sigma_{\alpha\beta}\sigma^{\alpha\beta} + \omega_{\alpha\beta}\omega^{\alpha\beta} - R_{\alpha\beta} u^\alpha u^\beta. \tag{7.61}$$

This is the famous *Raychaudhuri equation*. Recall that σ and ω are spatial tensors and hence $\sigma^2, \omega^2 \geq 0$. We further note that Einstein's equations imply

$$R_{\alpha\beta} = T_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} T. \tag{7.62}$$

Hence, if the SEC holds, then

$$R_{\alpha\beta} u^\alpha u^\beta \geq 0. \tag{7.63}$$

Finally, let's comment on geodesic congruences that satisfy $\omega_{\alpha\beta} = 0$. We say that a geodesic congruence is hypersurface orthogonal if u^α is orthogonal to a family of hypersurfaces *everywhere*. Note that while the set of normal vectors of a family of hypersurfaces always forms a geodesic congruence, the reverse is not true. According to a theorem due to Frobenius, hypersurface orthogonals satisfy $\omega = 0$ and vice versa.

7.6 Focussing theorem

The discussions above suggest the following theorem. If the SEC is satisfied and the geodesic congruence is hypersurface orthogonal, i.e., $\omega_{\alpha\beta} = 0$, then

$$\frac{d\theta}{d\tau} = -\frac{1}{3}\theta^2 - \sigma_{\alpha\beta}\sigma^{\alpha\beta} - R_{\alpha\beta}u^\alpha u^\beta \leq -\frac{1}{3}\theta^2. \tag{7.64}$$

This implies

$$\theta^{-1}(\tau) \geq \theta_0^{-1} + \frac{\tau}{3}. \tag{7.65}$$

Now, suppose that initially the test particles are converging, i.e., $\theta_0 \leq 0$. Then, eventually $\theta^{-1} \to 0$ and in turn, $\theta \to -\infty$ in at most proper time $\tau = 3|\theta_0^{-1}|$. Thus, if at any time, we have a converging set of particles, then the congruence reaches a singularity in a finite proper time.

Note that the singularity discussed here is a singularity in the congruence (known as a *caustic*), not in the spacetime. A caustic is not a cause for concern. For instance, the null congruence defining the lightcone of a point P has a caustic at P, but the spacetime at that point is perfectly regular.

So far, we have discussed timelike geodesic congruences. For the purposes of the area theorem, we need results for null congruences. A discussion of this is more subtle so we will simply state the results with mild explanations. In particular, a null congruence also has $B_{\alpha\beta}$ but this is now effectively a two-dimensional matrix, i.e., it has two null eigenvalues. This is because, essentially, boosting along τ does not change the nature of the congruence. To understand the quantities θ, ω and σ in this context, we can imagine shining a light on a two-dimensional screen and studying the pattern on the screen as time passes. This is shown in Figure 7.4.

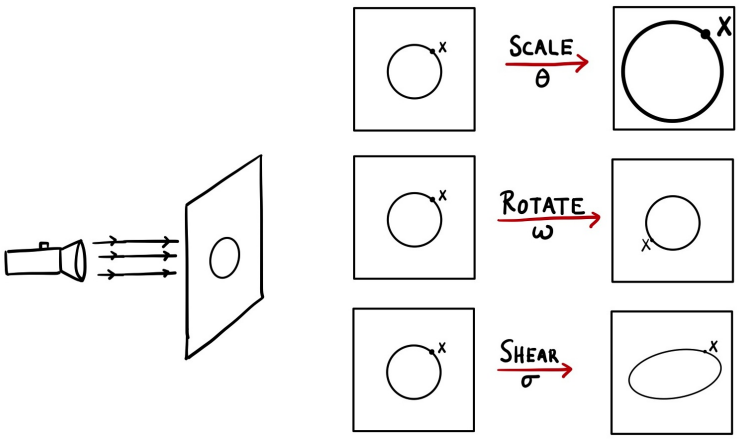


Figure 7.4: Effects of expansion, twist, and shear on a screen

The whole setup is therefore effectively two-dimensional. The corresponding Raychaudhuri equation then takes the form

$$\frac{d\theta}{d\lambda} = -\frac{1}{2}\theta^2 - \sigma_{\alpha\beta}\sigma^{\alpha\beta} + \omega_{\alpha\beta}\omega^{\alpha\beta} - R_{\alpha\beta}k^\alpha k^\beta. \tag{7.66}$$

Notice that the only difference is that coefficient of θ^2 changes from $1/3$ to $1/2$.

The focussing theorem in this context is the statement that if the NEC holds and $\omega = 0$, then

$$\theta^{-1}(\tau) \geq \theta_0^{-1} + \frac{\tau}{2}. \tag{7.67}$$

Suppose that initially the test particles are converging, i.e., $\theta_0 \leq 0$. Then eventually, $\lambda^{-1} \to 0$ and in turn, $\theta \to -\infty$ in at most proper time $\lambda = 2|\theta_0^{-1}|$. So again, if we have a converging set of particles, then the congruence reaches a caustic in a finite proper time.

7.7 Proof for the general case

First recall that the event horizon is a region of spacetime from which nothing can escape to infinity, i.e., $\mathscr{I}$. More precisely, it is the boundary of the past domain of dependence of future null infinity, i.e., $\partial J^-(\mathscr{I}^+)$.

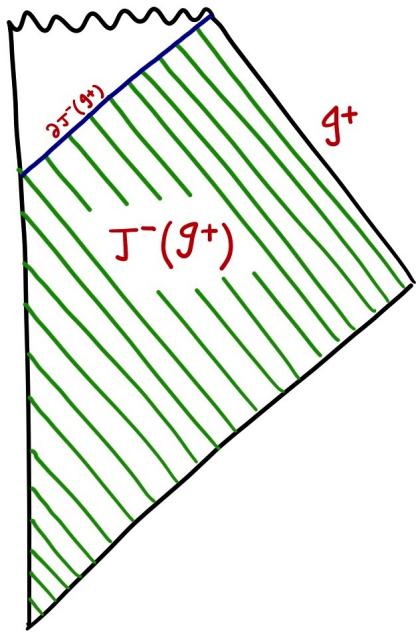


Figure 7.5: $J^-(\mathscr{I}^+)$ and its boundary

Therefore, event horizons are null surfaces. But null hypersurfaces are generated by null geodesics. So since any null hypersurface is generated by null geodesics, the event horizon is a geodesic congruence, which, in particular, satisfies the Raychaudhuri equation.

Another important property of the event horizons, proven by Penrose, is that they cannot have future caustics. The proof of this arises from some global analysis of spacetimes and is quite complex and in itself is not illuminating enough to reproduce here.

Finally, we have everything we need to prove the area theorem in a general setting. The event horizon is a hypersurface orthogonal geodesic congruence. Thus, if the NEC holds then we can use the focusing theorem, which tells us that a caustic must form if $\theta_0 < 0$. However, Penrose's theorem implies that no future caustic can form therefore showing that at all times we must have $\theta \geq 0$. Now, recall that θ is precisely the rate of change of local area. Thus, the area of the event horizon always increases. $\square$

7.8 Singularity theorem

We want to discuss the singularity theorems in the simple context of a solution with spherical symmetry. Before we get to this, we need to introduce some terminologies.

The black hole solutions we have discussed so far are *global*, in the sense that it is not possible to know whether a black hole exists in our future without having a complete evolution of our spacetime into the future. For instance,

at this very moment, there could be a Vaidya shockwave moving towards the origin to form a black hole and we could, at this very moment, be inside the event horizon of this future black hole. This implies that while there are several important statements one can make about event horizons, one cannot measure its existence from any local experiment. We would like to set up a more local definition of a black hole for this purpose. This is where the notion of the *apparent horizon* comes in.

To understand what this is, consider the Vaidya null shockwave solution shown in Figure 7.6.

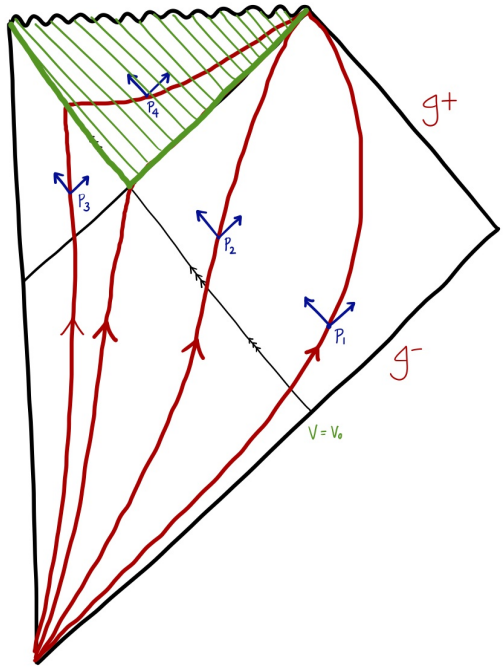


Figure 7.6: Vaidya null shockwave solution

Consider the lines of constant r as shown in red. At any point on the diagram, we have two future null directions. At a generic point P outside the global black hole, the area of the S^2, which is proportional to $e^{-2\phi}$, increases along one null direction and decreases along the other, i.e.,

$$\partial_+ e^{-2\phi} > 0, \qquad \partial_- e^{-2\phi} < 0. \tag{7.68}$$

However, deep within the black hole, the area decreases along both the null directions, i.e.,

$$\partial_\pm e^{-2\phi} < 0, \tag{7.69}$$

or,

$$\partial_\pm \phi > 0. \tag{7.70}$$

There must therefore be a region of spacetime where

$$\partial_+ e^{-2\phi\left(x^+, \hat{x}^-(x^+)\right)} = 0. \tag{7.71}$$

The surface described by the condition above is known as an *apparent horizon* or a future *trapped surface*. The area inside the apparent horizon is known as the *apparent black hole*, which is a local notion of a black hole. The spheres inside the apparent black hole are known as *trapped spheres*. Whenever a trapped surface exists, its causal future necessarily contains a singularity, since all spheres shrink along all future timelike and null directions. This is the statement of Penrose and Hawking's *singularity theorem*. In other words, *"A singularity lies in the future of any trapped surface"*. We prove this for the case where the spacetime is spherically symmetric.

As before, we consider a metric of the form

$$\mathrm{d}s^2 = -e^{2\rho(x^+,x^-)}\mathrm{d}x^+\mathrm{d}x^- + e^{-2\phi(x^+,x^-)}\mathrm{d}\Omega_2^2. \tag{7.72}$$

The $++$ component of the Einstein equation in this background takes the form

$$\nabla_+^2 e^{-\phi} = -4\pi e^{-\phi} T_{++}. \tag{7.73}$$

Consider this equation on the point $\left(x_T^+, x_T^-\right)$ on a trapped surface. At this point, we have

$$\frac{\partial}{\partial x_T^\pm}\phi\left(x_T^+, x_T^-\right) > 0. \tag{7.74}$$

Let's suppose that the trapped surface is defined by the equation $x_T^-(x^+)$. On the trapped surface, we have an affine parameter such that

$$\frac{ds(x^+)}{dx^+} = e^{2\rho\left(x^+, x_T^-(x^+)\right)}. \tag{7.75}$$

Then, on the trapped surface, we have

$$\partial_s^2 e^{-\phi} = -4\pi T_{ss} e^{-\phi}. \tag{7.76}$$

Again, since $T_{ss} > 0$ due to the energy conditions, we must have $\partial_s^2 e^{-\phi} < 0$, implying that the area of the apparent horizon goes to zero in finite affine time. Note that the apparent black hole is always inside the true black hole. As we will see, the area of the apparent horizon also obeys an area theorem.

For the more general situation for a Vaidya spacetime, not necessarily a shockwave, the Penrose diagram is portrayed in Figure 7.7.

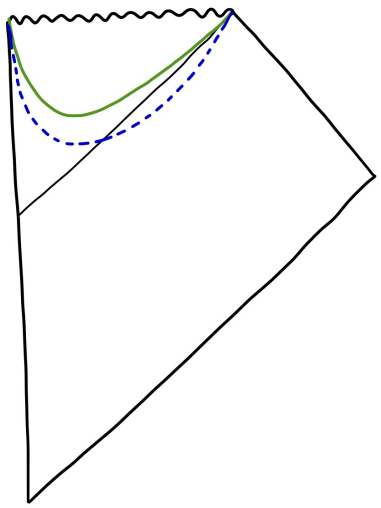


Figure 7.7: General Vaidya spacetime

Note that the solid green line, which is not a timelike surface, is the apparent horizon. It lies completely inside the event horizon. However, it turns out that quantum effects can and do move the apparent horizon outside the event horizon (dashed blue line). It is from the region inside the apparent horizon but outside the event horizon that Hawking radiation emerges.

We next prove that the area of the apparent horizon is non-decreasing for a Vaidya spacetime. Note that for a general spacetime, the apparent horizon is a non-timelike surface. Therefore, simply defining the area of an apparent horizon requires a foliation of spacetime. Thus, the statement about non-decreasing apparent horizon areas is not an invariant statement. However, such a statement can indeed be made in spherically symmetric spacetimes.

Recall that the apparent horizon is the surface $\hat{x}^-(x^+)$ satisfying

$$\frac{\partial}{\partial x^+}\phi(x^+,x^-)\Big|_{x^-=\hat{x}^-(x^+)}=0. \tag{7.77}$$

$\hat{x}^-$ satisfies the differential equation

$$\frac{d}{dx^+}\hat{x}^-(x^+)=-\frac{\nabla_+^2\phi}{\nabla_+\nabla_-\phi}. \tag{7.78}$$

Now, we use the Einstein equation which on the apparent horizon reads

$${}^4G_{+-}=-2\nabla_+\nabla_-\phi-e^{2\phi}g_{+-}=0, \tag{7.79}$$

where we have used the fact that for Vaidya metrics, only $T_{++}\neq 0$. Using this, we find

$$\frac{d\hat{x}^-}{dx^+}=-\frac{\nabla_+^2\phi}{\nabla_+\nabla_-\phi}=-4\pi T_{++}e^{2\rho-2\phi}\leq 0. \tag{7.80}$$

This shows that the apparent horizon is everywhere non-timelike, or that $\hat{x}^-$ is non-increasing. Next, the area

$$A=4\pi e^{-2\phi\left(x^+,\hat{x}^-(x^+)\right)}. \tag{7.81}$$

Thus, the change in the area is

$$\partial_+A=-8\pi e^{-2\phi\left(x^+,\hat{x}^-(x^+)\right)}\partial_{\hat{x}^-}\phi\left(x^+,\hat{x}^-(x^+)\right)\frac{d\hat{x}^-}{dx^+}\geq 0. \qquad \square \tag{7.82}$$

8 Quantum field theory in curved space

8.1 Dilaton black holes

Before we plunge into the big topic of quantum field theory in curved space, we first discuss the important scenario where scalar fields are coupled to Einstein gravity. So far we have discussed Einstein gravity described by the action

$$S=\int d^4x\sqrt{-g}R. \tag{8.1}$$

More general gravity theories that are also of interest and arise often in string theory are those where a scalar field is coupled to gravity in a non-trivial way. In particular, we consider theories of the form

$$S=\int d^4x\sqrt{-g}\left[A(\phi)R-\frac{1}{2}B(\phi)F^2+C(\phi)(\nabla\phi)^2\right]. \tag{8.2}$$

We call this the *Einstein frame*. By redefining the fields ϕ and g, we can set $A(\phi)=C(\phi)=1$. However, $B(\phi)$ describes a non-trivial coupling between the scalar field ϕ and the gauge field A_μ. A more interesting case though is the one in which $A(\phi)=B(\phi)=\frac{1}{4}C(\phi)=e^{-2\phi}$, with the action

$$S=\int d^4x\sqrt{-g}e^{-2\phi}\left[R+4(\nabla\phi)^2-\frac{1}{2}F^2\right]. \tag{8.3}$$

This action is known as *dilaton gravity* and it arises as a low energy effective theory of string theory.

Now, suppose that $F = 0$. Then the no-hair theorem implies that $\phi = \phi_0$, which is a constant. This leads to the Kerr solution. Another interesting possibility comes from conditions of the form $F^2 = \Box\phi$ and leads to dilaton black holes. For example, the action has a magnetically charged black hole solution if we take

$$\begin{aligned} ds^2 &= -4Q^2 \tanh^2\sigma d\tau^2 + \left(2M + \Delta_4 \sinh^2\sigma\right)^2 \left(4d\sigma^2 + d\Omega^2\right), \\ e^{2(\phi-\phi_0)} &= \frac{2M + \Delta_4 \sinh^2\sigma}{\Delta_4 \cosh^2\sigma}, \\ F &= Q \sin\theta d\theta \wedge d\phi, \end{aligned} \tag{8.4}$$

where

$$\Delta_4 = 2M - \frac{Q^2}{2M}. \tag{8.5}$$

The first thing we want to do is to check if the metric is asymptotically flat. The asymptotic region is when $\sigma \to \infty$. There,

$$ds^2 = -4Q^2 d\tau^2 + \frac{1}{4}\Delta_4^2 \, \mathrm{e}^{4\sigma}(4d\sigma^2 + d\Omega^2). \tag{8.6}$$

Next define $t \equiv -2Q\tau$, $r \equiv \mathrm{e}^{2\sigma}$. Then

$$ds^2 = -dt^2 + \frac{\Delta_4^2}{4}(4dr^2 + d\Omega^2). \tag{8.7}$$

So $\sigma \to \infty$ is indeed an asymptotically flat region. The horizon of this region, as usual, has a timelike Killing vector, ∂_τ. The horizon is where it is null, i.e., at $g_{\tau\tau} = -4Q^2 \tanh^2\sigma d\tau^2 = 0$, or, $\sigma = 0$. It is an event horizon.

Next we consider a different region where $\Delta_4 \sinh^2\sigma \ll 2M$. In this case, we have

$$ds^2 = -4Q^2 \tanh^2\sigma d\tau^2 + 4M^2 \left(4d\sigma^2 + d\Omega^2\right). \tag{8.8}$$

This is equivalent to taking $\Delta_4 \to 0$, which gives us $4M^2 = Q^2$. Hence,

$$ds^2 = -4Q^2 \tanh^2\sigma d\tau^2 + 4Q^2 d\sigma^2 + Q^2 d\Omega^2, \tag{8.9}$$

where the first term is nothing but a two-dimensional black hole and the second term is an S^2. Although we let $\Delta_4 \to 0$, we keep $\mathrm{e}^{-2\phi_0}\Delta_4$ fixed. We define $\mathrm{e}^{-2\phi_H} \equiv \frac{1}{2M}\mathrm{e}^{-2\phi_0}\Delta_4$ and take $\mathrm{e}^{-2\phi} = \mathrm{e}^{-2\phi_H}\cosh^2\sigma$. For $\Delta_4 \neq 0$, we have a two-dimensional black hole $\times S^2$.

We might worry that interesting phenomena like Hawking radiation and the information paradox will not be present in this simplified spherically symmetric case of black hole where only time and radial components are present. In fact they're still there. Moreover, in this kind of model, we could systematically glue in the corrections in the original four-dimensional problem. Now, this can all be described by the two-dimensional effective action

$$S = \int d^2\sigma \, \mathrm{e}^{-2\phi}\left(R + 4\nabla\phi^2 + 4\lambda^2\right), \tag{8.10}$$

where we set $\lambda^2 = \frac{1}{4Q^2}$. This is known as the two-dimensional dilaton gravity action.

In our analysis above, we have neglected the non-spherically symmetric perturbations. We can also do a reduction of the original Einstein gravity, without the scalar field or the Maxwell field. If we do that, we get

$$S_{\text{Reduced-Einstein}} = \int d^2\sigma\sqrt{-g}\left[\mathrm{e}^{-2\phi}(R + 4(\nabla\phi)^2) + 2\right]. \tag{8.11}$$

There are two reasons for us to prefer the former action. First, the latter action gives us the Schwarzschild black hole instead. It does stand on its own but doesn't arise as a limit of something in higher dimension. Second, out of sheer luck, in the former case we can solve the equations of motion exactly when matter is coupled to it.

The metric of a two-dimensional black hole is

$$ds^2 = -\frac{1}{\lambda^2}\tanh^2\sigma d\tau^2 + \frac{4}{\lambda^2}d\sigma^2, \tag{8.12}$$

and the resulting Penrose diagram for it is simply the standard Penrose diagram for the Schwarzschild geometry with each point being a point only instead of an S^2.

8.2 Curved background

We now embark on the study of quantum aspects of general relativity. As a first step, prior to quantising gravity itself, we consider the construction of quantum field theories on curved backgrounds. This is a big subject that we curtail due to time constraint. We will try to keep all the interesting conceptual issues using the $(1+1)$-dimension case and of course, discuss black holes.

We start with a free massless scalar f in a $(1+1)$-dimensional curved space. This is described by the action

$$S = -\frac{1}{4\pi}\int d^2x\sqrt{-g}g^{ab}\partial_a f\partial_b f. \tag{8.13}$$

f is the dynamical field that we will quantise, and g_{ab} is an arbitrary but fixed background metric. We introduce null coordinates

$$x^{\pm} = t \pm x. \tag{8.14}$$

With $g_{++} = g_{--} = 0$, we have

$$ds^2 = -\,\mathrm{e}^{2\rho}dx^+dx^-, \tag{8.15}$$

where $\rho = \rho(x^+, x^-)$. So we have $g_{+-} = -\frac{1}{2}\,\mathrm{e}^{2\rho}$. The action simplifies and we arrive at, with $d^2x = dx^+dx^-$,

$$S = -\frac{1}{4\pi}\int d^2x\left[g_{+-}g^{+-}\partial_+ f\partial_- f + g_{-+}g^{-+}\partial_- f\partial_+ f\right] \tag{8.16}$$

$$= -\frac{1}{2\pi}\int d^2x\,\partial_+ f\partial_- f. \tag{8.17}$$

It's interesting that ρ, which appears in the metric, doesn't actually appear in the action. Does this mean it's extremely simple to quantise the theory? The answer is in fact no. As we shall see later, due to an anomaly, the quantum theory does depend on the background.

Note that the form of metric we have doesn't fix all coordinate freedoms. We can take $x^+ \to \tilde{x}^+(x^+)$ and $x^- \to \tilde{x}^-(x^-)$. This is a residual symmetry generated by the so-called *Virasoro algebra*. We can check directly that our action is invariant under this transformation. This residual symmetry will give us a huge amount of control over the quantum theory.

8.3 Quantisation

Now, let's carry out quantisation. With the standard $t \equiv \frac{1}{2}(x^+ + x^-)$ and $x \equiv \frac{1}{2}(x^+ - x^-)$, we have

$$\dot{f} \equiv \partial_t f = \left(\frac{\partial x^+}{\partial t}\partial_+ + \frac{\partial x^-}{\partial t}\partial_-\right) f = \partial_+ f + \partial_- f, \tag{8.18}$$

$$f' \equiv \partial_x f = \left(\frac{\partial x^+}{\partial x}\partial_+ + \frac{\partial x^-}{\partial x}\partial_-\right) f = \partial_+ f - \partial_- f. \tag{8.19}$$

Hence the above action

$$S = -\frac{1}{8\pi}\int d^2x\left(\dot{f}^2 - f'^2\right). \tag{8.20}$$

We have the canonical commutation relations

$$\left[\dot{f}(x,t), f(x',t)\right] = -2\pi i\delta(x-x'). \tag{8.21}$$

And the equation of motion is simply

$$\partial_+\partial_- f = 0, \tag{8.22}$$

which has the general solution

$$f(x^+, x^-) = f_+(x^+) + f_-(x^-). \tag{8.23}$$

These can be seen as massless waves which are moving to the left or to the right at the speed of light. We now have the commutation relations for massless scalar fields as

$$[\partial_+ f_+(x^+), f_+(x^{+\prime})] = -i\pi\delta(x^+ - x^{+\prime}), \tag{8.24}$$

$$[\partial_- f_-(x^-), f_-(x^{-\prime})] = -i\pi\delta(x^- - x^{-\prime}). \tag{8.25}$$

We have a bit more information here compared to the previous commutator since these aren't equal time commutation relations. Note that we basically have two copies of the same problem, so let's just focus on the f_- one. We call them the *right movers* and now discuss their mode expansion. Define modes

$$u_\omega(x^-) \equiv \frac{1}{\sqrt{2\omega}}\, \mathrm{e}^{-i\omega x^-}, \tag{8.26}$$

so that we can expand in modes

$$f_- = \int_0^\infty d\omega \left[a_\omega u_\omega(x^-) + a_\omega^\dagger u_\omega^*(x^-)\right]. \tag{8.27}$$

And we postulate that

$$[a_\omega, a_\omega^\dagger] = \delta(\omega - \omega'). \tag{8.28}$$

Now, for any massless complex scalar field theory, we can find a current

$$J_\mu^{12} \equiv i\left(\phi_1^*\partial_\mu\phi_2 - (\partial_\mu\phi_1^*)\phi_2\right) \equiv -i\phi_1 \overset{\leftrightarrow}{\partial}_\mu \phi_2^*, \tag{8.29}$$

which is conserved, i.e.,

$$\nabla^\mu J_\mu^{12} = 0. \tag{8.30}$$

So if we consider spacelike surfaces Σ and take the normals to the surface, we can construct the quantity

$$\int d\Sigma^\mu J_\mu^{12}, \tag{8.31}$$

which is independent of the surfaces. And it can be used to define the norm

$$(\phi_1, \phi_2) = \int d\Sigma^\mu J_\mu^{12}. \tag{8.32}$$

To derive the norm of our interest, let Σ be defined as the surface $x^+ =$ constant. Then, $n_\mu = -\partial_\mu x^+ = -\delta_\mu^+$ implies that $n^\mu = e^{-2\rho}\delta_-^\mu$. Then,

$$d\Sigma^\mu = dx^-\sqrt{-g}n^\mu = dx^-\delta_-^\mu. \tag{8.33}$$

Plugging this into the norm previously defined, we get

$$(\phi_1, \phi_2) = -i\int_{-\infty}^\infty dx^-\phi_1 \overset{\leftrightarrow}{\partial}_- \phi_2^*. \tag{8.34}$$

This norm is linear in its first argument and anti-linear in its second argument. It satisfies

$$(\phi_1, \phi_2) = (\phi_2, \phi_1)^* = -\left(\phi_2^*, \phi_1^*\right). \tag{8.35}$$

So the norm for u_ω

$$(u_\omega, u_{\omega'}) = i\int dx^- u_\omega^* \overset{\leftrightarrow}{\partial}_- u_{\omega'} = 2\pi\delta(\omega - \omega'). \tag{8.36}$$

Similarly,

$$(u_\omega^*, u_{\omega'}^*) = -2\pi\delta(\omega - \omega'), \qquad (u_\omega, u_{\omega'}^*) = 0 \tag{8.37}$$

The u_ω's are orthogonal and above we simply have the orthogonality condition. They also form a complete set so for the completeness condition, write

$$A(x^-) = \frac{1}{2\pi}\int d\omega \left[(A, u_\omega)u_\omega(x^-) - (A, u_\omega^*)u_\omega^*(x^-)\right], \tag{8.38}$$

giving

$$(A, B) = \int_0^\infty d\omega \left[(A, u_\omega)\,(u_\omega, B) - (A, u_\omega^*)\,(u_\omega^*, B)\right]. \tag{8.39}$$

Likewise for f^+.

8.4 Defining the vacuum

Next, let's define vacuum state $|0\rangle$ such that $a_\omega |0\rangle = 0$ for all $\omega > 0$. We want to see if we have an invariance under the transformation to new coordinates

$$\tilde{x}^- = \tilde{x}^-(x^-). \tag{8.40}$$

We can define associated modes

$$\tilde{u}_\omega = u_\omega(\tilde{x}^-) = \frac{1}{\sqrt{2\omega}} \mathrm{e}^{-i\omega\tilde{x}^-}. \tag{8.41}$$

We then expand in these modes

$$f_- = \int_0^\infty d\omega \left[\tilde{a}_\omega \tilde{u}_\omega + \tilde{a}_\omega^\dagger \tilde{u}_\omega^*\right]. \tag{8.42}$$

And we find

$$[\tilde{a}_\omega, \tilde{a}_\omega^\dagger] = \delta(\omega - \omega') \tag{8.43}$$

given that

$$[\tilde{\partial}_- f(\tilde{x}^-), f(\tilde{x}^{-\prime})] = -\pi i \delta(\tilde{x}^- - \tilde{x}^{-\prime}). \tag{8.44}$$

We are now motivated to define a new state $\left|\tilde{0}\right\rangle$ such that $\tilde{a}_\omega \left|\tilde{0}\right\rangle = 0$. However, as we will see, we do *not* have $\left|\tilde{0}\right\rangle = |0\rangle$. The commutation relations work out fine, but in order to define a vacuum, we need to split the modes into those with positive frequencies and those with negative frequencies, where negative frequency modes annihilate the vacuum. And we need to find a way to write $\tilde{a}_\omega$ in terms of a_ω. Let's write

$$\begin{aligned}\tilde{u}_\omega &= \frac{1}{2\pi}\int_0^\infty d\omega' \left[(\tilde{u}_\omega, u_{\omega'})u_{\omega'} - (\tilde{u}_\omega, u_{\omega'}^*)u_{\omega'}^*\right] \\ &= \int_0^\infty d\omega' \left[\alpha_{\omega\omega'} u_{\omega'}(x^-) + \beta_{\omega\omega'} u_{\omega'}^*(x^-)\right],\end{aligned}$$

where we have defined

$$\alpha_{\omega\omega'} \equiv \frac{1}{2\pi}(\tilde{u}_\omega, u_{\omega'}) = -\frac{i}{2\pi}\int dx^- \frac{1}{2\sqrt{\omega\omega'}} \mathrm{e}^{-i\omega\tilde{x}^-(x^-)} \overleftrightarrow{\partial}_- \mathrm{e}^{i\omega' x^-}, \tag{8.45}$$

$$\beta_{\omega\omega'} \equiv -\frac{1}{2\pi}(\tilde{u}_\omega, u_{\omega'}^*) = \frac{i}{2\pi}\int dx^- \frac{1}{2\sqrt{\omega\omega'}} \mathrm{e}^{-i\omega\tilde{x}^-(x^-)} \overleftrightarrow{\partial}_- \mathrm{e}^{-i\omega' x^-}. \tag{8.46}$$

Using this, we can write

$$\tilde{u}_\omega(x^-) = \int_0^\infty d\omega' \left[\alpha_{\omega\omega'} u_{\omega'}(x^-) + \beta_{\omega\omega'} u_{\omega'}^*(x^-)\right]. \tag{8.47}$$

Before proceeding further, let us pause and derive some properties of the Bogoliubov coefficients.

- Alternative forms: Using (8.35), we can write

$$\begin{aligned}\alpha_{\omega\omega'} &= \frac{1}{2\pi}\left(\tilde{u}_\omega, u_{\omega'}\right) = \frac{1}{2\pi}\left(u_{\omega'}, \tilde{u}_\omega\right)^* = -\frac{1}{2\pi}\left(u_{\omega'}^*, \tilde{u}_\omega^*\right), \\ \beta_{\omega\omega'} &= -\frac{1}{2\pi}\left(\tilde{u}_\omega, u_{\omega'}^*\right) = -\frac{1}{2\pi}\left(u_{\omega'}^*, \tilde{u}_\omega\right)^* = \frac{1}{2\pi}\left(u_{\omega'}, \tilde{u}_\omega^*\right).\end{aligned} \tag{8.48}$$

- Completeness: Using the completeness relations (8.39), we find

$$\begin{aligned}\int_0^\infty d\omega' \left[\alpha_{\omega\omega'}\alpha_{\omega''\omega'}^* - \beta_{\omega\omega'}\beta_{\omega''\omega'}^*\right] &= \delta(\omega - \omega''), \\ \int_0^\infty d\omega'' \left[\alpha_{\omega\omega''}\beta_{\omega'\omega''} - \beta_{\omega\omega''}\alpha_{\omega'\omega''}\right] &= 0.\end{aligned} \tag{8.49}$$

- Inversion: Since $\tilde{u}$ are complete modes, we can also expand

$$\begin{aligned} u_\omega(x^-) &= \frac{1}{2\pi}\int_0^\infty d\omega' \left[(u_\omega, \tilde{u}_{\omega'})\,\tilde{u}_{\omega'}(x^-) - (u_\omega, \tilde{u}^*_{\omega'})\,\tilde{u}^*_{\omega'}(x^-)\right] \\ &= \int_0^\infty d\omega' \left[\alpha^*_{\omega'\omega}\tilde{u}_{\omega'}(x^-) - \beta_{\omega'\omega}\tilde{u}^*_{\omega'}(x^-)\right]. \end{aligned} \tag{8.50}$$

Now, inserting the expansion of $\tilde{u}$ in the above expression, we find

$$\begin{aligned} u_\omega(x^-) = \int_0^\infty d\omega'' \Bigg[&\int_0^\infty d\omega' \left(\alpha^*_{\omega'\omega}\alpha_{\omega'\omega''} - \beta_{\omega'\omega}\beta^*_{\omega'\omega''}\right) u_{\omega''}(x^-) \\ &+ \int_0^\infty d\omega' \left(\alpha^*_{\omega'\omega}\beta_{\omega'\omega''} - \beta_{\omega'\omega}\alpha^*_{\omega'\omega''}\right) u^*_{\omega''}(x^-)\Bigg]. \end{aligned} \tag{8.51}$$

This implies the properties

$$\begin{aligned} &\int_0^\infty d\omega' \left(\alpha^*_{\omega'\omega}\alpha_{\omega'\omega''} - \beta_{\omega'\omega}\beta^*_{\omega'\omega''}\right) = \delta(\omega-\omega''), \\ &\int_0^\infty d\omega' \left(\alpha^*_{\omega'\omega}\beta_{\omega'\omega''} - \beta_{\omega'\omega}\alpha^*_{\omega'\omega''}\right) = 0. \end{aligned} \tag{8.52}$$

We now have all the properties we need. Before we proceed further, we introduce a notation that simplifies the formulae of this section. We treat ω as in infinite matrix index. a and u are treated as a column vectors and the Bogoliubov coefficients α, β are matrices. Usual rules for matrix multiplication apply, e.g.,

$$\begin{aligned} (u^T\alpha)_\omega &\equiv \int_0^\infty d\omega' u_{\omega'}\alpha_{\omega'\omega}, \\ a^\dagger\alpha a &\equiv \int_0^\infty d\omega \int_0^\infty d\omega' a^\dagger_\omega \alpha_{\omega\omega'} a_{\omega'}. \end{aligned} \tag{8.53}$$

The identity matrix in this notation is

$$\mathbb{1} \equiv \mathbb{1}_{\omega\omega'} = \delta(\omega-\omega'). \tag{8.54}$$

Using this notation, α and β satisfy the properties

$$\begin{aligned} \alpha\alpha^\dagger - \beta\beta^\dagger &= \alpha^\dagger\alpha - \beta^\dagger\beta = \mathbb{1}, \\ \alpha\beta^T - \beta\alpha^T &= \alpha^\dagger\beta - \beta^T\alpha^* = 0. \end{aligned} \tag{8.55}$$

The modes are expanded as

$$\tilde{u} = \alpha u + \beta u^*, \qquad u = \alpha^\dagger\tilde{u} - \beta^T\tilde{u}^*. \tag{8.56}$$

Further, the mode expansion of f_- is

$$f_- = \tilde{u}^T\tilde{a} + \tilde{u}^\dagger\tilde{a}^* = u^T a + u^\dagger a^*. \tag{8.57}$$

Now, using (8.56), we find

$$u^T\left(\alpha^T\tilde{a} + \beta^\dagger\tilde{a}^*\right) + u^\dagger\left(\beta^T\tilde{a} + \alpha^\dagger\tilde{a}^*\right). \tag{8.58}$$

Hence,

$$a = \alpha^T\tilde{a} + \beta^\dagger\tilde{a}^*. \tag{8.59}$$

If $\beta \neq 0$, then there is a mixing between a and $\tilde{a}^*$ which immediately implies that $|0\rangle$ and $|\tilde{0}\rangle$ are not the same vacuum states. To derive a precise relationship between the two, we look for a state such that

$$a\,|0\rangle = \left(\alpha^T\tilde{a} + \beta^\dagger\tilde{a}^*\right)|0\rangle = 0. \tag{8.60}$$

Thus, in the tilde'd Fock space, $|0\rangle$ is some sort of squeezed state. To determine it in terms of $|\tilde{0}\rangle$, we make the ansatz

$$|0\rangle = N e^{\tilde{a}^\dagger A\tilde{a}^*}\,|\tilde{0}\rangle. \tag{8.61}$$

for some normalisation factor N and some matrix A satisfying $A^T = A$. So to fix A, we require

$$\left(\alpha^T \tilde{a} + \beta^\dagger \tilde{a}^*\right) e^{\tilde{a}^\dagger A \tilde{a}^*} \left|\tilde{0}\right\rangle = 0. \tag{8.62}$$

To simplify this, we will need the identity

$$Ae^B = e^B \left(A + [A, B]\right) \qquad \text{if} \qquad [[A, B], B] = 0. \tag{8.63}$$

Using this identity, we find

$$\alpha^T \tilde{a} e^{\tilde{a}^\dagger A \tilde{a}^*} = e^{\tilde{a}^\dagger A \tilde{a}^*} \left(\alpha \tilde{a} + \left[\alpha^T \tilde{a}, \tilde{a}^\dagger A \tilde{a}^*\right]\right) = e^{\tilde{a}^\dagger A \tilde{a}^*} \left(\alpha^T \tilde{a} + 2\alpha^T A \tilde{a}^*\right). \tag{8.64}$$

This implies

$$\begin{aligned} \left(\alpha^T \tilde{a} + \beta^\dagger \tilde{a}^*\right) e^{\tilde{a}^\dagger A \tilde{a}^*} \left|\tilde{0}\right\rangle &= e^{\tilde{a}^\dagger A \tilde{a}^*} \left(\alpha^T \tilde{a} + 2\alpha^T A \tilde{a}^* + \beta^\dagger \tilde{a}^*\right) \left|\tilde{0}\right\rangle \\ &= 2 e^{\tilde{a}^\dagger A \tilde{a}^*} \alpha^T \left(A + \frac{1}{2} \alpha^{-T} \beta^\dagger\right) \tilde{a}^* \left|\tilde{0}\right\rangle. \end{aligned} \tag{8.65}$$

So we must have

$$A = -\frac{1}{2} \alpha^{-T} \beta^\dagger = -\frac{1}{2} \beta^* \alpha^{-1}. \tag{8.66}$$

Thus, we find

$$\left|0\right\rangle = N \, \mathrm{e}^{-\frac{1}{2} \tilde{a}^\dagger \beta^* \alpha^{-1} \tilde{a}^*} \left|\tilde{0}\right\rangle, \tag{8.67}$$

where N is a normalisation factor worked out to be

$$N = \left(1 - \left(\frac{\beta^*}{\alpha}\right)^2\right)^{-1/4}. \tag{8.68}$$

This kind of calculations, while being more complicated, can be done in (3+1)-dimension. Now, will the "real" vacuum stand out? This is perhaps a strange question to ask since in a quantum field theory class, we start off having *the* vacuum to work with.

Let's use the coordinates $x^\pm = t \pm x$ and take the $x^\pm(a)$ vacuum. Observers with worldlines at $x^+ = c_0 x^-$ detect no particle, where c_0 is just some constant. We also take the $\tilde{x}^\pm(a)$ vacuum. Then observers with worldlines at $\tilde{x}^+ = c_0 \tilde{x}^-$ also detect no particle. By "detecting no particle", we mean that the observers think that they're in a vacuum. To decide which one the real vacuum is, we have to know the metric. Suppose we have

$$ds^2 = -dx^+ dx^- \neq -d\tilde{x}^+ d\tilde{x}^-. \tag{8.69}$$

Then $x^+ = c_0 x^-$ is an inertial trajectory while $\tilde{x}^+ = c_0 \tilde{x}^-$ is not. So if we specify such a flat metric, we do have a preferred vacuum. However, not all two-dimensional metric are flat. We can take for example

$$ds^2 = -\,\mathrm{e}^{2\rho(x^+, x^-)} dx^+ dx^-. \tag{8.70}$$

Then as we have seen, the Ricci tensor

$$R_{+-} \propto \partial_+ \rho \, \partial_- \rho \neq 0. \tag{8.71}$$

So in general, there isn't a preferred set of coordinates and there is no preferred vacuum. In other words, a particle is not a coordinate invariant concept. One can associate coordinates with families of observers, and different observers will disagree as to whether there are particles. We see that a particle is associated with the mode of a field, created by positive frequency modes and annihilated by negative frequency modes. And the modes are in turn associated with the coordinate system we use.

8.5 Unruh detector

The notion that a particle is not a coordinate invariant concept was understood in the late sixties and caused a fair bit of consternation. But then came William George Unruh, whose work allowed us to now discuss more carefully what it means for a particle to be detected.

Take what's called the monopole detector, a field $m(\tau)$, which is excited whenever a particle is observed at a point τ on some worldline with affine parameter τ. We can for instance take $m(\tau)$ to be a scalar field with an action of the form

$$S = \frac{1}{2}\int d\tau \left[(\partial_\tau m(\tau))^2 + m^2(\tau)\right]. \tag{8.72}$$

We postulate that the detector which moves along the worldline couples to the quantum field in $(1+1)$-dimension, and we introduce a coupling constant g which is taken to be small. Then we can write down an interaction Lagrangian which couples the Hilbert space of the detector to that of the free quantum fields.

$$S_{\text{int}} = g\int d\tau\, m(\tau)\partial_\tau f(x^+(\tau), x^-(\tau)). \tag{8.73}$$

The full Hilbert space of the system is a tensor product of the Hilbert space of the detector and that of the scalar field. Now, suppose that the detector lies in an initial state $|M_i\rangle \otimes |f_{\text{in}}\rangle$ and after an interaction is $|M_f\rangle \otimes |f_{\text{out}}\rangle$. By the first Born approximation, we have the relation

$$|M_f\rangle\,|f_{\text{out}}\rangle = (1 + iS_{\text{int}})\,|M_i\rangle\,|f_{\text{in}}\rangle + \mathcal{O}(g^2). \tag{8.74}$$

Hence, since we know the interactions explicitly, we have

$$|M_f\rangle\,|f_{\text{out}}\rangle = \left(1 + ig\int_{-\infty}^{\infty} d\tau\, m(\tau)\partial_\tau f\right)|M_i\rangle\,|f_{\text{in}}\rangle\,. \tag{8.75}$$

Moreover, we assume that the detector starts in its ground state, i.e.,

$$\int d\tau\, m(\tau)\,\mathrm{e}^{i\omega\tau}\,|M_i\rangle = 0, \qquad \omega > 0. \tag{8.76}$$

We now deal with the simplest case in flat space. For some constant c_0, set $x^+ \equiv c_0^2 x^-$. Then

$$ds^2 = -dx^+dx^- = -c_0^2 dx^{-2}. \tag{8.77}$$

Hence we have the proper time $\tau = c_0 x^-$. Let's also take $|f_{\text{in}}\rangle$ be the standard Minkowski vacuum $|0_{\text{Mink}}\rangle$ defined with respect to x^+, x^- so that it is annihilated by all negative frequency modes of the field f. Then, using our usual mode expansions for f and neglecting the x^+ modes, we have

$$|M_f\rangle\,|f_{\text{out}}\rangle = \left[1 + g\int_{-\infty}^{\infty} d\tau\, m(\tau)\int_0^\infty \sqrt{\frac{\omega}{2}}\left(a_\omega\,\mathrm{e}^{-i\omega\tau/c} - a_\omega^\dagger\,\mathrm{e}^{i\omega\tau/c}\right)\right]|M_i\rangle\,|0_{\text{Mink}}\rangle\,. \tag{8.78}$$

The first term annihilates $|0_{\text{Mink}}\rangle$ so it can be ignored. But the second term also vanishes because they are positive frequency modes of $m(\tau)$, which, by (8.77), annihilates the incoming state. So nothing happens and the incoming and outgoing states are the same. Of course, if we aren't moving on some inertial trajectory and τ was some complicated function of x^-, then in general we'll get something non-trivial.

We now look at the example where we have a uniformly accelerated (i.e., non-inertial) trajectory

$$x^+x^- = t^2 - x^2 = -\frac{1}{a^2}. \tag{8.79}$$

We have an accelerated trajectory since, with tangent vector $u^a = dx^a/d\tau$, its acceleration vector $a^b = u^c\nabla_c u^b$, which has norm a. Then, the proper time from the distance of closest approach, $-\frac{1}{a}$, to an arbitrary point $\tau(x^-)$ is

$$\tau(x^-) = \int d\tau = \int_{x^{-\prime}=-\frac{1}{a}}^{x^{-\prime}=x^-} \sqrt{dx^+dx^{-\prime}} = \int_{-\frac{1}{a}}^{x^-} \frac{1}{ax^{-\prime}}dx^{-\prime} = -\frac{1}{a}\ln(-ax^-). \tag{8.80}$$

Hence we have

$$x^- = \frac{1}{a}\,\mathrm{e}^{-a\tau}, \qquad x^+ = \frac{1}{a}\,\mathrm{e}^{a\tau}. \tag{8.81}$$

Notice that they are periodic under

$$\tau \to \tau + \frac{2\pi i n}{a}, \qquad a \in \mathbb{Z}. \tag{8.82}$$

Now, suppose that the detector has eigenstates $|E_i\rangle$. We can compute the transition probability to go from $|E_i\rangle$ to $|E_j\rangle$. The outgoing state

$$|\psi_{\text{out}}\rangle = \left(1 + ig\int d\tau\, m\partial_\tau f\right)|0_{\text{in}}\rangle\,|E_i\rangle\,. \tag{8.83}$$

So there's some Hamiltonian H such that $H\,|E_i\rangle = E_i\,|E_i\rangle$, and we can write

$$m(\tau) = \mathrm{e}^{iH\tau} m(0)\,\mathrm{e}^{-iH\tau}. \tag{8.84}$$

Take the out state to be $|\alpha\rangle$ and some other detector state to be $|E_j\rangle$. Then,

$$\langle\alpha|\,\langle E_j|\,\psi_{\text{out}}\rangle = \int d\tau\,\mathrm{e}^{i(E_j-E_i)\tau}\,\langle E_j|\,gm(0)\,|E_i\rangle\,\langle\alpha|\,\partial_\tau f(x(\tau))\,|0_{\text{in}}\rangle\,. \tag{8.85}$$

We define the matrix element $m_{ij} \equiv \langle E_j|\,gm(0)\,|E_i\rangle$. We want to find the transition probability of starting the scalar field in the vacuum state with the detector in the state $|E_i\rangle$ and going to the state $|E_i\rangle$. Note that we don't care about the final states of the scalar field. So we square the amplitude and sum over all final states of the scalar field to find that the transition probability

$$P_{E_i\to E_j} = \sum_\alpha \int d\tau d\tau'\,\mathrm{e}^{i(E_j-E_i)(\tau-\tau')} m_{ij} m_{ij}^*\,\langle 0_{\text{in}}|\,\partial_{\tau'} f(x(\tau'))\,|\alpha\rangle\,\langle\alpha|\,\partial_\tau f(x(\tau))\,|0_{\text{in}}\rangle\,. \tag{8.86}$$

We can use the fact that we're not measuring anything in the final state of the scalar field to take $\sum_\alpha |\alpha\rangle\,\langle\alpha| = \mathbb{1}$. Then, integrating over τ' and dividing it out, we have that the transition probability per unit time

$$\Gamma_{E_i\to E_j} = |m_{ij}|^2 \int d\tau\,\mathrm{e}^{i(E_j-E_i)}\,\langle 0_{\text{in}}|\,\partial_\tau f(x(0))\partial_\tau f(x(\tau))\,|0_{\text{in}}\rangle\,. \tag{8.87}$$

After a bit of work, we see that

$$\Gamma_{E_i\to E_j} = \frac{C(E_j-E_i)|m_{ij}|^2}{\mathrm{e}^{2\pi(E_j-E_i)/a}-1}, \tag{8.88}$$

where C is an even function of $(E_i - E_j)$.

Now, if we wait for a very long time, what would be the asymptotic distributions of states of the detector? Not every detector will eventually equilibrate. The density matrix

$$\rho = \sum_j p_j\,|E_j\rangle\,\langle E_j|\,, \qquad \sum_j p_j = 1. \tag{8.89}$$

Then we have

$$P_{E_i\to E_j} = p_i \Gamma_{E_i\to E_j} \tag{8.90}$$

But if we reach equilibrium, then

$$p_i\Gamma_{E_i\to E_j} = P_{E_j\to E_i} = p_j\Gamma_{E_j\to E_i}. \tag{8.91}$$

So we can calculate the ratio

$$\frac{p_i}{p_j} = \frac{\Gamma_{E_j\to E_i}}{\Gamma_{E_i\to E_j}} = \frac{1-\mathrm{e}^{2\pi(E_j-E_i)/a}}{\mathrm{e}^{2\pi(E_i-E_j)/a}-1} = \mathrm{e}^{2\pi(E_j-E_i)/a}. \tag{8.92}$$

Hence we have, at temperature $T = a/2\pi$,

$$p_i = \frac{\mathrm{e}^{-2\pi E_i/a}}{\sum_j \mathrm{e}^{-2\pi E_j/a}}. \tag{8.93}$$

So we see that the detector starts off in the ground state, moves along this accelerated trajectory, becomes equilibrated with the surrounding scalar field, and ends up in a state at $T = a/2\pi$, the Rindler temperature. Note that this is independent of the details of the detector. However, the rate of equilibration does depend on these details. On the other hand, an inertial observer observes that the detector simply moves along an accelerated trajectory and radiates out particles. He will nevertheless come to the same conclusion that the thermal temperature at the end is $T = a/2\pi$. So at the end of the day, the coordinate invariant statement is that we have a thermally populated detector.

Note the similarity between the Rindler temperature and the Hawking temperature $T = \kappa/2\pi$. In fact, these two formulae have the same origin: The former corresponds to the case of an accelerated detector near the Rindler horizon and the latter, a black hole event horizon.

Further, recall that the equivalence principle relates observers in uniformly accelerated fields with those in constant gravitational fields. The principle is manifested here as the temperature measured by a uniformly accelerated detector in Minkowski space is the same as that measured by a stationary detector in a uniform gravitational field.

8.6 Another point of view

Next we introduce Rindler coordinates

$$y^+ = \ln(x^+), \qquad y^- = -\ln(-x^-). \tag{8.94}$$

The signs are such that they both increase towards the future. This corresponds to the Rindler wedge as shown in Figure 8.1.

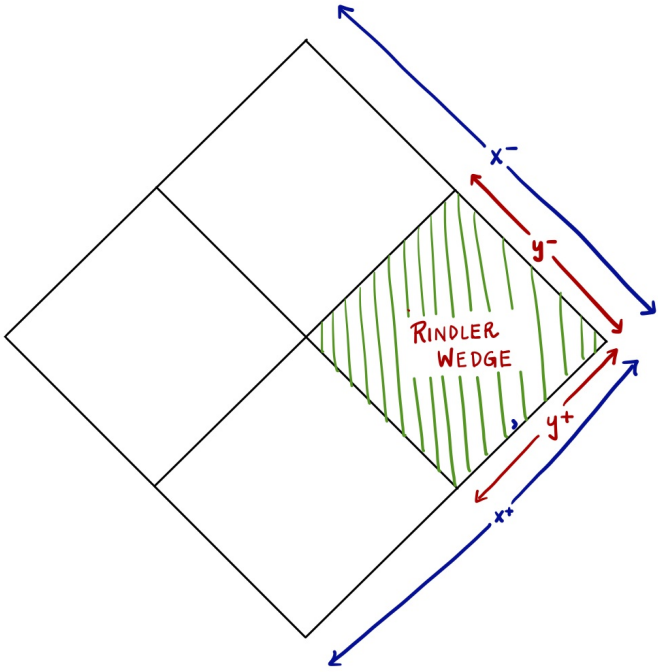


Figure 8.1: Rindler wedge

They are the natural coordinates for studying accelerated coordinates since

$$x^+x^- = -\frac{1}{a^2} \tag{8.95}$$

gives $y^+ - y^- = -2\ln a$, and we can take

$$y^- = \ln a + a\tau, \qquad y^+ = -\ln a + a\tau \tag{8.96}$$

as our trajectory.

Moreover, the Rindler time

$$t_{\text{Rindler}} = \frac{1}{2}(y_+ + y_-) = a\tau. \tag{8.97}$$

We next introduce Rindler modes

$$u_\omega = \frac{1}{\sqrt{2\omega}}\, \mathrm{e}^{i\omega y^-}. \tag{8.98}$$

We'd like to see what happens when we use a new coordinate system. We want to decompose Minkowski modes x^+, x^- in terms of Rindler modes but the Rindler coordinates cover only half of the Minkowski space. We can introduce another set of modes through the left Rindler wedge, which, when taken along with the right Rindler wedge, gives us a complete set of modes. We have the Minkowski annihilation operator

$$a = \alpha^T \tilde{a} + \beta^\dagger \tilde{a}^*, \tag{8.99}$$

where, more precisely,

$$\alpha^T \tilde{a} = \alpha_L^T \tilde{a}_L + \alpha_R^T \tilde{a}_R. \tag{8.100}$$

Then the Minkowski vacuum is such that

$$a_\omega \left|0_{\text{Mink}}\right\rangle = 0. \tag{8.101}$$

Similarly, we define the Rindler vacuum $\left|0_R\right\rangle = \left|0_{RR}\right\rangle \left|0_{RL}\right\rangle$ such that

$$\tilde{a}_R \left|0_{RR}\right\rangle = 0, \qquad \tilde{a}_L \left|0_{RL}\right\rangle = 0. \tag{8.102}$$

Recall that

$$\left|0_{\text{Mink}}\right\rangle = \mathrm{e}^{-\frac{1}{2}\tilde{a}^\dagger \beta^* \alpha^{-1} \tilde{a}^*} \left|0_{RR}\right\rangle \left|0_{RL}\right\rangle. \tag{8.103}$$

This works out to be

$$\left|0_{\text{Mink}}\right\rangle = N \prod_j \exp\left[\mathrm{e}^{-\pi\omega_j} a_{Rj}^\dagger a_{Lj}^\dagger\right] \left|0_{RR}\right\rangle \left|0_{RL}\right\rangle, \tag{8.104}$$

which is a sum of product states. Suppose that we're only interested in experiments which are out of causal contact with the left Rindler wedge. Then we don't need any information about anything that's going on in the left Hilbert space. So we can find the thermal density matrix by taking the trace over the unobservable left Hilbert space as follows.

$$\rho = \mathrm{Tr}_L \left|0_{\text{Mink}}\right\rangle \left\langle 0_{\text{Mink}}\right| = N^2 \prod_j \mathrm{Tr}_L \exp\left[\mathrm{e}^{-\pi\omega_j} a_{Rj}^\dagger a_{Lj}^\dagger\right] \left|0_{RR}\right\rangle \left|0_{LL}\right\rangle \left\langle 0_{LL}\right| \left\langle 0_{RR}\right| \exp\left[\mathrm{e}^{-\pi\omega_j} a_{Rj}^\dagger a_{Lj}^\dagger\right]. \tag{8.105}$$

This is an important calculation so we'll show the steps. For one mode,

$$\begin{aligned}
&\mathrm{Tr}_L \exp\left[\mathrm{e}^{-\pi\omega_j} a_{Rj}^\dagger a_{Lj}^\dagger\right] \left|0_{RR}\right\rangle \left|0_{LL}\right\rangle \left\langle 0_{LL}\right| \left\langle 0_{RR}\right| \exp\left[\mathrm{e}^{-\pi\omega_j} a_{Rj}^\dagger a_{Lj}^\dagger\right] \\
&= \mathrm{Tr}_L \sum_{n,n'} \frac{1}{n!} \left[\mathrm{e}^{-\pi\omega_j} a_{Rj}^\dagger a_{Lj}^\dagger\right]^n \left|0_{RR}\right\rangle \left|0_{LL}\right\rangle \left\langle 0_{LL}\right| \left\langle 0_{RR}\right| \frac{1}{n'!} \left[\mathrm{e}^{-\pi\omega_j} a_{Rj}^\dagger a_{Lj}^\dagger\right]^{n'} \\
&= \sum_{m_L} \left\langle m_L\right| \sum_{n,n'} \frac{1}{n!} \left[\mathrm{e}^{-\pi\omega} a_R^\dagger a_L^\dagger\right]^n \left|0_{RR}\right\rangle \left|0_{LL}\right\rangle \left\langle 0_{LL}\right| \left\langle 0_{RR}\right| \frac{1}{n'!} \left[\mathrm{e}^{-\pi\omega} a_R^\dagger a_L^\dagger\right]^{n'} \left|m_L\right\rangle \\
&= \sum_m \frac{1}{m!}\, \mathrm{e}^{-m\pi\omega} (a_R^\dagger)^m \left|0_{RR}\right\rangle \left\langle 0_{RR}\right|\, \mathrm{e}^{-m\pi\omega} (a_R)^m \\
&= \sum_m \mathrm{e}^{-2\pi m\omega} \left|m\right\rangle \left\langle m\right| \\
&= \sum_m \mathrm{e}^{-H/T_R} \left|m\right\rangle \left\langle m\right|,
\end{aligned} \tag{8.106}$$

where H is the Hamiltonian and $T_R = 1/2\pi$. So the conclusion is that we have a product of these modes and the density matrix is a thermal state at $T_R = 1/2\pi$. This agrees with our Unruh detector result of $T = a/2\pi$ since $\tau = t_R/a$, i.e., $\omega = a\omega_R$, giving $T = aT_R = a/2\pi$.

8.7 A note on curved spacetimes

We start with the generally covariant action

$$S = \int d^2x\sqrt{-g}g^{ab}\nabla_a f\nabla_b f = \int d^2x\partial_+ f\partial_- f, \tag{8.107}$$

where the last equality holds in the coordinates

$$ds^2 = -e^{2\rho}dx^+dx^-. \tag{8.108}$$

The action and the canonical commutation relations do not depend on ρ. However, as we will see, while the classical theory is ρ-independent, it appears in the quantum theory through an anomaly. To understand this, let us consider a situation where $R \neq 0$ inside a certain compact region of the spacetime and zero outside, as in Figure 8.2.

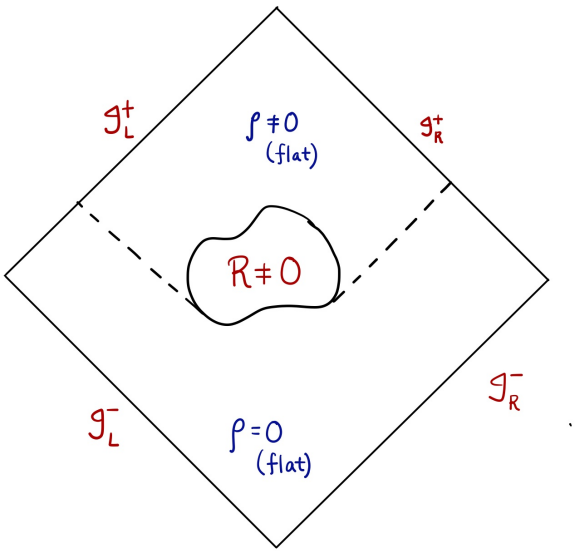


Figure 8.2: Flat and curved spacetimes

Now, since the spacetime is flat at early times, we can choose coordinates $x^\pm$ so that

$$ds^2 = -dx^+dx^-, \quad \text{when } x^- \text{ or } x^+ \to -\infty. \tag{8.109}$$

Further, since it is flat at late times, we can also choose coordinates $\tilde{x}^\pm$ so that

$$ds^2 = -d\tilde{x}^+d\tilde{x}^-, \quad \text{when } \tilde{x}^- \text{ or } \tilde{x}^+ \to -\infty. \tag{8.110}$$

However, since the spacetime is not globally flat, the two coordinates are not identical. We can see this as follows. In any set of coordinates,

$$R_{+-} = -2\partial_+\rho\partial_-\rho \tag{8.111}$$

implies that

$$\rho(x^+, x^-) = -\frac{1}{2}\int_{-\infty}^{x^+} dx^+ \int_{-\infty}^{x^-} dx^- R_{+-}, \tag{8.112}$$

where in the last equality, we have used the fact that $\rho(-\infty, x^-) = \rho(x^+, -\infty) = 0$. This equation tells us unless $R = 0$ everywhere we cannot have $\rho(+\infty, x^-) = \rho(x^+, +\infty) = 0$. Therefore, $x^\pm$ and $\tilde{x}^\pm$ are distinct coordinate systems. We focus below on x^- and $\tilde{x}^-$.

Let's define an in vacuum (defined at early times) using the modes

$$a_\omega = (f, u_\omega)\,, \qquad a_\omega \left|0_{\text{in}}\right\rangle = 0. \tag{8.113}$$

At late times, the inertial coordinate is $\tilde{x}^-$ and we define the out vacuum as

$$\tilde{a}_\omega = (f, \tilde{u}_\omega)\,, \qquad \tilde{a}_\omega \left|0_{\text{out}}\right\rangle = 0. \tag{8.114}$$

Since x^- and $\tilde{x}^-$ are distinct coordinates, $|0_{\text{in}}\rangle$ and $|0_{\text{out}}\rangle$ are distinct states. Now, consider an observer travelling along inertial worldlines in the past, i.e., along lines of changing x^- but fixed large x^+. Also suppose that the system is originally in the in vacuum, $|0_{\text{in}}\rangle$. In this case, the observer never encounters the curvature and is him/herself in flat spacetime. However, when the observer crosses the point P as shown, his trajectory is no longer inertial. To maintain the inertial path, he/she must change to the $\tilde{x}^-$ coordinates. In this new coordinate, the system is no longer in a vacuum with respect to the in states and the observer will observe particles. Thus, we see that there is particle production due to curvature of spacetime. This is an extremely interesting phenomena and one we believe is the cause of particles today!

8.8 Dilaton gravity continued

We can ask a similar question in the presence of black holes, i.e. since black holes produce curvature, they must also produce particles. We now ask, what kind of particle production occurs when black holes are present. We will now study this in the context of dilaton gravity. Recall the dilaton gravity action

$$S = \frac{1}{2\pi} \int d^2x \sqrt{-g} \left[\mathrm{e}^{-2\phi}(R + 4(\nabla\phi)^2 + 4\lambda^2) - \frac{1}{2}(\nabla f)^2 \right]. \tag{8.115}$$

This theory does have black holes with a Penrose diagram similar to Schwarzschild as in Figure 8.3.

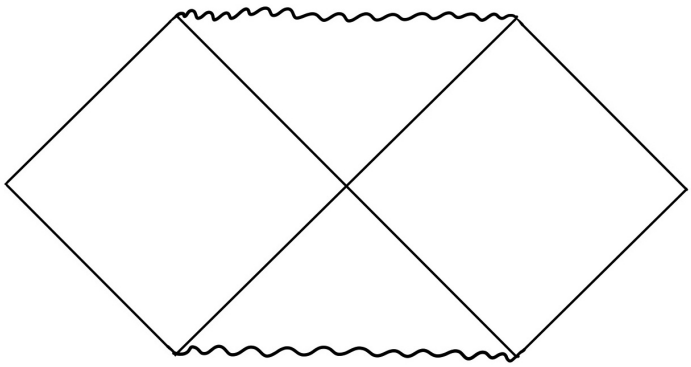


Figure 8.3: Schwarzschild black holes

Now, instead of studying eternal black holes, we study black holes formed by gravitational collapse as portrayed in Figure 8.4.

This is the (1+1)-dimensional analogue of of the Vaidya geometry that we've studied. We start from a Minkowski vacuum in the past and send in a shockwave with

$$T^f_{++} = \frac{1}{2}\partial_+ f \partial_+ f = M\delta(\sigma^+). \tag{8.116}$$

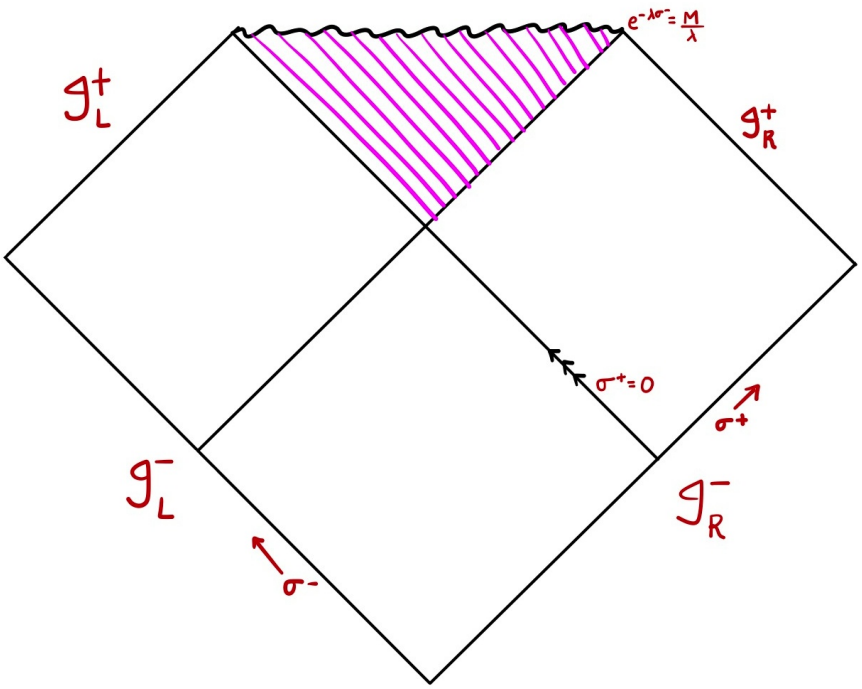


Figure 8.4: Black holes formed by gravitational collapse

We end up making a black hole. So what happens to the quantum field in this background? Is black hole formation accompanied by particle production? Luckily, we at least have a natural starting point of the quantum state for us to tackle the problem. Our situation at hand corresponds to the case where for $\sigma^+ < 0$, we have the flat space

$$ds^2 = -d\sigma^+ d\sigma^-. \tag{8.117}$$

And for $\sigma^+ > 0$,

$$ds^2 = \frac{-d\sigma^+ d\sigma^-}{1 - \frac{M}{\lambda}\, e^{\lambda\sigma^-} + \frac{M}{\lambda}\, e^{\lambda(\sigma^- - \sigma^+)}}. \tag{8.118}$$

If $\sigma^+ = 0$ along the shockwave, we see that the two cases are consistent with each other. On the other hand, at the singularity where $\sigma^+ = \infty$, $\sigma^- = \frac{1}{\lambda} \ln \frac{\lambda}{m}$, the metric blows up as expected. We also see that the metric is asymptotically flat for $\sigma^+ \to -\infty, \sigma^- \to -\infty$. Since we have Minkowski space there, we can talk about $|0_{\text{Mink}}\rangle$.

Now, on $\mathscr{I}_R^+$, i.e., when $\sigma^+ \to \infty$,

$$ds^2 = -d\sigma^+ d\tilde{\sigma}^-, \qquad d\tilde{\sigma}^- = \frac{d\sigma^-}{1 - \frac{M}{\lambda}\, e^{\lambda\sigma^-}}. \tag{8.119}$$

So we see that

$$\tilde{\sigma}^- = -\frac{1}{\lambda} \ln\left(e^{-\lambda\sigma^-} - \frac{M}{\lambda} \right). \tag{8.120}$$

Note that $\tilde{\sigma}^+, \tilde{\sigma}^- \in (-\infty, \infty)$ so the surface is geodesically complete. Since σ^+, σ^- are inertial coordinates, the incoming state with no incoming particles is just the vacuum in σ^+, σ^- coordinates. Therefore, a detector on a worldline with $\sigma^+ = \sigma^-$ observes nothing whereas inertial detectors with $\tilde{\sigma}^+ = \tilde{\sigma}^-$ do detect particles. We will set $\lambda \equiv 1$ for simplicity from now on.

Focusing on the right movers, we have the incoming and outgoing modes as

$$u_\omega = \frac{1}{\sqrt{2\omega}}\, e^{i\omega\sigma^-}, \qquad \tilde{u}_\omega = \frac{1}{\sqrt{2\omega}}\, e^{i\omega\tilde{\sigma}^-}. \tag{8.121}$$

Now, although $\tilde{\sigma}^-$ runs up to positive infinity, σ^- doesn't. So we introduce a second set of modes and denote them with $\hat{}$. We can then write the annihilation modes as

$$a = \alpha^T \tilde{a} + \beta^\dagger \tilde{a}^* + \hat{\alpha}^T \hat{a} + \hat{\beta}^\dagger \hat{a}^*. \tag{8.122}$$

Now we have a complete relationship between incoming and outgoing modes. And if we define

$$\bar{\alpha} \equiv \begin{pmatrix} \alpha \\ \hat{\alpha} \end{pmatrix}, \qquad \bar{\beta} \equiv \begin{pmatrix} \beta \\ \hat{\beta} \end{pmatrix}, \tag{8.123}$$

then we can write the incoming vacuum as

$$|0_{\text{in}}\rangle = N \, \mathrm{e}^{-\frac{1}{2}\bar{\alpha}^\dagger \bar{\beta}^* \bar{\alpha}^{-1} \bar{a}^*} \left|\tilde{0}\right\rangle |\bar{0}\rangle \, . \tag{8.124}$$

Then the outgoing density matrix

$$\rho_{\text{out}} = \mathrm{Tr}_{\hat{H}} \, |0_{\text{in}}\rangle \langle 0_{\text{in}}| \, . \tag{8.125}$$

As $\tilde{\sigma}^- \to -\infty$, we have a pure state $\left|\tilde{0}\right\rangle \left\langle\tilde{0}\right|$, and as $\tilde{\sigma}^+ \to +\infty$, we have a thermal state $\sum_i \mathrm{e}^{-E_i/T_H} \, |E_i\rangle \langle E_i|$ where the Hawking temperature $T_H = \lambda/2\pi$. Hawking himself, by doing an analogous but more complicated calculation in the four-dimensional theory, found that black holes have a mass-dependent temperature

$$T_H = \frac{\hbar}{8\pi G M}. \tag{8.126}$$

Defining black hole entropy S_{BH} through

$$\frac{dS_{\text{BH}}}{dM} = \frac{1}{T_H} = \frac{8\pi G_N M}{\hbar}, \tag{8.127}$$

we sees that

$$S_{\text{BH}} = \frac{A}{4G_N \hbar} = \frac{4\pi G_N M^2}{\hbar}. \tag{8.128}$$

It's useful to characterise the quantum states of the field f by its energy flow. So if we have the stress-energy tensor

$$T^f_{ab} = \frac{1}{2}\left(\partial_a f \partial_b f - \frac{1}{2} g_{ab} (\partial f)^2\right), \tag{8.129}$$

we want to find the expectation value $\langle 0_{\text{in}}| \, T^f_{ab} \, |0_{\text{in}}\rangle$. Classically, we have

$$T_{++} = M\delta(\sigma^+), \qquad T_{--} = \frac{1}{2}\partial_- f \partial_- f, \qquad T_{+-} = 0. \tag{8.130}$$

$T_{+-} = 0$ tells us the action for the f field is invariant under global scalings of the metric. In quantum mechanics, we need to introduce a cutoff to regulate ultraviolet divergences, so in principle, a quantum theory can depend on the scalings. To understand the resulting conformal (trace) anomaly, we consider the most general form

$$\langle 0_{\text{in}}| \, T^f_{ab} \, |0_{\text{in}}\rangle = \alpha g_{+-} + \beta R_{+-} + \gamma R R_{+-} + \cdots \tag{8.131}$$

But we are dealing with a theory with no dimensionful constants of the Lagrangian and only the β term is dimensionless. The other terms might appear in some regularisation scheme, but can be subtracted off. The best way to regulate it is to go to non-integer dimension $D = 2 + \epsilon$, which makes the loop corrections finite. Then the action becomes

$$S = -\frac{1}{2\pi} \int d^{2+\epsilon} x \, \sqrt{-g} \, g^{+-} \partial_+ f \partial_- f, \tag{8.132}$$

where $ds^2 = -\mathrm{e}^{2\rho} dx^+ dx^-$ gives

$$g_{+-} = -\frac{1}{2}\mathrm{e}^{2\rho}, \qquad \sqrt{-g} = \left(\frac{1}{2} e^{2\rho}\right)^{\frac{2+\epsilon}{2}}. \tag{8.133}$$

Hence,

$$S \simeq \frac{1}{4\pi} \int d^{2+\epsilon} x \, \partial_+ f \partial_- f (1 + \epsilon\rho). \tag{8.134}$$

And we find

$$T_{+-} = \frac{\epsilon}{4}\partial_+ f \partial_- f. \tag{8.135}$$

Now, we are not quantising ρ; it's only a background. So at fixed ρ,

$$\langle T_{+-} \rangle_\rho = -\frac{1}{12}\partial_+\partial_-\rho. \tag{8.136}$$

The generalisation of this result is as follows. With $R_{+-} = -2\partial_+\partial_-\rho$,

$$\langle T_{+-} \rangle = -\frac{c}{12}\partial_+\partial_-\rho = \frac{c}{24}R_{+-}, \tag{8.137}$$

where c is the *central charge*. For a free scalar field, $c = 1$, and for n free scalar fields, $c = n$. This is the famous trace anomaly formula for a two-dimensional quantum field theory on a curved space. We'll see that this anomaly corresponds to Hawking radiation. Eventually we will have to solve

$$G_{ab} = \#T_{ab}, \tag{8.138}$$

where by G_{ab} we now mean the variation of the gravitational part of the action given above. For there to be a well-defined classical limit we need

$$\nabla^a T_{ab} = 0. \tag{8.139}$$

Using this, we get

$$\partial_+ \langle T_{--} \rangle + \partial_- \langle T_{+-} \rangle - \Gamma^-_{--} \langle T_{+-} \rangle = \nabla^a \langle T_{a-} \rangle = 0. \tag{8.140}$$

And substituting in various expressions, we have

$$\partial_+ \langle T_{--} \rangle + \partial_- \left(-\frac{1}{12}\partial_+\partial_-\rho \right) - 2\partial_-\rho \left(-\frac{1}{12}\partial_+\partial_-\rho \right) = 0, \tag{8.141}$$

which gives

$$\langle T_{--} \rangle = -\frac{1}{12}\left((\partial_-\rho)^2 - \partial_-^2\rho \right) + T_{--}^{\text{classical}} + t(x^-), \tag{8.142}$$

where $t(x^-)$ is related to normal ordering ambiguities. But in our case, $T_{--}^{\text{classical}} = 0$. Further, the condition

$$\langle 0_{\text{in}} | \, T_{--}(\sigma^+ = -\infty, \sigma^-) \, | 0_{\text{in}} \rangle = 0 \tag{8.143}$$

fixes $t(x^-) = 0$. Hence, rewriting the first term, we have

$$\langle T_{--} \rangle = -\frac{1}{12}\, \mathrm{e}^{\rho}\partial_-^2 \, \mathrm{e}^{-\rho}. \tag{8.144}$$

Going back to the dilaton black hole, we have for $\sigma \to \infty$, $\mathrm{e}^{-\rho} \to \sqrt{1 - M\,\mathrm{e}^{\sigma^-}}$. In this limit then,

$$\langle T_{--} \rangle = \frac{M\,\mathrm{e}^{\sigma^-}(2 - M\,\mathrm{e}^{\sigma^-})}{48(1 - M\,\mathrm{e}^{\sigma^-})^2}. \tag{8.145}$$

We've been using σ^- coordinates so far. In inertial coordinates,

$$\left\langle \tilde{T}_{--} \right\rangle = \left(\frac{\partial \sigma^-}{\partial \tilde{\sigma}^-} \right)^2 \langle T_{--} \rangle = \frac{M}{48} \frac{1}{M + \mathrm{e}^{-\tilde{\sigma}^-}} \left(2 - \frac{M}{M + \mathrm{e}^{-\tilde{\sigma}^-}} \right). \tag{8.146}$$

This is an explicit formula with interesting properties. As $\tilde{\sigma}^- \to -\infty$, we end up with

$$\left\langle \tilde{T}_{--} \right\rangle = \frac{M}{24}\, \mathrm{e}^{\tilde{\sigma}^-}, \tag{8.147}$$

i.e., in the far past, there is no energy flux. And as $\tilde{\sigma}^- \to \infty$, restoring λ, we get

$$\left\langle \tilde{T}_{--} \right\rangle = \frac{\lambda^2}{48}. \tag{8.148}$$

This is nothing but the energy density of thermal radiation at the temperature $T = \lambda/2\pi$. So black holes emit with a purely thermal spectrum.

Now, because black holes emit energy and conservation laws hold, we conclude that they must shrink. We now discuss the process of black hole shrinking in a systematic way. This is referred to as *back reaction*, in the sense that the stress-energy tensor back reacts and sources the metric. We start by varying the action with respect to g^{++}.

$$\begin{aligned} 2\pi \frac{\delta S}{\delta g^{++}} &= \mathrm{e}^{-2\phi}(2\partial_+^2\phi - 4\partial_+\rho\partial_+\phi) \\ &= \frac{1}{2}\partial_+ f \partial_+ f - \frac{\hbar}{12}\mathrm{e}^{\rho}\partial_+^2\,\mathrm{e}^{-\rho}, \end{aligned} \tag{8.149}$$

where the first term is the classical $T_{++}^{\text{classical}}$ and the second term, i.e., the quantum correction we calculated earlier, gives us the back reaction. Note that we restore the factor of $\hbar$ for clarity. Now, it doesn't make sense to take a classical equation of motion, add a one-loop quantum correction, and solve the loop-corrected new equation. This is because there's no systematic small parameter involved. In our case, the small perturbative correction of order $\hbar$ isn't enough for seeing black hole evaporation, which is a huge effect. So how can we find a small parameter to expand in? The idea is to introduce N fields and expand around the large-N limit. We do so by modifying the action to

$$S = \frac{1}{2\pi}\int d^2x\sqrt{-g}\left(N\,\mathrm{e}^{-2\phi}(R + 4(\nabla\phi)^2 + 4) - \frac{1}{2}\sum_{i=1}^{N}(\nabla f_i)^2\right). \tag{8.150}$$

In terms of Feynman diagrams, this corresponds to tree diagrams involving ρ and ϕ and loop corrections involving f. The previous equation picks up a factor of N and becomes

$$\begin{aligned} 2\pi \frac{\delta S}{\delta g^{++}} &= N\,\mathrm{e}^{-2\phi}(2\partial_+^2\phi - 4\partial_+\rho\partial_+\phi) \\ &= \frac{1}{2}\sum_{i=1}^{N}\partial_+ f_i \partial_+ f_i - \frac{\hbar}{12}N\,\mathrm{e}^{\rho}\partial_+^2\,\mathrm{e}^{-\rho}, \end{aligned} \tag{8.151}$$

Solving this gives us the Penrose in Figure 8.5. The red lines are lines of constant r. Inside black holes, lines of constant r are spacelike, and outside, timelike. Starting from the vacuum, we again send a shockwave, which makes a black hole, and we get a singularity. We have an event horizon which no longer coincides with the apparent horizon. (Recall that the apparent horizon is the location where lines of constant r turn from being timelike to being spacelike. It's what looks like the event horizon of a black hole locally.)

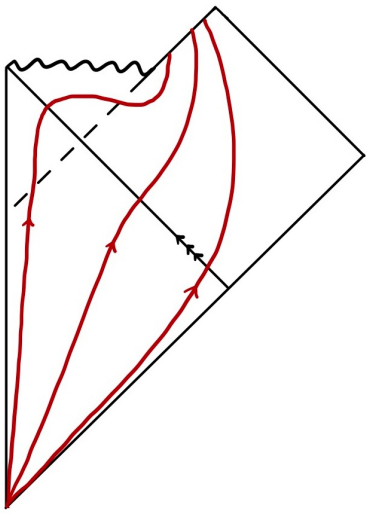


Figure 8.5: Black hole evaporation

We see that towards the left of the diagram (smaller values of r), lines of constant r start off being timelike, cross the shockwave, become spacelike, and finally return to being timelike again. We see that the apparent horizon shrinks since, as we move along it towards the future, it crosses smaller and smaller values of r. This is expected as back reaction causes black holes to shrink. Note that T_{++} doesn't have to be positive here; it is in fact negative along the apparent horizon and we can think of it as a flux of negative energy entering the black hole. This calculation substantiates the picture of Hawking radiation in which we have a pair production of a positive and a negative energy particle outside and the negative energy one enters the black hole and the process keeps repeating.

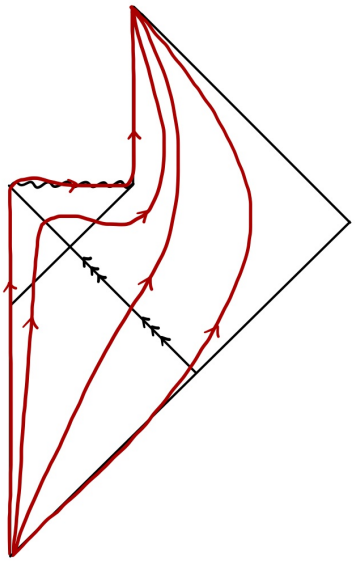


Figure 8.6: Black hole evaporation and beyond

Now, space becomes flat at the null surface where the red lines of constant r end, so it is natural to patch our Minkowski Penrose diagram there. This is shown in Figure 8.6 and leads to our upcoming discussion on the *information paradox*.

So in our two-dimensional context, we have obtained this picture as a solution of our well-defined differential equation. An interesting fact about this two-dimensional model is that when we quantise ρ and ϕ, the causal structure doesn't fluctuate. So the causal relationship between the points is not subject to the uncertainty principle.

9 Information paradox

Now we make an attempt to understand what the future of a singularity holds, i.e., what goes beyond the null surface in Figure 8.5. There are four main proposals.

- Information is lost;
- There exist long-lived remnants;
- Information is returned to $\mathscr{I}^+$;
- None of the above.

9.1 Information is lost

We found earlier that for a black hole of mass M, its temperature $T \sim 1/M$. So the power

$$P = \frac{dM}{dt} \sim T^4 A \sim T^2 \sim \frac{1}{M^2}, \tag{9.1}$$

which we can solve to obtain

$$M(t) \sim (t - t_f)^{1/3}. \tag{9.2}$$

Hence, the evaporation time

$$t_{\text{evap}} \sim M^3. \tag{9.3}$$

We can also estimate entropy production. It does depend on quantities such as the number of species of particles, but parametrically,

$$S_{\text{produced}} \sim M^2 > \frac{A}{4}. \tag{9.4}$$

So this is not an adiabatic process. If we now graph entropy S at $\mathscr{I}^+$ verses time, then we have Figure 9.1 (solid line). S rises to order M^2 in the amount of time of order M^3 then simply remains unchanged.

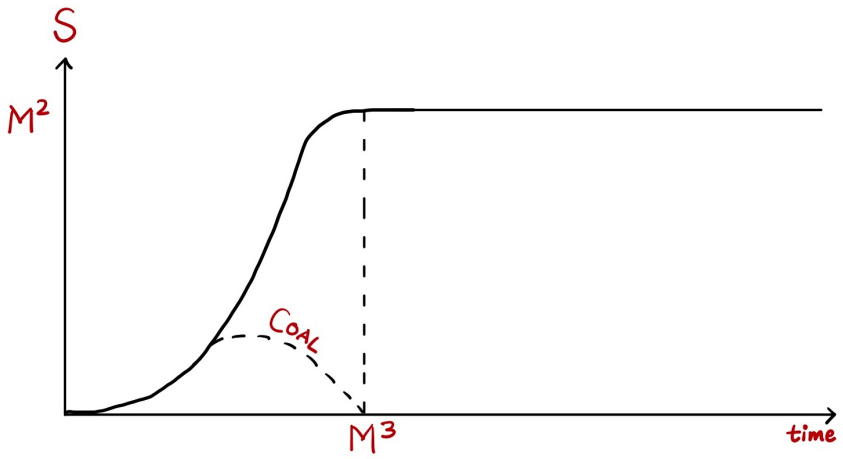


Figure 9.1: Entropy change of black hole evaporation verses coal burning

Let's contrast the situation with an idealised lump of coal (dotted line) that starts off having $S = 0$. Let's burn it and consider the radiation. Initially it would look thermal and the entropy would rise according to some thermal curve. But if we wait long enough, we see that emissions at later times correlate with emissions at earlier times, and the entropy returns to zero. Note that the lump of coal is initially in a pure state, but if we know the initial density matrix exactly, then unitary time evolutions in quantum mechanics imply that the final state is unique.

Recall that the Schrödinger equation says

$$|\psi(t)\rangle = \mathrm{e}^{iHt} |\psi(0)\rangle = U(t) |\psi(0)\rangle . \tag{9.5}$$

So the density matrix

$$\rho(t) = U(t)\rho(0)U^\dagger(t), \tag{9.6}$$

which implies that

$$S = \mathrm{Tr}\left[\rho(t) \ln \rho(t)\right] \tag{9.7}$$

is invariant under time evolution, i.e., entropy is conserved. But here, entropy increases. So instead of unitary matrices, we can use what's called the *dollar matrix* \$ such that

$$\$\rho_{\text{in}} = \rho_{\text{out}}. \tag{9.8}$$

However, general transformation matrices of this kind don't preserve $\mathrm{Tr}\rho = 1$. Further, symmetries no longer imply conservation laws: we only get a conservation of the average of some quantity. So we don't have a good alternative to unitary evolutions in quantum mechanics.

In summary...

- Advantage: It follows from the semiclassical calculations.
- Disadvantage: It violates unitarity and energy conservation.

9.2 Long-lived remnants

In this proposition, we have some Planck-sized object, a *remnant*, that can live for a very long time and will send the information back out. Well, how long does it live? First of all, if they live forever, then we can say that the universe contains an infinite number of Planckian objects and they have an infinite number of states. The information isn't lost; it's simply in the Planckian objects.

The problem of this is that first, it violates the generalised second law because these Planckian objects have an infinite amount of entropy and thus an infinite number of microstates. If we take one of these objects and throw it into the black hole, then the entropy of the black hole suddenly becomes infinite and cannot be related to the area so simply anymore. Second, quantum mechanically, the pair production rate (of particles of mass M) is associated with a Boltzmann suppression factor, which is e^{-M}. This is a small number, but when we multiply it by the an infinite number of species of remnants, we get an infinity, and the suppression factor serves no purpose.

Now we discuss the case of non-eternally existing remnants. Suppose that they disappear after an amount of time of order T_{rem}. If we have a conserved amount of energy, we can ask how much information can be encoded. In fact, it depends on how big of a box we have, in the sense that if our energy is distributed among photons in a box, then if the box is infinitely large then we have an infinite amount of energy. So there is no limit on the amount of information we can encode in a fixed amount of energy. Hence, when the remnant disappears, all the energy, which we take to be of order one in Planck units, is in some radially moving particles confined in a box of radius of order T_{rem}.

The entropy of the one-dimensional gas then is $S = \sqrt{EL} = \sqrt{T_{\mathrm{rem}}}$, which tells us that $T_{\mathrm{rem}} \sim M^4$, which is much longer than $t_{\mathrm{evap}} \sim M^3$. So the entropy against time plot is now as in Figure 9.2.

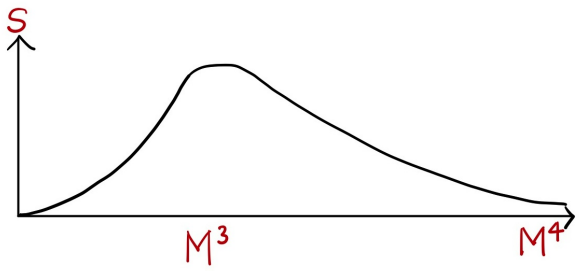


Figure 9.2: Entropy change in the remnant scenario

There is a variation, the "non-local remnant" scenario, that doesn't get discussed much. This is where we say that information is spread across the null surface. Then we have $T_{\mathrm{rem}} \sim M^{7/2}$ instead. Of course, it is still greater than the evaporation time.

In summary...

- Advantage: There is no need for an information escape mechanism.

- Disadvantage: Remnants may need to contain an infinite amount of entropy hence violating the generalised second law.

9.3 Information is returned

This is perhaps the most popular proposal since it's rather intuitive. It says that half way through the evaporation process, something goes wrong with Hawking's calculations. The entropy against time curve turns over and heads back to zero as in the case of the coal. Here unitarity in quantum mechanics is preserved.

So what went wrong? Perhaps the $1/N$ expansion is invalid: it could have a zero radius of convergence; perhaps the model is simply not good for the four-dimensional case; perhaps we should've have assumed causality or locality; perhaps it's invalid to use here an effective field theory, i.e., the approach where we approximately describe things by semiclassical gravity.

In summary...

- Advantage: Unitarity in quantum mechanics is preserved.
- Disadvantage: There is a large deviation from semiclassical results in the two-dimensional calculation.

9.4 General comments

There are two parts to the information paradox: the information storage problem, and the information retrieval problem.

If the information ends up coming back out, it certainly has to be stored. In string theory, as we'll discuss next semester, one can solve the information storage problem: By a counting of microstates we can check explicitly for a wide array of black holes that $S = A/4$, in agreement with the proposal that information is returned. This calculation was a precursor to the AdS/CFT correspondence. However, since AdS/CFT only requires information to be eventually returned, it is compatible with the remnant proposal. Note that this correspondence between things that live on the boundary and the bulk only applies well to certain situations. If we had a small black hole with radius smaller than the AdS radius, then we don't understand the boundary state hence have no dictionary to link the two.

Finally, the *firewall* has been discussed a lot recently. People often take locality, causality, unitarity, low-energy effective field theory and the equivalence principle, keep four of them and give up one, and conclude that since the four holds, the fifth doesn't apply. In the firewall story, information returns, but there is an infinite energy density at the horizon: a firewall. In this scenario, the equivalence principle is violated. No one has yet suggested a dynamical mechanism to allow firewall.

To us, as physicists, computing how entropy evolves with time is the key to solving the information paradox.