Lectures on black hole information and spacetime wormholes

Henry Maxfield*

February 1, 2021

Abstract

Lectures.

Contents

*hmaxfield@physics.ucsb.edu

2 Quantum black holes

2.1 The geometry of horizons

We begin with a discussion of the geometry of black hole horizons in classical general relativity, pointing out the key qualitative features for the physics of Hawking radiation.

After a black hole forms, if it is isolated (not rapidly accreting matter, for example), it will rapidly settle into a stationary state. Stationary solutions to Einstein's equations are typically characterised only by conserved quantities: the mass, the angular momentum, and the charge under any gauge fields ('black holes have no hair' [1]). Assuming spherical symmetry for simplicity (no angular momentum), we can write any stationary metric using ingoing coordinates r, v as

$$ds^2 = -f(r)dv^2 + 2drdv + (\text{transverse}) \tag{2.1}$$

for some function $f(r)$. The 'transverse' piece describes the metric on the orbits of spherical symmetry at constant v and r, and is given by the metric on a sphere of some radius (a function of r). We will henceforth drop this spherical part of the metric, concentrating on the two-dimensional geometry at constant angle on the sphere.

The coordinate v is an ingoing time, constant on radial null geodesics falling into the black hole. In general, we will define v to match the proper time of a distant observer, at fixed but large r (see the discussion of asymptotics in a moment). For a black hole formed from collapse, (2.1) will be a good approximation for ingoing times v sufficiently long after the black hole has formed.

A black hole corresponds to a region $r < r_+$ from which light cannot escape. From (2.1), outgoing null geodesics satisfy $\frac{dr}{dv} = \frac{1}{2}f(r)$, so they will fail to escape to infinity if $f(r) < 0$. The event horizon of the black hole is at $r = r_h$, where r_h is the largest r satisfying $f(r_h) = 0$. Generically, f will vanish linearly at the horizon, so the surface gravity $\kappa = \frac{1}{2}f'(r_h)$ will be positive, and near to the horizon we can approximate the metric with $f(r) \sim 2\kappa(r - r_h)$.

This immediately tells us something important: near to the horizon, the outgoing null geodesics diverge exponentially, satisfying $r - r_h \sim Ae^{\kappa v}$. This simple fact is at the root of much of the interesting physics of black holes.

More generally, outgoing null geodesics outside the horizon are at constant values of the outgoing time u, defined analogously to v with respect to the proper time of a distant observer. Concretely, it is given for the static metric (2.1) by $u = v - 2r_*(r)$, where $r_*'(r) = \frac{1}{f(r)}$ (often called the 'tortoise coordinate'). In terms of these lightcone coordinates u, v, we can then write the metric in the exterior of the black hole as

$$ds^2 = -f(r)dudv, \tag{2.2}$$

where r is defined implicitly via $r_*(r) = \frac{v-u}{2}$. Close to the horizon, the exponential divergence of outgoing null geodesics tells us that

$$r - r_h \sim \frac{A}{2\kappa} e^{\kappa(v-u)} \tag{2.3}$$

for some A (the factor of 2κ inserted for later convenience). In particular, the horizon $r = r_h$ itself lies at $u \to \infty$, so our coordinates u, v cover only the region $r > r_h$.

At this point, let us be more specific about the asymptotic metric far from the black hole, and the definitions of u and v. We will be interested in two possibilities. First, for asymptotically flat spacetimes, we have $f(r) \to 1$, so

$$\begin{aligned} ds^2 \sim -dv^2 + 2drdv &= -dudv, \\ r \sim \tfrac{v-u}{2} &\to \infty. \end{aligned} \qquad \text{(asymptotically flat)} \tag{2.4}$$

In asymptotically AdS spacetimes, we have $f(r) \sim \frac{r^2}{\ell^2}$, and we define u, v so that

$$\begin{aligned} ds^2 \sim -\frac{r^2}{\ell^2}dv^2 + 2drdv &= -\left(\tfrac{2\ell}{u-v}\right)^2 dudv, \\ r \sim \tfrac{2\ell^2}{u-v} &\to \infty. \end{aligned} \qquad \text{(asymptotically AdS)} \tag{2.5}$$

In particular, the asymptotic boundary is timelike, at $u = v = t$, where t is a 'renormalised' proper time (in AdS/CFT, t would correspond to the time in the boundary dual theory).

Using our near-horizon coordinate change (2.3), we can write our near-horizon metric in lightcone coordinates as

$$ds^2 \sim -Ae^{\kappa(v-u)}dudv \qquad (r \to r_h, \quad u \gg \kappa^{-1}). \tag{2.6}$$

But, like any metric in a region much smaller than the curvature scale, this is just two-dimensional Minkowski space in disguise. To see this very explicitly, define new 'Kruskal' coordinates U, V, which in the near-horizon region $u \to \infty$ behave as

$$U \sim -\tfrac{1}{\kappa}e^{-\kappa u}, \quad V \sim \tfrac{A}{\kappa}e^{\kappa v} \tag{2.7}$$

$$\implies ds^2 \sim -dUdV. \tag{2.8}$$

The key result here is not that the metric is locally flat (since that is always true), but rather the exponential relationship between the flat spacetime coordinates U, V in the vicinity of the horizon and the coordinates u, v adapted to the asymptotic region far from the black hole. The region of the near horizon visible to an asymptotic observer for an unbounded range of outgoing times u is exponentially compressed to a finite region. This is essentially the same point that we emphasised before: outgoing geodesics diverge exponentially near the event horizon.

This fact is responsible for the phenomenon of Hawking radiation. Namely, for any state of quantum fields that is regular at the event horizon, an observer far from the black hole will see a flux of energy radiating from the black hole.

To get some intuition for this, let's see how time evolution acts. On the coordinates u, v adapted to the asymptotic region, time translation by a time t simply acts by addition of t. But in terms of our coordinates U, V adapted to the near-horizon region, this becomes a boost:

$$u \mapsto u + t, \quad v \mapsto v + t \tag{2.9}$$

$$U \mapsto e^{-\kappa t} U, \quad V \mapsto e^{+\kappa t} V. \tag{2.10}$$

So the Hamiltonian $H = \partial_u + \partial_v$ for matter fields on the black hole background acts near the horizon like a boost generator $K = V\partial_V - U\partial_U$, with proportionality constant given by the surface gravity: $H \sim \kappa K$.

Now, the vacuum state of any relativistic QFT on a half of two-dimensional Minkowski space times any transverse manifold looks like a thermal state with respect to the boost generator: $\rho_{\mathrm{half-space}} \propto e^{-2\pi K}$. The underlying reason is that the boost generator becomes a rotation in Euclidean signature, so we can think of $\rho_{\mathrm{half-space}}$ as enacting a rotation through angle 2π. More precisely, suppose we would like to compute the expectation value $\mathrm{Tr}(\mathcal{O} e^{-2\pi K})$ of some local operator $\mathcal{O}$ acting on the half space. The operator $e^{-2\pi K}$ rotates the half-space by an angle 2π around its boundary, sweeping out the whole of Euclidean space, so this expectation value is computed by the path integral on flat Euclidean space. But (up to normalisation) this is precisely the vacuum expectation value $\langle 0|\mathcal{O}|0\rangle$.

This is relevant because any nonsingular state in QFT looks like the vacuum state at short distances, so the state of fields just outside the horizon looks thermal in terms of the local boost generator. Identifying the near-horizon boost with the asymptotic Hamiltonian, we see that the vacuum in the near-horizon region appears to the asymptotic observer like a thermal state at inverse temperature $\beta = \frac{2\pi}{\kappa}$, giving us the Hawking temperature

$$T_H = \frac{\kappa}{2\pi}. \tag{2.11}$$

This heuristic argument suggests that for any state of quantum fields that looks like the vacuum near to the horizon, it will be interpreted as thermal to the asymptotic observer.

Indeed, this conclusion is is borne out by more careful calculations, and results in the physical effect of Hawking radiation: a flux of thermal radiation escaping to infinity. To illustrate this, we now restrict to a class of theories where the argument becomes extremely simple.

2.2 Conformal matter: energy dynamics

As a simple illustrative model, we will take the matter to be described by a conformally invariant theory in two dimensions. If you would like a concrete model, take some number of free massless bosons or fermions, though everything we say will also apply to interacting matter. We do not require any background in two-dimensional conformal field theory, and will introduce the facts we need along the way (though feel free to consult [2, 3, 4] for a more systematic and detailed account).

The dynamics of energy-momentum is extremely simple in these theories, since it is determined completely by the symmetries. The energy-momentum tensor has three independent components in two dimensions, and we have two equations constraining their evolution from conservation. Adding conformal invariance provides us with a third equation fixing the trace, so we can have a complete set of equations governing its evolution.

The defining property of conformal field theories is that there is a symmetry under changing the spacetime metric by a Weyl transformation: $g \mapsto \Omega^2 g$, where Ω is a local rescaling, which can be any function of spacetime. This means that we change the notion of distance, but leave angles invariant. In Lorentzian spacetime, this means that the causal structure (the space of null vectors) remains invariant.

The energy-momentum tensor is defined (classically) through the variation of the CFT action I_{CFT} as

$$T^{ab}(x) = -\frac{2}{\sqrt{-g}} \frac{\delta I_{\mathrm{CFT}}}{\delta g_{ab}(x)} \tag{2.12}$$

(and in the quantum theory by metric variations of correlation functions).[1] In particular, Weyl transformations (with $\delta g_{ab} \propto g_{ab}$) are generated by the

[1]Note that this is the usual definition in most contexts (in particular, with this nor-

trace $\operatorname{Tr} T$. To leave the correlation functions invariant, this means that $\operatorname{Tr} T$ should vanish, and indeed it does in flat spacetime.

However, things are slightly more complicated in curved spacetime, since the conformal symmetry is anomalous: that is, correlation functions are not left invariant, but instead change in a predictable way. The anomaly is determined only by an anomaly coefficient, in this case the 'central charge' c, some fixed positive constant (for a unitary theory). The upshot is that the trace of T depends on the spacetime curvature:

$$\operatorname{Tr} T = \frac{c}{24\pi}\mathcal{R}, \tag{2.13}$$

where $\mathcal{R}$ is the Ricci scalar. For a full account of this anomaly, see §3.4 of [3] for example. Note that this is an operator equation, meaning that $\operatorname{Tr} T$ is proportional to the identity operator in CFTs (except for contact terms where $\operatorname{Tr} T$ becomes coincident with other operator insertions), and in particular it holds in any state of the theory.

Writing a general metric in terms of lightcone coordinates u, v as

$$ds^2 = -e^{2\omega(u,v)} du dv, \tag{2.14}$$

the equation for the trace (2.13) becomes

$$T_{uv} = -\frac{c}{12\pi}\partial_u\partial_v\omega. \tag{2.15}$$

Using this, the conservation equations $\nabla_a T^a{}_b = 0$ become equations for $\partial_v T_{uu}$ and $\partial_u T_{vv}$ with solutions

$$T_{uu} = \frac{c}{12\pi}\left(\partial_u^2\omega - (\partial_u\omega)^2\right) + F_u(u), \tag{2.16}$$

$$T_{vv} = \frac{c}{12\pi}\left(\partial_v^2\omega - (\partial_v\omega)^2\right) + F_v(v), \tag{2.17}$$

where F_u, F_v are functions only of u, v respectively.

Now, in either asymptotically flat or asymptotically AdS spacetimes, the inhomogeneous terms (depending on u, v) vanish at large r, so F_u and F_v tell us the asymptotic value of the stress-energy:

$$\begin{aligned} F_u(u) &= \text{asymptotic outgoing energy flux}, \\ F_v(v) &= \text{asymptotic incoming energy flux}. \end{aligned} \tag{2.18}$$

malisation T_{00} is interpreted as the local energy density), but differs from a common 2D CFT convention by a factor of -2π.

Exercise 1. *Check these results. For what spacetimes can we ignore the inhomogeneous terms, so T_{uu} is just a function of u and T_{vv} is a function of v? (Hint: compare the conservation equations to the derivatives of the curvature.)*

In the quantum theory, the trace equation (2.13) and the conservation equations apply to correlation functions when stress-energy tensor is inserted away from any other operators or matter sources. In particular, our results (2.16), (2.17) and (2.15) hold for the one-point function $\langle T_{ab}\rangle$ in any state.

2.3 Hawking radiation

Let's now apply this result to conformal matter in a fixed static black hole background, looking at the stress-energy in the near-horizon region. From (2.6) we have $\omega \sim \frac{\kappa}{2}(v-u)$, so

$$T_{uu} \sim -\frac{c}{48\pi}\kappa^2 + F_u(u), \tag{2.19}$$

$$T_{vv} \sim -\frac{c}{48\pi}\kappa^2 + F_v(v). \tag{2.20}$$

Now, u is not a good coordinate at the horizon (where it goes to infinity). To get an idea of the physical energy density near the horizon, we can use instead the r coordinate. We then have

$$T_{rr} \sim \frac{F_u(u) - \frac{c}{48\pi}\kappa^2}{\kappa^2(r-r_h)^2}. \tag{2.21}$$

But for any quantum state which is nonsingular at the horizon, T_{rr} must remain finite there. This requires that F_u must approach $\frac{c}{48\pi}\kappa^2$ as $u \to \infty$ to cancel the inhomogeneous term in (2.19), plus terms that go to zero at least as fast as $(r-r_h)^2$. We therefore have

$$F_u(u) = 2\pi \frac{c}{24} T_H^2 + O(e^{-2\kappa u}), \tag{2.22}$$

where we used (2.11) to rewrite the surface gravity κ in terms of the Hawking temperature T_H. This is precisely the expectation value of T_{uu} for outgoing thermal radiation at temperature T_H. The conclusion is that for any state that is nonsingular at the horizon, asymptotic observer will see a thermal flux of radiation at the Hawking temperature!

The exponential correction to (2.22) at late times corresponds to the energy density at the horizon in the immediate aftermath of the formation of the black hole. The coefficient of the correction term is proportional to

$e^{2\kappa v_0}$ if the black hole is formed at time v_0, times the energy density at the horizon at that time. This exponentially decaying transient energy can be thought of as a simple example of a quasinormal mode, which describes the black hole's approach to equilibrium.

Now we look at the near-horizon value of T_{vv}, which gives the flux of energy falling through the horizon into the black hole. From (2.20), this flux is given by the energy coming in from infinity, minus a constant which matches the outgoing flux from Hawking radiation. This is in accord with conservation of energy: the energy absorbed by the black hole is given by the incoming energy flux minus the outgoing energy flux as measured at infinity.

If there is no energy incoming from infinity ($F_v = 0$), the negative flux $T_{vv} = -\frac{c}{48\pi}\kappa^2$ of energy across the horizon will provide a source for Einstein's equations, altering the metric. Specifically, it will cause the black hole to shrink as it loses energy via Hawking radiation. This is the process of black hole evaporation, which we will describe explicitly in the next section, after we introduce a model for the gravitational dynamics.

Alternatively, we can keep the black hole in equilibrium by producing an ingoing flux of energy $F_v = +\frac{c}{48\pi}\kappa^2$ from infinity, balancing the outgoing Hawking radiation. One way to do this is by 'putting the black hole in a box': rather than letting the Hawking radiation escape, we reflect it back so that the black hole attains equilibrium with its own Hawking radiation. This is the usual situation in asymptotically AdS spacetimes: energy-conserving boundary conditions require $F_u(t) = F_v(t)$, so black holes do not evaporate.[2]

2.4 A reservoir for Hawking radiation.

Our simple model will describe black holes in AdS_2, but would like them to be able to evaporate. To this end, we will not use the usual reflecting boundary conditions, but instead couple the matter theory to an auxiliary 'reservoir' system at the boundary. This reservoir will simply be described by the same matter CFT, living on a fixed flat spacetime, given by half of Minkowski space $ds^2 = -du dv$, with $v > u$. The boundary of this space at $u = v = t$ will be joined to the boundary of our dynamical asymptotically AdS spacetime, with asymptotic metric $ds^2 \sim -\left(\frac{2\ell}{u-v}\right)^2 du dv$ for $u > v$ as

[2]This need not apply to very small black holes in more than two spacetime dimensions. The energy takes a time of order the AdS time ℓ to travel to infinity and reflect back, which can be longer than the black hole's lifetime if it is sufficiently small.

in (2.5).

$$ds^2 = -e^{2\omega(u,v)} dudv \tag{2.23}$$

$$\begin{aligned} &\text{Reservoir:} && v > u, && \text{fixed background } \omega = 0. \\ &\text{Black hole:} && u > v, && \omega \text{ dynamical.} \end{aligned} \tag{2.24}$$

This metric is not continuous at the boundary where we couple the two systems: indeed, it is singular as we approach it from the AdS side! But because our matter is conformally invariant, we may first define the theory on any smooth, flat metric related by a Weyl transformation, $ds^2 = -e^{2\omega'(u,v)}dudv$. Correlation functions in the physical metric are then determined by applying an appropriate Weyl transformation (including the anomaly (2.13)). Indeed, we have already seen one example of this Weyl transformation in (2.16), (2.17), where the inhomogeneous terms are the contribution of the Weyl anomaly transforming from the flat metric $-dudv$ to the physical black hole metric $-e^{2\omega}dudv$. This allows matter excitations to pass freely between the gravitational black hole spacetime and the reservoir. In particular, Hawking radiation can escape from the black hole into the reservoir, where it will propagate freely away to future null infinity ($v \to \infty$ at fixed u).

2.5 Black hole thermodynamics

We saw above that a black hole will remain stationary if we put it in contact with a system of temperature T_H. But, essentially by definition (the zeroth law of thermodynamics), this means that the black hole itself has temperature T_H! We can then use the first law (along with the energy of the black hole, determined from the metric far away) to assign an entropy,

$$S_{\mathrm{BH}}(E) = \int \frac{dE}{T_H(E)}. \tag{2.25}$$

This is the Bekenstein-Hawking entropy, and (as we will see later) it can always be described in terms of the geometry of the event horizon. For Einstein gravity, it is given by the famous formula

$$S_{\mathrm{BH}} = \frac{A}{4G_N}, \tag{2.26}$$

where A is the area of the event horizon.

Deriving the entropy from the temperature is somewhat backwards from a historical point of view. Bekenstein proposed that the area of a black hole should be assigned an entropy, by considerations of the second law [5].

Hawking set out to prove him wrong by showing that black holes didn't have a temperature, so couldn't possibly be assigned an entropy: of course, he instead discovered that phenomenon of Hawking radiation we just described above [6].

We usually think of this sort of equilibrium thermodynamics as arising from a quantum statistical description: namely, an entropy is the logarithm of the number of quantum states in a small window of energy. Does the Bekenstein-Hawking entropy count states in this way? This is the key question that we will be testing later.

2.6 JT gravity

So far, we have been discussing only the fixed geometry of a static black hole, and the physics of quantum fields on that background. It is time now to move beyond this, and to incorporate gravitational dynamics. We will make our lives as easy as we can by discussing perhaps the simplest theory possible: Jackiw-Teitelboim (JT) gravity [7, 8, 9].

At first you may like to study pure Einstein gravity, with action

$$I_{EH} = \frac{1}{16\pi G_N}\left[\int_{\mathcal{M}} d^2x\sqrt{-g}\,\mathcal{R} + 2\int_{\partial\mathcal{M}} ds\,\mathcal{K}\right]. \tag{2.27}$$

But in two dimensions this is purely topological: the variation with respect to the metric (the Einstein tensor) is identically zero. This does not give sensible and interesting dynamics, and cannot be coupled to matter (classically, at least: Einstein's equations set the matter stress-energy to zero, $T_{ab} = 0$).

We must therefore add another term to the action. The simplest possibility is to introduce a scalar 'dilaton' field ϕ, and add the term

$$I_{JT} = \frac{1}{2}\int_{\mathcal{M}} d^2x\sqrt{-g}\,\phi\,(\mathcal{R}+2) + \int_{\partial\mathcal{M}} ds\,\phi\,(\mathcal{K}-1)\,. \tag{2.28}$$

This may look like an arbitrary choice, but in fact this theory emerges very naturally from studying near-extremal (low temperature) black holes in higher dimensions, for example Reissner-Nordstrom black holes in four dimensions. In such cases, the Einstein term is proportional to the area of the event horizon of the extremal black hole, and ϕ describes deviations of the area of the transverse sphere from its extremal value. The JT action is then the leading order approximation (linear in ϕ) when these deviations are small. This has a description in terms of a weakly broken symmetry, which we will briefly describe later.

We can now begin to understand why this theory is simple. Since ϕ appears linearly, its equation of motion simply becomes the constraint

$$\mathcal{R} = -2, \tag{2.29}$$

so our solutions have constant negative curvature. We have chosen units to set the curvature scale ℓ to unity. In fact, something stronger is true, since in the quantum theory we can integrate out ϕ and impose (2.29) as a delta-function, constraining the path integral only to such constant-curvature metrics.

The total action we will study is given by

$$I = I_{EH} + I_{JT} + I_{\text{matter}} + I_{\text{counterterm}}. \tag{2.30}$$

I_{matter} is the action of our (conformal) matter theory, which propagates on the metric g: in particular, it does not couple directly to the dilaton, so we do not spoil the property above. We have allowed for the possibility of counterterms (which will not be too important for us, though we will have occasional cause to mention them).

Lastly, we discuss the boundary conditions, which we impose asymptotically. Take large dilaton, $\phi = \frac{\gamma}{\epsilon}$ where we will take $\epsilon \to 0$. Writing the intrinsic metric on the boundary as $ds^2 = -\frac{1}{\epsilon^2}dt^2$ defines the physical boundary time t. This time defines our lightcone coordinates u, v at infinity ($u = v = t$), and our coupling to the reservoir as described in section 2.4.

2.7 Black holes in JT gravity

Writing our metric in ingoing coordinates r, v as above, the curvature is given by $\mathcal{R} = -\partial_r^2 f$. We can therefore write the most general constant curvature metric as

$$ds^2 = -(r^2 - r_h(v)^2)dv^2 + 2drdv \tag{2.31}$$

for some function $r_h(v)$, where we have used the freedom to shift r by a function of v alone to remove a term linear in r. The boundary is then at large r, $r = \frac{1}{\epsilon} + O(\epsilon)$.

Now, on constant curvature metrics $\mathcal{R} = -2$, the bulk part of the JT action vanishes, and the action is purely a boundary term. Using the boundary condition $\phi_\partial = \frac{\gamma}{\epsilon}$, we can evaluate this on the metric (2.31), finding

$$I_{JT} = -\frac{\gamma}{2}\int_{t_0}^{t} dv\, r_h(v)^2 \tag{2.32}$$

in the limit $\epsilon \to 0$. In particular, from this we can read off the energy: this is given by the Hamilton-Jacobi equation $E = -\frac{\partial I}{\partial t}$, where t is the endpoint of the integral computing the action I, so

$$E = \tfrac{1}{2}\gamma r_h^2. \tag{2.33}$$

Let us look now at the classical solutions, first without matter. In addition to the constant curvature condition, we need to satisfy the equation of motion from varying the metric,

$$\nabla_a \nabla_b \phi - g_{ab}\nabla^2\phi + g_{ab}\phi = 0. \tag{2.34}$$

The rr component and rv component, along with the boundary conditions, uniquely specify a simple solution:

$$\phi = \gamma r. \tag{2.35}$$

The vv component then fixes $\partial_v r_h = 0$, so the energy (2.33) is independent of time as we would expect, giving us a static black hole solution.

The surface gravity is given by $\kappa = r_h$, so in the presence of of matter we expect the black hole to radiate at the Hawking temperature

$$T = \frac{r_h}{2\pi}. \tag{2.36}$$

Now that we have the energy and temperature, we can use the first law $dE = TdS$ to find the entropy:

$$S = S_0 + 2\pi\gamma r_h = S_0 + 2\pi\phi_h = S_0 + \gamma(2\pi)^2 T. \tag{2.37}$$

We have here included an integration constant S_0, which is not fixed by the first law. We will later find a natural way to fix this using the Einstein-Hilbert term I_{EH} in the action, but for now it is arbitrary.

2.8 The Schwarzian description

Before adding matter, we will give another description for the metric (2.31), which relates JT gravity to the 'Schwarzian theory' [10, 11, 12].

To arrive at this, we can observe that the constant curvature constraint $\mathcal{R} = -2$ uniquely fixes the metric to be AdS_2, since the scalar curvature completely fixes the geometry in two dimensions (up to topology). There must therefore be some other coordinates U, V such that the metric can be written in the usual 'Poincaré' form

$$ds^2 = -\left(\frac{2}{U-V}\right)^2 dU dV. \tag{2.38}$$

The asymptotic boundary is at $U = V = T$, but T will not coincide with our physical time t: instead, we have some diffeomorphism F relating them as $T = F(t)$. This diffeomorphism determines U, V in terms of our usual lightcone coordinates as $U = F(u)$, $V = F(v)$, so we can write the metric as

$$ds^2 = -\left(\frac{2}{F(u) - F(v)}\right)^2 F'(u)F'(v)dudv. \tag{2.39}$$

Finally, we can recover the form of our ingoing coordinates (2.31) by defining

$$r = \frac{F''(v)}{F'(v)} + 2\frac{F'(v)}{F(u) - F(v)}. \tag{2.40}$$

The final step is to identify the horizon radius r_h in terms of the diffeomorphism F, which gives us

$$r_h(t)^2 = -2\,\mathrm{Schw}(F, t), \tag{2.41}$$

where Schw is the 'Schwarzian derivative':

$$\mathrm{Schw}(F, t) = \frac{F'''(t)}{F'(t)} - \frac{3}{2}\left(\frac{F''(t)}{F'(t)}\right)^2. \tag{2.42}$$

Putting this expression for r_h^2 into our action (2.32) gives us the *Schwarzian action*

$$I_{\mathrm{Schw}} = \gamma \int dt\,\mathrm{Schw}(F, t), \tag{2.43}$$

and the energy is

$$E = -\gamma\,\mathrm{Schw}(F, t). \tag{2.44}$$

The constant γ is not a dimensionless coupling constant (it has units of time), so we should think of it as setting a scale in the theory. Specifically, γ^{-1} gives an energy or temperature scale below which the theory becomes strongly coupled. For large energies $\gamma E \gg 1$, the action will be large, which suppresses quantum fluctuations so we may treat the theory classically. We will study such large energy black holes, so gravitational quantum fluctuations are a small effect. Since we have fixed units in AdS by setting the curvature scale ℓ to unity, we take r_h to be of order one and $\gamma \gg 1$ in these AdS units.

We can now describe the gravitational dynamics in terms of this diffeomorphism F with the Schwarzian action. However, importantly, not every diffeomorphism gives rise to a different metric: the physical configuration is invariant under the fractional linear transformations

$$F \mapsto \frac{aF+b}{cF+d}, \quad a,b,c,d \in \mathbb{R}, ad-bc=1, \tag{2.45}$$

which form the group $PSL(2,\mathbb{R})$. This redundancy arises from the symmetries of AdS_2. In particular, this means that the Schwarzian action is not in fact a local functional of time, despite appearances: the apparent locality is spoiled by the nonlocal action of the $PSL(2,\mathbb{R})$ gauge redundancy.

For avoidance of possible confusion, we should note that the Lorentzian signature description in terms of diffeomorphisms modulo $PSL(2,\mathbb{R})$ is valid only locally, since the range of F will generically not be $\mathbb{R}$, and fractional linear transformations introduce poles. Things are slightly more straightforward in the Euclidean description, discussed briefly in a moment.

We can now write our static classical solutions without matter in terms of the Schwarzian variable. Up to a fractional linear transformation, the solution is

$$F(t) = -\exp\left(-r_h t\right). \tag{2.46}$$

Note that F describes the change of variables $U = F(u)$ between the outgoing coordinate u and an outgoing 'Kruskal' coordinate U which is smooth at the horizon. The exponential form of F is a signature of the exponential divergence of outgoing null geodesics near the horizon.

2.9 Symmetry origin of the Schwarzian

This section provides a little background about the Schwarzian theory: it is not necessary for what follows.

The Schwarzian theory described above makes an appearance in other contexts, most notably in the SYK model, a quantum mechanical model of $N \gg 1$ Majorana fermions with random interactions. The fact that it crops up in somewhat generic circumstances can be understood from symmetry: the Schwarzian can be thought of as a pseudo-Nambu-Goldstone boson for a particular pattern of symmetry breaking.

To do this, it will be slightly simpler to discuss the Euclidean theory, with periodic Euclidean time $t_E \sim t_E + \beta$ describing the theory at temperature $T = \beta^{-1}$. The metric is then Euclidean AdS_2, which we think of as the Poincaré disk. The Schwarzian diffeomorphism gives the map from physical time t_E to an angular coordinate ϕ on the boundary, where $F = \tan\frac{\phi}{2}$, and ϕ is 2π periodic. This means that we have a map from the t_E circle to the ϕ circle, so the configurations of the theory (the gauge orbits of the diffeomorphism) take values in the coset

$$[\phi] \in \mathrm{Diff}(S^1)/PSL(2,\mathbb{R}). \tag{2.47}$$

If we were interested in locally AdS spacetimes with only the Einstein-Hilbert action, this coset describes spontaneous symmetry breaking. Since the Einstein theory is topological, it has zero Hamiltonian, and can only describe degenerate states at zero energy.[3] All diffeomorphisms of time them become symmetries. But any given solution does not respect this full symmetry group, but only the isometry group of AdS: in other words, it spontaneously breaks $\mathrm{Diff}(S^1)$ to $PSL(2,\mathbb{R})$, and F is a Goldstone boson for that breaking.

To get an interesting theory with dynamics, we must include a leading order correction which explicitly breaks the symmetry. That comes from writing down an action on the coset with the lowest dimension possible, which is precisely the Schwarzian action. In JT gravity, this explicit breaking comes from the dilaton. F is then only a 'pseudo-Goldstone'.

This can be compared to chiral symmetry breaking in QCD. With N_F massless quarks, there is a chiral symmetry $SU(N_F)\times SU(N_F)$, which is broken spontaneously to the diagonal $SU(N_F)$ subgroup. We then have Goldstone bosons — pions — living on the coset $SU(N_F)\times SU(N_F)/SU(N_F)$. Unlike the Schwarzian, this is an interesting theory because we can write down an action with kinetic terms on the coset (the chiral Lagrangian). Giving the quarks small masses (compared to the scale of the symmetry breaking) provides a small explicit symmetry breaking, which can be described by adding a term which does not respect the full symmetry of the coset. The pions are then described as pseudo-Goldstone bosons.

2.10 The semiclassical approximation

We would now like to incorporate matter. Before getting to the specific details, we first outline the general strategy. Ideally, we would like to compute the full path integral over gravitational variables (the metric g and dilaton ϕ) and matter fields X. But even for such a simple model, that's not so

[3]This is related to the fact that there is no nontrivial scale-invariant theory in one dimension, so we must break that symmetry [13]. Without introducing a scale, the density of states can only be of the form $e^{S_0}\delta(E)+\frac{C}{E}$. The second term is unphysical, describing a continuum of infinitely many states, so we only have the first: a space of degenerate zero-energy states. This is avoided in higher dimensions, since the volume of space sets a scale.

easy, so we will make use of a semiclassical limit, looking for saddle-points where the action is stationary. But simply looking for saddle-points of the classical action will not quite be sufficient, because over a long period of time, quantum effects from the matter (namely, production of Hawking radiation) build up to become important. We therefore use an intermediate 'semiclassical' approach, treating the matter fully quantum mechanically, but looking for saddle-points in the integral over gravitational variables.

The full quantum mechanical matter theory is captured by the quantum effective action I_{eff}. This is a functional only of the metric, obtained by integrating out the CFT fields X:

$$e^{iI_{\text{eff}}[g]} = \int \mathcal{D}X \, e^{iI_{\text{CFT}}[X,g]}. \tag{2.48}$$

This implicitly depends on our boundary conditions, in particular the state of the matter and any operators we are inserting to compute an expectation value. We then look for stationary points of the action I_{JT} for the metric and dilaton plus the quantum effective action I_{eff},

$$\delta I_{JT} + \delta I_{\text{eff}} = 0, \tag{2.49}$$

for any metric variation.

By definition, varying the effective action with respect to the metric g gives us the expectation value of the stress tensor (inserted in whatever correlation function we want to compute). Saddle-points therefore correspond to solutions to the metric equation of motion (2.34) sourced by the expectation value of the CFT stress tensor:

$$\nabla_a \nabla_b \phi - g_{ab} \nabla^2 \phi + g_{ab} \phi + \langle T_{ab} \rangle = 0. \tag{2.50}$$

Our task is to find solutions to this equation, giving us the quantum-corrected geometry.

2.11 Evaporation in JT

Now, to determine the expectation value $\langle T_{ab} \rangle$ we have to choose some state. However, as we saw from studying the static black hole spacetime, the dynamics quickly becomes independent of this choice. Specifically, any state which is nonsingular at the horizon at some time v_0 will have exponentially small $\langle T_{rr} \rangle$ a little later, decaying as $e^{-2\kappa(v-v_0)}$. The timescale κ^{-1} for this decay is much shorter than the time over which appreciable evaporation takes place, so our static near-horizon analysis from earlier remains appropriate. We will therefore choose a state with $\langle T_{rr}\rangle = 0$, as an excellent approximation to any state with a smooth horizon at earlier times.

It will also be convenient to add a counterterm to cancel the source from the trace of the stress tensor, given by the anomaly (2.13). Because the metric is constant curvature, the anomaly gives just a constant source: this can be cancelled with a finite cosmological constant counterterm, proportional to $\int d^2x\sqrt{-g}$. This amounts only to a finite redefinition of parameters, which can be absorbed by shifting the dilaton. The upshot is that we will simply set $\langle \mathrm{Tr}\, T\rangle = 0$.

That leaves the only nonzero component of $\langle T_{ab}\rangle$ as the ingoing energy

$$\langle T_{vv}\rangle = F_v(v) - \frac{c}{48\pi} r_h(v)^2, \tag{2.51}$$

where we have used the conservation equation (2.17) with conformal factor $\omega(u,v) = \frac{1}{2}\log\left(\frac{4F'(u)F'(v)}{(F(u)-F(v))^2}\right)$ to write it in terms of the flux from infinity.

With this simple expression for the stress tensor expectation value, it is simple to solve the equation of motion (2.50) as before. In fact, we still have the same result as (2.35) relating the dilaton to r ($\phi = \gamma r$), but now we have interesting dynamics for r_h, or equivalently the energy E:

$$\dot{E}(t) = F_v(t) - \frac{c}{24\pi\gamma} E(t). \tag{2.52}$$

This simply tells us that the change in the energy equals the ingoing flux, minus the outgoing flux of Hawking radiation.

Let us now suppose that there is no incoming energy, $F_v = 0$. The energy then decays exponentially:

$$E(t) = E(0)e^{-2kt} \implies r_h(t) = r_h(0)e^{-kt}, \tag{2.53}$$

where the decay rate is given by

$$k = \frac{c}{48\pi\gamma} \ll 1. \tag{2.54}$$

Now, the exact solution for the diffeomorphism F is slightly complicated, involving Bessel functions. But there is a very simple approximate solution, which is extremely close to the exact solution whenever the black hole remains in the semiclassical regime $\gamma E \gg 1$:

$$F(t) \sim -\exp\left(\tfrac{1}{k} r_h(0)e^{-kt}\right) = -\exp\left(\tfrac{1}{k} r_h(t)\right). \tag{2.55}$$

In fact, this solves (2.52) with a very weak constant incoming flux $F_v = \frac{c}{48\pi}k^2$ of radiation at temperature $\frac{k}{2\pi} \ll 1$, which is negligible until extremely late times. Note that over any short period of time, this is well approximated by our static solution (2.46). Specifically, write $t = t_0 + \Delta t$ where t_0 can be of order k^{-1}, but $\Delta t \ll k^{-1}$. Then $r_h(t) \sim r_h(t_0)(1 - k\Delta t)$, so

$$F(t) \sim -\exp\left(\tfrac{1}{k} r_h(t_0)\right) e^{-r_h(t_0)\Delta t}. \tag{2.56}$$

2.12 Horizons

To complete our discussion of the geometry of the evaporating black hole, we can look at two notions of 'horizon'. First, in this dynamical geometry, $r = r_h$ is no longer the event horizon: outgoing light from there will escape to infinity. Instead, it is an 'apparent horizon'. In the JT context, the apparent horizon is defined by the criterion that the dilaton is stationary to first order variation along outgoing null geodesics: $\left.\frac{\partial\phi}{\partial v}\right|_u = 0$. In higher dimensions, we define an apparent horizon as a surface with stationary area along outgoing null rays (a marginally trapped surface).

If the null energy condition is satisfied, an apparent horizon is destined to lie inside an event horizon. The reason is that lightrays focus (since gravity is attractive), so if the area is stationary to first order it can only decrease. This focusing theorem is the key idea behind the Hawking-Penrose singularity theorems.

The event horizon is defined by the surface where $u \to \infty$. Using our approximate solution (2.55) and (2.40), the event horizon lies at

$$r_{\mathrm{EH}}(v) \sim r_h(v) - k, \tag{2.57}$$

just inside the apparent horizon. This is possible because quantum matter does not obey the null energy condition. Indeed, we saw earlier that there is a flux of negative energy through the horizon, a violation of the null energy condition.

Exercise 2. *Let $N^a = \frac{d}{d\lambda}$ be the tangent vector of a null geodesic with affine parameter λ, and define the 'divergence' of the geodesic by $\theta = \frac{d}{d\lambda}\phi$. Using the equation of motion* (2.50)*, find the JT gravity version of the 'Raychaudhuri equation':*

$$\frac{d\theta}{d\lambda} = -N^a N^b T_{ab}\,. \tag{2.58}$$

Show that under the null energy condition $N^a N^b T_{ab} \geq 0$, a trapped surface ($\theta < 0$ for the outgoing direction) must lie inside an event horizon.

3 Entropy and information in quantum systems

We now take a brief detour, leaving black holes to one side and considering entropy in quantum systems more generally. This will set the stage for studying entropy and information for the evaporating black hole.

3.1 Fine-grained and coarse-grained entropy

First, there are several different notions of entropy, so we should be precise about what we mean.

The main thing we will be interested in is the *von Neumann entropy* of a quantum state with density matrix ρ:

$$S(\rho) = -\operatorname{Tr}(\rho \log \rho). \tag{3.1}$$

In many cases we will be looking at the state of a subsystem, in which case we use the reduced density matrix. In that context the von Neumann entropy is often called the *entanglement entropy*, since if the full state is pure, the von Neumann entropy of a subsystem is a measure of the quantum entanglement between the subsystem and its complement.

The von Neumann entropy is distinct from the entropy we meet in thermodynamics. In particular, the second law tells us that the thermodynamic entropy will typically increase over time, even if we have an isolated system. But the von Neumann entropy of an isolated system does not change: the density matrix evolves as $\rho \mapsto U(t)\rho U(t)^\dagger$ for the unitary time-evolution operator $U(t) = e^{-iHt}$, which does not affect $S(\rho)$. We need a different notion of 'coarse-grained' entropy.

Coarse-graining means that we regard many states as indistinguishable if they are the same for some set $\mathbb{O}$ of simple 'macroscopic' observables $\mathcal{O}$:

$$\mathbb{O} = \{\text{operators } \mathcal{O} \text{ corresponding to macroscopic observables}\}\,. \tag{3.2}$$

This clearly involves a choice, and we will mention a couple of possibilities in a moment. Having chosen $\mathcal{O}$, we can then define a coarse-grained entropy $S_{\mathbb{O}}$ by maximising the von Neumann entropy $S(\tilde{\rho})$ over all states $\tilde{\rho}$ with the specified expectation values for macroscopic observables $\mathcal{O} \in \mathbb{O}$:

$$S_{\mathbb{O}}(\rho) = \max\left\{S(\tilde{\rho}) : \operatorname{Tr}(\mathcal{O}\tilde{\rho}) = \operatorname{Tr}(\mathcal{O}\rho) \text{ for all } \mathcal{O} \in \mathbb{O}\right\}. \tag{3.3}$$

We can define a coarse-grained density matrix $\rho_{\mathbb{O}}$ as the (unique) $\tilde{\rho}$ that attains this maximum. A 'coarse-graining' means choosing a smaller set of macroscopic observables.

Exercise 3. *Show that the coarse-grained density matrix* $\rho_{\mathbb{O}}$ *(attaining the maximum in* (3.3)*) is unique.*

In a thermodynamic limit, for which there are many orthogonal states $N \gg 1$ with observables close to the specified values, we may think of the entropy as providing us with a count of those states: $S_{\mathbb{O}} \sim \log N$.

The simplest example, most relevant to systems close to equilibrium, is the canonical entropy, where $\mathbb{O} = \{H\}$, so we fix only the energy $\mathrm{Tr}(\rho H) = E$. A state maximising the entropy under such a constraint is given by the canonical density matrix

$$\rho_{\mathrm{can}}(\beta) = \frac{e^{-\beta H}}{Z(\beta)}, \qquad Z(\beta) = \mathrm{Tr}(e^{-\beta H}) \tag{3.4}$$

for some inverse temperature $\beta = \frac{1}{T}$. The canonical entropy is then the von Neumann entropy of this state:

$$S_{\mathrm{can}}(\beta) = S(\rho_{\mathrm{can}}(\beta)). \tag{3.5}$$

This can also be expressed in terms of the free energy F:

$$S_{\mathrm{can}} = -\frac{\partial F}{\partial T}, \quad F = -T \log Z. \tag{3.6}$$

If there are other conserved quantities (like charge or angular momentum) which do not equilibrate, we might like to add them to the list $\mathcal{O}$, considering a grand canonical ensemble.

If our system is not in equilibrium, we might like to use a slightly more refined choice of coarse-graining. For example, we might specify the density of energy and momentum (and perhaps other conserved quantities) as a function of space, as in hydrodynamics where equilibrium is only attained locally.

For us, coarse-grained entropy will be useful to provide a bound on fine-grained (von Neumann) entropy. By definition, entropy increases under coarse graining. That is, if $\mathcal{O}$ is more coarse-grained that $\mathcal{O}'$, meaning that it contains fewer macroscopic observables $\mathbb{O} \subset \mathbb{O}'$, then $S_{\mathbb{O}} \geq S_{\mathbb{O}'}$. This follows simply because in (3.3), we are maximising with fewer constraints. The extreme example is the von Neumann entropy, for which we take $\mathbb{O}'$ to contain all operators:

$$S(\rho) \leq S_{\mathbb{O}}(\rho). \tag{3.7}$$

We will suppress the choice of coarse-graining, from here on denoting $S_{\mathbb{O}}$ as a 'thermal entropy' S_{thermal}. We assume some choice which is sufficiently detailed to capture the local properties and dynamics of the state, so that our bound (3.7) approaches the best possible without considering detailed correlations.

For a black hole, the 'no hair theorem' suggests that the canonical (or grand canonical) entropy is a good choice of coarse-graining, and we will work under the hypothesis that this is given by the Bekenstein-Hawking formula obtained from the first law in section 2.5. For the Hawking radiation, a better choice involves the local stress-tensor $\langle T_{ab}(x) \rangle$, since radiation emitted at different times does not equilibrate.

3.2 The Page curve

Let us now apply these ideas to learn about the state of a quantum system as it loses energy to its surroundings. Specifically, prepare a many-body system A in some excited pure state, and weakly couple it to a large reservoir (or 'bath') B prepared in its ground state. Then, energy will typically flow from A to B, until (if B is sufficiently large) A cools to its own ground state. Ultimately, we will be interested in an evaporating black hole, but for now it might be helpful to think of a more ordinary system. Perhaps A could be a lump of metal, and B a lab in which it sits: as the metal cools, it fills the lab with radiation.

Now, since the coupled system evolves unitarily and the final state of A is pure (the ground state), the final state of B must also be pure. So the von Neumann entropy of A or B (i.e., their entanglement entropy) begins at zero and ends at zero. But unitarity tells us more about the evolution of the von Neumann entropy over time, since we have the bound (3.7) by the coarse-grained thermodynamic entropy. This in fact gives us two separate bounds, by the thermal entropy of either subsystem:

$$S_{\mathrm{vN}} \leq \min \left\{ S_{\mathrm{th}}(A), S_{\mathrm{th}}(B) \right\}. \tag{3.8}$$

At early times, the thermal entropy of B will be small and the thermal entropy of A large, so the second bound will be relevant. At late times, the first bound will be relevant instead. Moreover, we typically expect the tighter bound to be extremely close to saturation (for a sensible choice of coarse-graining), so that (3.8) becomes an approximate equality.

The resulting function

$$S_{\mathrm{Page}}(t) = \min \left\{ S_{\mathrm{th}}(\rho_A(t)), S_{\mathrm{th}}(\rho_B(t)) \right\} \tag{3.9}$$

is called the 'Page curve', since it was introduced by Don Page as a quantitative expectation of unitary black hole evaporation. A typical example is shown in a figure. There is an initial steady increase (following $S_{\text{th}}(B)$), until the 'Page time' t_{Page} at which there is a sharp transition to a steady decrease (following $S_{\text{th}}(A)$).

Exercise 4. *Take system A to be a box containing weakly-interacting relativistic bosons (photons, perhaps) in d dimensions of space, so the energy at temperature T is $E = CT^{d+1}$ for some constant C (proportional to the volume of the box). Allow the photons to escape through a small hole into the outside world (system B), so some small fraction λ of them are emitted per unit time. Assume that this process is slow compared to the equilibration time of the system, so that the photons remain in equilibrium as the energy leaks out. (Note that for $d = 1$, this is also describes our evaporating black hole model in section 2.1!)*

What is the rate of increase of thermal entropy for system B? What proportion of the energy has escaped at the 'Page time'? Compare the final thermal entropy of B to the initial thermal entropy of A.

To provide some justification for why we might expect the bound (3.8) to be close to saturation, we can make a crude model for the dynamics. Namely, we model systems A and B as finite-dimensional Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$ with dimensions $\dim \mathcal{H}_A \approx e^{S_{\text{th}}(A)}$, $\dim \mathcal{H}_B \approx e^{S_{\text{th}}(B)}$ given by their respective thermal entropies, and take the state to be chosen at random from the tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$. Roughly, this model says that the dynamics produce a pure state on AB with the correct macroscopic observables, but which is otherwise completely generic. In this model, the average entropy is very close to maximal: the difference between the two sides of (3.8) is exponentially small in the difference $|S_{\text{th}}(A) - S_{\text{th}}(B)|$ between the two thermal entropies.

We expect this to be a reasonable approximation if the dynamics of A is sufficiently chaotic. The time-evolution operator should be sufficiently generic after some timescale — the 'scrambling time' t_{scr} — on which local information is spread over the whole system. As long as this timescale is much shorter than the time over which evaporation is significant, we expect the Page curve to be an excellent approximation. A reasonable estimate for the scrambling time of a strongly-interacting system is $t_{\text{scr}} \sim \beta \log S_{\text{th}}(A)$. The number of degrees of freedom effected by a small local perturbation will tend to grow exponentially at a rate set by the timescale of local interactions, which might typically be of order β. The effect of the perturbation will then be spread over the whole system after a time which is logarithmic in the number of excited degrees of freedom.

References

[1] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*. W. H. Freeman, San Francisco, 1973. 3

[2] P. Di Francesco, P. Mathieu, and D. Senechal, *Conformal Field Theory*. Graduate Texts in Contemporary Physics. Springer-Verlag, New York, 1997. 6

[3] J. Polchinski, *String theory. Vol. 1: An introduction to the bosonic string*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 12, 2007. 6, 7

[4] X. Yin, "Aspects of Two-Dimensional Conformal Field Theories," *PoS* **TASI2017** (2017) 003. 6

[5] J. D. Bekenstein, "Black holes and entropy," *Phys. Rev. D* **7** (1973) 2333–2346. 10

[6] S. Hawking, "Particle Creation by Black Holes," *Commun. Math. Phys.* **43** (1975) 199–220. [Erratum: Commun.Math.Phys. 46, 206 (1976)]. 11

[7] R. Jackiw, "Lower Dimensional Gravity," *Nucl. Phys. B* **252** (1985) 343–356. 11

[8] C. Teitelboim, "Gravitation and Hamiltonian Structure in Two Space-Time Dimensions," *Phys. Lett. B* **126** (1983) 41–45. 11

[9] A. Almheiri and J. Polchinski, "Models of AdS_2 backreaction and holography," *JHEP* **11** (2015) 014, `1402.6334`. 11

[10] J. Maldacena, D. Stanford, and Z. Yang, "Conformal symmetry and its breaking in two dimensional Nearly Anti-de-Sitter space," *PTEP* **2016** (2016), no. 12 12C104, `1606.01857`. 13

[11] K. Jensen, "Chaos in AdS_2 Holography," *Phys. Rev. Lett.* **117** (2016), no. 11 111601, `1605.06098`. 13

[12] J. Engelsöy, T. G. Mertens, and H. Verlinde, "An investigation of AdS_2 backreaction and holography," *JHEP* **07** (2016) 139, `1606.03438`. 13

[13] K. Jensen, S. Kachru, A. Karch, J. Polchinski, and E. Silverstein, "Towards a holographic marginal Fermi liquid," *Phys. Rev. D* **84** (2011) 126002, 1105.1772. 16