Lectures on Quantum Gravity and Black Holes

Thomas Hartman

Cornell University

Please email corrections and suggestions to: hartman@cornell.edu

Abstract These are the lecture notes for a one-semester graduate course on black holes and quantum gravity. We start with black hole thermodynamics, Rindler space, Hawking radiation, Euclidean path integrals, and conserved quantities in General Relativity. Next, we rediscover the AdS/CFT correspondence by scattering fields off near-extremal black holes. The final third of the course is on AdS/CFT, including correlation functions, black hole thermodynamics, and entanglement entropy. The emphasis is on semiclassical gravity, so topics like string theory, D-branes, and super-Yang Mills are discussed only very briefly.

Course Cornell Physics 7661, Spring 2015

Prereqs This course is aimed at graduate students who have taken 1-2 semesters of general relativity (including: classical black holes, Penrose diagrams, and the Einstein action) and 1-2 semesters of quantum field theory (including: Feynman diagrams, path integrals, and gauge symmetry.) No previous knowledge of quantum gravity or string theory is necessary.

Contents

1 The problem of quantum gravity

These are lectures on quantum gravity. To start, we better understand clearly what problem we are trying to solve when we say 'quantum gravity.' At low energies, the classical action is

$$S = \frac{1}{16\pi G_N}\int \sqrt{-g}\,(R - 2\Lambda + \mathcal{L}_{matter}) \ . \tag{1.1}$$

Why not just quantize this action? The answer of course is that it is not renormalizable. This does not mean it is useless to understand quantum gravity, it just means we have to be careful about when it is reliable and when it isn't. In this first lecture we will consider gravity as a low-energy effective field theory,* see when it breaks down, and make some general observations about what we should expect or not expect from the UV completion.

1.1 Gravity as an effective field theory

The rules of effective field theory are:

1. Write down the most general possible action consistent with the symmetries;
2. Keep all terms up to some fixed order in derivatives;
3. Coefficients are fixed by dimensional analysis, up to unknown order 1 factors (unless you have a good reason to think otherwise);
4. Do quantum field theory using this action, including loops;
5. Trust your answer only if the neglected terms in the derivative expansion are much smaller than the terms you kept.

This works for renormalizable or non-renormalizable theories. Let's follow the steps for gravity. Our starting assumption is that nature has a graviton — a massless spin-2 field. This theory can be consistent only if it is diffeomorphism invariant.

*Donaghue gr-qc/9512024.

Counting metric degrees of freedom

This can be argued various ways*; we'll just count degrees of freedom. In 4D, a massless particle has two degrees of freedom (2 helicities). Similarly, the metric $g_{\mu\nu}$ has†

$$10 \text{ components} - 4 \text{ diffeos} - 4 \text{ non-dynamical} = 2 \text{ dof} . \tag{1.2}$$

In D dimensions, we count dof of a massless particle by looking at how the particle states transform under $SO(D-2)$, the group of rotations that preserve a null ray.‡ A spin-2 particle transforms in the symmetric traceless tensor rep of $SO(D-2)$, which has dimension $\frac{1}{2}(D-2)(D-1)-1 = \frac{1}{2}D(D-3)$. Similarly, assuming diffeomorphism invariance, the metric has

$$\tfrac{1}{2}D(D+1) - D - D = \tfrac{1}{2}D(D-3) \tag{1.3}$$

degrees of freedom.

Note that in $D=3$, the metric has no (local) dof. It turns out that it does have some nonlocal dof; this will be useful later in the course.

Back to effective field theory: Steps 1 and 2, the derivative expansion

The only things that can appear in a diff-invariant Lagrangian for the metric are objects built out of the Riemann tensor $R_{\mu\nu\rho\sigma}$ and covariant derivatives ∇_μ. Each Riemann contains $\partial\partial g$, so the derivative expansion is an expansion in the number of R's and ∇'s. Up to 4th order in derivatives,

$$S = \frac{1}{16\pi G_N}\int \sqrt{-g}\left(-2\Lambda + R + c_1 R^2 + c_2 R_{\mu\nu}R^{\mu\nu} + c_3 R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} + \cdots\right) \tag{1.4}$$

So the general theory is the Einstein-Hilbert term plus higher curvature corrections. We have ignored the matter terms $\mathcal{L}_{matter}$ and matter-curvature couplings, like ϕR.

*See Weinberg QFT V1, section 5.9, and the discussion of the Weinberg-Witten theorem below, and Weinberg Phys. Rev. 135, B1049 (1964).

†In more detail: a 4x4 symmetric matrix has 10 independent components. In 4D we have 4 functions worth of diffeomorphisms, $x^\mu \to x^{\mu'}(x^\mu)$. And $\ddot{g}_{0\mu}$ cannot appear in a 2nd order diff-invariant equation of motion, so these components are non-dynamical. For more details, see the discussion of gravitational waves in any introductory GR textbook, which should show that in transverse-traceless gauge the linearized Einstein equation have two independent solutions (the '+' and '×' polarizations).

‡See Weinberg QFT V1, section 2.5.

Step 3: Coefficients = scale of new physics

Coefficients should be fixed by dimensional analysis, up to $O(1)$ factors. This doesn't work for the cosmological constant: experiment (ie the fact the universe is not Planck-sized) indicates that Λ is unnaturally small. This is the cosmological constant problem. We will just sweep this under the rug, take this fine-tuning as an experimental fact, and proceed to higher order.

For these purposes let's take the coordinates to have dimensions of length, so the metric is dimensionless, and R has mass dimension 2. The action should be dimensionless (since $\hbar = 1$). Looking at the Einstein-Hilbert term, that means $[G_N] = 2 - D$, so in terms of the Planck scale,

$$\frac{1}{G_N} \equiv (M_P)^{D-2} \ . \tag{1.5}$$

In $D > 2$, this term is not renormalizable. This means that the theory is strongly coupled at the Planck scale. If we try to compute scattering amplitudes using Feynman diagrams, we would find non-sensical, non-unitary answers for $E \gtrsim M_P$. The rules of effective field theory tell us that we must include the R^2 terms, with coefficients $c_{1,2,3} \sim 1/M_P^2$. Higher curvature terms should also be included, suppressed by more powers of M_P. More generally, the rule is that these coefficients should be suppressed by the scale of new physics, which we will call M_s. New physics must appear at or below the Planck to save unitarity, so $M_s < M_P$, but it's possible that $M_s \ll M_P$. So to allow for this possibility, we set

$$c_{1,2,3} \sim \frac{1}{M_s^2} \ . \tag{1.6}$$

In string theory M_s would be the mass of excited string states:

$$M_s \sim \frac{1}{\ell_s} \ , \qquad c_{1,2,3} \sim \alpha' \tag{1.7}$$

where ℓ_s is the string length and $\alpha' = \ell_s^2$ is the string tension. In this context the R^2 and higher curvature terms in the action are called 'stringy corrections.'

Steps 4 and 5: Do quantum field theory, but be careful what's reliable

In this section to be concrete we will work in $D = 4$. To do perturbation theory (about flat space) with the action (1.4), we set

$$g_{\mu\nu} = \eta_{\mu\nu} + \frac{1}{M_P} h_{\mu\nu} \tag{1.8}$$

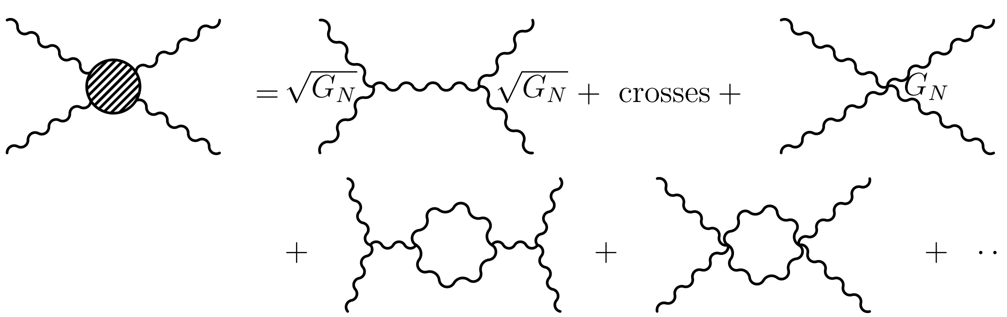
and expand in h, or equivalently in $1/M_P$. The factor of M_P is inserted here so that the quadratic action is canonically normalized; schematically, the perturbative action looks like

$$S \sim \int \partial h \partial h + \frac{1}{M_P} h \partial h \partial h + \cdots + \frac{1}{M_s^2} \left(\partial^2 h \partial^2 h + \frac{1}{M_P} h \partial^2 h \partial^2 h + \cdots \right) \tag{1.9}$$

where the first terms come from expanding the Einstein action, and the other terms come from the higher curvature corrections.[*] In curved space, the higher curvature terms would also contribute to the terms like $h\partial h \partial h$ since $R \sim \text{const} + \partial^2 h + \cdots$.

Scattering and the strong coupling scale

As expected in a non-renormalizable theory, the perturbative expansion breaks down at high energies. First consider the case where we set $M_s = M_P$ (or, we keep only the Einstein term in the Lagrangian) and calculate the amplitude for graviton scattering in perturbation theory:



(1.10)

This is an expansion in the coupling constant $\sqrt{G_N} = 1/M_P$; but this is dimensionful, so it must really be an expansion in E/M_P. That is, each diagram contributes something

[*]The term written here comes from a term in the action $\left(\frac{M_s}{M_P}\right)^2 R^2$, where we pick off the terms $\sqrt{-g} \sim 1 + \delta g$, $R \sim \partial^2 \delta g$, then rescale $\delta g = \frac{1}{M_P} h$.

of order

$$\left(\frac{E^2}{M_P^2}\right)^{1+\text{number of loops}} . \tag{1.11}$$

So the strong coupling scale, where loop diagrams are the same size as tree diagrams, is

$$E_{strong} \sim M_P \ . \tag{1.12}$$

Below the strong coupling scale, this is a perfectly good quantum theory. We can use it to make reliable predictions about graviton-graviton scattering, including calculable loop corrections.

If there are two different scales M_s and M_P with $M_s \ll M_P$ the situation is slightly more complicated. The Einstein term is strongly coupled at the Planck scale, but looking at (1.9), the higher curvature terms become strongly coupled at a lower scale somewhere between M_s and M_P,

$$E_{strong} \sim M_P^x M_s^{1-x} \tag{1.13}$$

for some $x \in (0, 1)$. (This is a known number that you can find by examining all the diagrams.) It is important, however, that interactions in the higher curvature terms still come with powers of $1/M_P$, so even in this case E_{strong} contains some factor of M_P (ie, $x > 0$): the theory is still weakly coupled at the scale of new physics, $E \sim M_s$.

Classical corrections to the Newtonian potential and ghosts

Returning to the classical theory, consider for example the theory with just the first term in (1.4), so $c_1 = (M_s)^{-2}$ and $c_2 = c_3 = 0$. The equations of motion derived from this action are schematically

$$\Box h + \left(\frac{1}{M_s}\right)^2 \Box\Box h = 8\pi G_N T \ . \tag{1.14}$$

Going to momentum space $\Box \sim E^2$, so clearly the higher curvature term is negligible at low energies $E \ll M_s$. The propagator looks like

$$\frac{1}{q^2 + M_s^{-2} q^4} = \frac{1}{q^2} - \frac{1}{q^2 + M_s^2} \tag{1.15}$$

The 2nd term looks like a massive field with mass M_s, but the wrong sign. It is a new, non-unitary degree of freedom, or 'ghost', in the classical theory. Although the only field is still the metric, it makes sense that we've added a degree of freedom because we need more than 2 functions worth of initial data to solve the 4th order equation of motion (1.14).

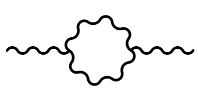
The 'ghost' should not bother us, because it appears at the scale M_s. This is the scale of new physics where we should not trust our effective field theory anyway. And at energies $E \ll M_s$, the ghost has no effect on classical gravity. To see this, let's compute the classical potential between two massive objects. The first term in (1.15) gives the Newtonian $1/r$ potential. The second term looks like a massive Yukawa force, so the classical potential is

$$V(r) = -G_N m_1 m_2 \left[\frac{1}{r} - \frac{e^{-rM_s}}{r} \right] \ . \tag{1.16}$$

This tiny for distances $r > 1/M_s$.

Loop corrections to the Newtonian potential

The 2pt function has calculable, reliable quantum corrections:

 $+$ 

$$+ \cdots \quad \sim \quad \frac{1}{q^2} - \frac{1}{M_s^2} + \frac{1}{q^2}\frac{a}{M_P^2} q^4 \log \frac{q^2}{\Lambda^2}\frac{1}{q^2} + \cdots \tag{1.17}$$

The first two terms are the classical part,

$$\frac{1}{q^2 + M_s^{-2} q^4} = \frac{1}{q^2} - \frac{1}{M_s^2} + \cdots \ , \tag{1.18}$$

and the log term is the loop diagram (including external legs!). a is an order 1 number that can been calculated from this diagram, and we've dropped some terms to simplify the discussion (see Donaghue for details). To calculate the attractive potential between two stationary masses, we set the frequency to zero $q^0 = 0$ and go to position space,*

$$\int d^3\vec{q}\left(\frac{1}{\vec{q}^2} + \frac{1}{M_s^2} - \frac{1}{M_P^2}\log\vec{q}^2 + \cdots\right)e^{i\vec{q}\cdot\vec{x}} \sim \frac{1}{r} + \frac{1}{M_s^2}\delta(r) + \frac{1}{M_P^2 r^3} \tag{1.19}$$

This Fourier transform is just done by dimensional analysis (we've dropped numerical coefficients). The first term is the classical Newtonian potential, the second term is the classical higher-curvature correction, and the last term is a quantum correction. The delta function does not matter at separated points; it is UV physics and does not affect the potential. It came from the same physics as the Yukawa term e^{-rM_s} in our discussion above— the difference is that the Yukawa term is the exact classical contribution whereas the delta function comes from expanding out the propagator in a derivative expansion.

The last term in (1.19) is a reliable prediction of quantum gravity, with small but non-zero effects at low energies.

Does $M_s = M_P$?

So does $M_s = M_P$, or is there a new scale $M_s \ll M_P$? This is basically the question of whether the new UV physics that fixes the problems of quantum gravity is weakly coupled ($M_s \ll M_P$) or strongly coupled (no new scale). Both options are possible, and both are realized in different corners of string theory.

If we ask the analogous question about other effective field theories that exist in nature, then sometimes the new physics is strongly coupled (for example, QCD as the UV completion of the pion Lagrangian) and sometimes it's weakly coupled (for example, electroweak theory as the UV completion of Fermi's theory of beta decay).

Breakdown

As argued above, the effective field theory breaks down at (or below) M_P. It is conceivable that this is just a problem with perturbation theory, and that the theory makes sense non-perturbatively, for example by doing the path integral on a computer. The problem is that the theory is UV-divergent and must be renormalized; this means we do not have the option of just plugging in a particular action, say just the Einstein term

*See for example Peskin and Schroeder section 4.7.

$\int \sqrt{-g}R$, and using this to define a quantum theory. We must include the full series of higher curvature terms. Each comes with a coupling constant, so we have an infinite number of tunable parameters and lose predictive power. The only way this theory can make predictions is if these infinite number of running couplings flow under RG to a UV fixed point with a finite number of parameters. This idea is called 'asymptotic safety,' and although it's a logical possibility there is little evidence for it. It is not how other effective fields theories in nature (eg pions) have been UV completed.

1.2 Quantum gravity in the Ultraviolet

We have seen that at $E \ll M_P$, ordinary methods in quantum field theory can be applied to gravity without any problems. So the real problem, of course, is how to find a UV completion that reduces at low energy to the effective field theory we've just described. This problem is unsolved, but there has been a lot of progress. The point of this section is to make a few comments about what we should and should not expect in a theory of quantum gravity, and to introduce the idea of emergent spacetime.

Quantum gravity has no local observables

Gauge symmetry is not a symmetry. It is a fake, a redundancy introduced by hand to help us keep track of massless particles in quantum field theory. All physical predictions must be gauge-independent.

In an ordinary quantum field theory without gravity, in flat spacetime, there two types of physical observables that we most often talk about are correlation functions of gauge-invariant operators $\langle O_1(x_1)\cdots O_n(x_n)\rangle$, and S-matrix elements. The correlators are obviously gauge-independent. S-matrix elements are also physical, even though electrons are not gauge invariant. The reason is that the states used to define the S-matrix have particles at infinity, and gauge transformations acting at infinity are *true* symmetries. They take one physical state to a different physical state — unlike local gauge transformations, which map a physical state to a different description of the same physical state.

In gravity, local diffeomorphisms are gauge symmetries. They are redundancies. This means that local correlation functions like $\langle O_1(x_1)\cdots O_n(x_n)\rangle$ are not gauge invariant, and so they are not physical observables.[*] On the other hand, diffeomorphisms that reach infinity (like, say, a global translation) are physical symmetries — taking states in the Hilbert space to different states in the Hilbert space — so we get a physical observable by taking the insertion points to infinity. This defines the S-matrix, so it is sometimes said that 'The S-matrix is the only observable in quantum gravity.'

This is not quite true, since there are also non-local physical observables. For example, suppose we send in an observer from infinity, along a worldline $x_0^\mu(\tau)$, with τ the proper time along the path. Although the coordinate value $x_0^\mu(\tau)$ depends on the coordinate system, it unambiguously labels a physical point on the manifold; that is, $x_0^\mu(\tau_1)$ labels the same physical point as $x^{\mu'}(\tau_1)$ in some other coordinates. Therefore $\langle O_1(x_0(\tau_1)) \cdots O_n(x_0(\tau_n))\rangle$ should be a physical prediction of the theory, which answers the physical question 'If I follow the path $x_0^\mu(\tau)$, carrying an O-meter, what do I measure?'. So apparently, to construct diff-invariant physical observables, we need to tie them to infinity. Although this sounds like a straightforward fix, it is actually a radical departure from ordinary, local quantum field theory.

The graviton is not composite (Weinberg-Witten theorem)

In QCD, there are quarks at high energies, and pions are composite degrees of freedom that appear at low energy where the quarks are strongly coupled. The pion Lagrangian is non-renormalizable; it breaks down at the QCD scale and must be replaced by the full UV-complete theory of QCD.

Based on this analogy, we might guess that the UV completion of gravity is an ordinary, $D = 4$ quantum field theory with no graviton, and that the graviton is a emergent degree of freedom at low energies. This is wrong. The graviton may be an emergent degree of freedom, but it cannot come from an ordinary $D = 4$ quantum field theory in the UV. The reason is the Weinberg-Witten theorem:[†]

---

[*]This is true in the effective field theory of gravity too, not just in the UV. However in perturbation theory it is not a problem. In perturbation theory, coordinates x^μ label points on the fixed background manifold, which are meaningful. It is only when we allow the geometry to fluctuate wildly that this really becomes a problem.

[†] See the original paper for the proof, it is short and clear. The basic idea is to first argue that, since states carry energy, $\langle p|T^{00}(0)|p\rangle \neq 0$. However, if these are single-particle states for a particle with helicity ± 2, then there is just now way for $\langle p'|T_{\mu\nu}|p\rangle$ to transform properly under rotations unless it vanishes.

A 4D Lorentz-invariant QFT with a conserved, gauge-invariant stress tensor $T_{\mu\nu}$ cannot have massless particles with spin > 1.

This does not rule out general relativity itself because $T_{\mu\nu}$ is not gauge invariant in GR (or, equivalent, the physical part of $T_{\mu\nu}$ is not really a Lorentz tensor). It does rule out a composite graviton: If gravity emerged from an ordinary QFT, then in the UV there is no diffeomorphism symmetry, the stress tensor is gauge invariant, so there can be no graviton in the spectrum.

Emergent spacetime

So the theory of the graviton is sick in the UV, but if we stick to ordinary QFT we cannot eliminate the graviton in the UV. This leaves two possibilities. One is that the graviton appears in the UV theory, along with other degrees of freedom which cure the problems seen in effective field theory. The other is that the graviton is an emergent degree of freedom, but the UV theory is not an ordinary 4D QFT. These are not mutually exclusive, and in fact both of these possibilities are realized in string theory (simultaneously!).

In this course we will focus on the second possibility. We will discuss models where not only the graviton, but spacetime itself is emergent. The fundamental degrees of freedom of the theory do not live in the same spacetime as the final theory, or in some cases do not live in any spacetime at all. Spacetime is an approximate, collective description of these underlying degrees of freedom, and makes sense only the infrared. The graviton is emergent, but evades Weinberg-Witten because the way it emerges is outside the usual framework of QFT.

A look ahead

There are many ways to approach this subject. In this course we will take a route that begins and ends with black holes. Unlike other EFTs (eg the pion Lagrangian), the Einstein action contains an enormous amount of information about the UV completion — infrared hints about the ultraviolet. Much of this information is encoded in the thermodynamics of black holes, so that is our starting point, and will be the basis of the first half of the course. As it turns out, black holes also lead to emergent spacetime and the AdS/CFT correspondence, which are the topics of the second half of the course.

1.3 Homework

Review the chapter on black holes in Carroll's textbook (or online lecture notes) on General Relativity.

2 The Laws of Black Hole Thermodynamics

In classical GR, black holes obey 'laws' that look analogous to the laws of thermodynamics. These are classical laws that follow from the Eintsein equations. Eventually, we will see that in quantum gravity, this is not just an analogy: these laws are the *ordinary* laws of thermodynamics, governing the microsopic UV degrees of freedom that make up black holes.

2.1 Quick review of the ordinary laws of thermodynamics

The first law of thermodynamics is conservation of energy,

$$\Delta E = \mathcal{Q} \tag{2.1}$$

where $\mathcal{Q}$ is the heat transferred to the system.* For quasistatic (reversible) changes from one equilibrium state to a nearby equilibrium state, $\delta\mathcal{Q} = TdS$ so the 1st law is

$$TdS = dE \ . \tag{2.2}$$

Often we will turn on a potential of some kind. For example, in the presence of an ordinary electric potential Φ, the 1st law becomes

$$TdS = dE - \Phi dQ \tag{2.3}$$

where Q is the total electric charge. If we also turn on an angular potential, then the 1st law is

$$TdS = dE - \Omega dJ - \Phi dQ \ . \tag{2.4}$$

The second law of thermodynamics is the statement that in any physical process, entropy cannot decrease:

$$\Delta S \geq 0 \ . \tag{2.5}$$

These laws can of course be derived (more or less) from statistical mechanics. In the microscopic statistical theory, the laws of thermodynamics are not exact, but are an

*Often the rhs is written $Q + W$ where $-W$ is the work done by the system. We'll set $W = 0$.

extremely good approximation in a system with many degrees of freedom.

2.2 The Reissner-Nordstrom Black Hole

The Riessner-Nordstrom solution is a charged black hole in asymptotically flat space. It will serve as an example many times in this course.

Consider the Einstein-Maxwell action (setting units $G_N = 1$),*

$$S = \frac{1}{16\pi}\int d^4x\sqrt{-g}\,(R - F_{\mu\nu}F^{\mu\nu}) \tag{2.6}$$

where $F_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu = \partial_\mu A_\nu - \partial_\nu A_\mu$. This describes gravity coupled to electromagnetism. The equations of motion derived from this action are

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = 8\pi T_{\mu\nu} \tag{2.7}$$

$$\nabla_\mu F^{\mu\nu} = 0 \tag{2.8}$$

with the Maxwell stress tensor

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S^{matter}}{\delta g^{\mu\nu}} = \frac{1}{4\pi}\left(-\frac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta} + F_{\mu\gamma}F_\nu{}^\gamma\right) . \tag{2.9}$$

The Reissner-Nordstrom solution is

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega_2^2 \tag{2.10}$$

with

$$f(r) = 1 - \frac{2M}{r} + \frac{Q^2}{r^2} , \tag{2.11}$$

and an electromagnetic field

$$A_\mu dx^\mu = -\frac{Q}{r}dt, \qquad \text{so} \qquad F_{rt} = \frac{Q}{r^2} . \tag{2.12}$$

This component of the field strength is the electric field in the radial direction, so this is exactly the gauge field corresponding to a point source of charge Q at $r = 0$.

*See Carroll Chapter 6 for background material.The factors of 2 in that chapter are confusing; see appendix E of Wald for a consistent set of conventions similar to the ones we use here.

This is a static, spherically symmetric, charged black hole. There is nothing on the rhs of the Maxwell equation (2.8), so the charge is carried by the black hole itself; there are no charged particles anywhere. The parameter Q in the solution is the electric charge; this can be verified by the Gauss law,

$$Q_{electric} = \frac{1}{4\pi}\int_{\partial\Sigma} \star F = \frac{1}{4\pi} r^2 \int d\Omega_2 F_{rt} = Q \ . \tag{2.13}$$

This integral is over the boundary of a fixed-time slice Σ, ie a surface of constant t and constant $r \gg 1$.

Horizons and global structure

Write

$$f(r) = \frac{1}{r^2}(r - r_+)(r - r_-), \quad r_\pm = M \pm \sqrt{M^2 - Q^2} \ . \tag{2.14}$$

Then r_+ is the event horizon and r_- is the Cauchy horizon (also called the outer and inner horizon). The coordinates (2.10) break down at the event horizon, though the geometry and field strength are both smooth there. There is a curvature singularity at $r = 0$. See Carroll's textbook for a detailed discussion, and for the Penrose diagram of this black hole.

We will always consider the case $M > Q > 0$. If $|Q| > M$, then $r_+ < 0$, so the curvature singularity is not hidden behind a horizon. This is called a naked singularity, and there are two reasons we will ignore it: First, there is a great deal of evidence for the *cosmic censorship conjecture*, which says that reasonable initial states never lead to the creation of naked singularities.* Second, if there were a naked singularity, then physics outside the black hole depends on the UV (since the naked singularity can spit out visible very heavy particles), and we should not trust our effective theory anyway.

*There are also interesting violations of this conjecture in some situations, but in mild ways [Pretorious et al.]

2.3 The 1st law

Now we will check that this black hole obeys an equation analogous to (2.3), if we define an 'entropy' proportional to the area of the black hole horizon:

$$S \equiv \frac{1}{4\hbar G_N} \times \text{Area of horizon} \ . \tag{2.15}$$

We've temporarily restored the units in order to see that this is the area of the horizon in units of the Planck length $\ell_P = \sqrt{\hbar G_N}$. (Now we'll again set $G_N = \hbar = 1$.) For now this is just a definition but we will see later that there is a deep connection to actual entropy. The horizon has metric $ds^2 = r_+^2 d\Omega_2^2$, so the area is simply

$$A = 4\pi r_+^2 = 4\pi(M + \sqrt{M^2 - Q^2})^2 \ . \tag{2.16}$$

Varying the entropy gives

$$dS = \left(\frac{4\pi}{f'(r_+)}\right) dM - \left(\frac{4\pi Q}{f'(r_+) r_+}\right) dQ \ . \tag{2.17}$$

Rearranging, this can be written as the 1st law in the form

$$T dS = dM - \Phi dQ \tag{2.18}$$

with

$$T \equiv \frac{r_+ - r_-}{4\pi r_+^2} = \frac{\sqrt{M^2 - Q^2}}{2\pi(M + \sqrt{M^2 - Q^2})^2}, \quad \Phi = \frac{Q}{r_+} = \frac{Q}{M + \sqrt{M^2 - Q^2}} \ . \tag{2.19}$$

M, the mass of the black hole, is the total energy of this spacetime, so this makes sense. Φ also has a natural interpretation:

$$\Phi = -A_0|_{r=r_+} \ . \tag{2.20}$$

It is the electric potential of the horizon.

But, we have no good reason yet to call T the 'temperature' or S the 'entropy' (this will come later).

The 1st law relates two nearby equilibrium configurations. There are two ways we can think about it: *(i)* as a mathematical relation on the space of solutions to the equations, or *(ii)* dynamically, as what happens to the entropy if you throw some energy and charge into the black hole.

T is related to the *surface gravity* of the black hole

$$T = \frac{\kappa}{2\pi} \ , \tag{2.21}$$

which is defined physically as the acceleration due to gravity near the horizon (which goes to infinity) times the redshift factor (which goes to zero). If you stand far away from the black hole holding a fishing pole, and dangle an object on your fishing line near so it hovers near the horizon, then you will measure the tension in your fishing line to be κM_{object}. It can be shown that κ is constant everywhere on the horizon of a stationary black hole. This is analogous to the '0th law of thermodynamics': in equilibrium, temperature is constant.

If we restore units, then note that $S \propto \hbar$, so

$$T \propto \hbar \ . \tag{2.22}$$

Exercise: Thermodynamics of 3d Black Holes

Difficulty level: easy

Three-dimensional gravity has no true graviton, since a massless spin-2 particle has $\frac{1}{2}D(D-3) = 0$ local degrees of freedom. However, with a negative cosmological constant, there are non-trivial black hole solutions, found by Banados, Teitelboim, and Zanelli. The metric of the non-rotating BTZ black hole is

$$ds^2 = \ell^2 \left[-(r^2 - 8M)dt^2 + \frac{dr^2}{r^2 - 8M} + r^2 d\phi^2 \right] \ , \tag{2.23}$$

where ϕ is an angular coordinate, $\phi \sim \phi + 2\pi$. This is a black hole of mass M in a spacetime with cosmological constant $\Lambda = -\frac{1}{\ell^2}$.

(a) Compute the area of the black hole horizon to find the entropy.

(b) Vary the entropy, and compare to the 1st law $TdS = dM$ to find the temperature of the black hole.

(c) Put all the factors of G_N and $\hbar$ back into your formulas for S and T. S should be dimensionless and T should have units of energy (since by T we always mean $T \equiv k_B T_{thermodynamic}$). Does T have any dependence on G_N?

Exercise: Thermodynamics of rotating black holes.

Difficulty level: Straightforward, if you do the algebra on a computer

The Kerr metric is

$$ds^2 = -\frac{\Delta(r)}{\rho^2}(dt - a\sin^2\theta d\phi)^2 + \frac{\rho^2}{\Delta(r)}dr^2 + \rho^2 d\theta^2 + \frac{1}{\rho^2}\sin^2\theta(adt - (r^2+a^2)d\phi)^2 \ , \quad (2.24)$$

where

$$\Delta(r) = r^2 + a^2 - 2Mr \ , \qquad \rho^2 = r^2 + a^2\cos^2\theta \ , \quad (2.25)$$

and $-M < a < M$. This describes a rotating black hole with mass M and angular momentum

$$J = aM \ . \quad (2.26)$$

(a) Show that the entropy is

$$S = 2\pi M r_+ = 2\pi M(M + \sqrt{M^2 - a^2}) \ . \quad (2.27)$$

(b) The first law of thermodynamics, in a situation with an angular potential Ω, takes the form

$$TdS = dM - \Omega dJ \ . \quad (2.28)$$

Use this to find the temperature and angular potential of the Kerr black hole in terms of M, a. (*Hint:* The angular potential can also be defined as the angular velocity of the horizon: $\Omega = -\frac{g_{tt}}{g_{t\phi}}|_{r=r_+}$.)

2.4 The 2nd law

The second law of thermodynamics says that entropy cannot decrease: $\Delta S \geq 0$. This law does not require a quasistatic process; it is true in any physical process, including those that go far from equilibrium. (For example, if gas is confined to half a box, and we remove the partition.)

Hawking proved, directly from the Einstein equation, that in any physical process *the area of the event horizon can never decrease.* This parallels the second law of thermodynamics! This is a very surprising feature of these complicated nonlinear PDEs. We will not give the general proof; see Wald's textbook.

Exercise: Black hole collision

Difficulty level: 2 lines

The 2nd law also applies to multiple black holes. In this case the statement is that the total entropy – ie the sum of the areas of all black holes – must increase. Argue that if two uncharged, non-rotating black holes collide violently to make one bigger black hole, then at most 29% of their initial rest energy can be radiated in gravitational waves.

Exercise: Perturbative 2nd law

Difficulty level: Easy if familiar with particle motion on black holes

The 2nd law applies to the full nonlinear Einstein equation. In most cases, like a black hole collision, it is hopeless to actually solve the Einstein equations explicitly and check that it holds. But one special case where this can be done is for small perturbations of a black hole. In this exercise we will drop a charged, massive particle into a Reissner-Nordstrom black hole, and check that the entropy increases.*

Suppose we drop a particle of energy ϵ and charge q into a Reissner-Nordstrom black hole along a radial geodesic (to avoid adding angular momentum), with $\epsilon \ll M$ and

*Reference: MTW section 33.8.

$q \ll Q$. This will change the mass and charge of the black hole,

$$M \to M + \epsilon, \qquad Q \to Q + q \ . \tag{2.29}$$

Although initially there will be some fluctuations in the spacetime and ripples on the horizon from the particle that just passed through, these will quickly decay so that we have once again the Reissner-Nordstrom solution, now with the new energy and charge. Therefore, in this process the area of the black hole horizon changes according to the 1st law (2.18),

$$\delta S = \frac{1}{T}(\epsilon - \Phi q) \ . \tag{2.30}$$

(a) The infalling particle follows a trajectory $x^\mu(\tau)$ where τ is proper time. Its 4-momentum is

$$p^\mu = \frac{dx^\mu}{d\tau} \ . \tag{2.31}$$

In a spacetime with a time-translation Killing vector $\zeta^{(t)}$, the energy of a charged particle

$$\epsilon = -(p + qA) \cdot \zeta^{(t)} \ . \tag{2.32}$$

This is conserved along the path of the particle (which is not a geodesic, since it feels an electromagnetic force). For a charged particle on the Reissner-Nordstrom black hole, find ϵ in terms of $f(r)$ and the components of p^μ.

(b) Assume $Q > 0$. For one sign of q, the energy ϵ can be negative. Which sign?

If we drop a negative-energy particle into a black hole, the mass of the black hole decreases. Therefore it is possible to extract energy using this process. For uncharged but rotating black holes, a similar procedure can be used to extract energy in what is called the *Penrose process.* Particles far from the black hole cannot have negative energy, so negative-energy orbits are always confined to a region near the horizon. This region is called the *ergosphere.*

(c) Although we can decrease the energy of the charged black hole, we cannot decrease the entropy. To show this, we need to find the minimal energy of an orbit crossing the horizon. Assume the particle enters the horizon along a purely radial orbit,* $p^\theta = p^\phi =$

*This assumption is not necessary. In the general case, the particle can add angular momentum to the black hole, so we need to consider the charged, rotating Kerr-Newman spacetime. This is treated

0. The proper time along the orbit is

$$d\tau^2 = -ds^2 = f(r)dt^2 - \frac{dr^2}{f(r)} \ . \tag{2.33}$$

Use this equation to write ϵ in terms of p^r, q, and $f(r)$.

(d) ϵ is conserved along the orbit, so you can evaluate it where the particle crosses the horizon, $r = r_+$. Show that the minimal value of ϵ is

$$\epsilon_{min} = qA_t(r = r_+) \ . \tag{2.34}$$

Conclude that the 2nd law of thermodynamics is obeyed.

(e) Reversible processes are those in which $\Delta S = 0$. How would you reversibly drop a charged particle into a Reissner-Nordstrom black hole? (*i.e.,* what charge would it have and how would you drop it?)

2.5 Higher curvature corrections

Everything in this section so far has assumed the gravity action is $\int \sqrt{-g}(R - 2\Lambda)$. As discussed in section 1.1, this is incomplete: there should be higher curvature corrections suppressed by the scale of new physics.

In a general theory of gravity including curvature corrections, the formula for the entropy also receives corrections,

$$S = \frac{\text{Area}}{4} \ + \ \text{higher curvature corrections} \ . \tag{2.35}$$

The more general formula is called the 'Wald entropy'. We will postpone the general discussion of the Wald entropy until later; for now suffice it to say that the 1st law still holds. The 2nd law, however, does not. There are known counterexamples involving black hole collisions.* To my knowledge this is not fully understood. A likely explana-

in detail in [MTW section 33].

*See arXiv: [hep-th/9305016], [0705.1518], [1011.4988].

tion is that this signals a breakdown of the effective field theory — *i.e.,* that when these violations occur we must include higher corrections or corrections from new physics in the UV.

2.6 A look ahead

We have seen that the classical Einstein equations lead to laws of black hole mechanics that are analogous to the laws of thermodynamics. In quantum gravity, it is not just an analogy.

Temperature

What we called 'T' is a true temperature: black holes radiate as blackbodies with temperature T. This is Hawking radiation. It does not rely on quantizing gravity itself — it is a feature of quantum field theory in curved space, which will be derived in the next couple lectures.

Generalized second law

The entropy S is also a real entropy. This means that the total entropy of a system is the ordinary entropy (of whatever gas is present, or a cup of tea, etc) *plus* the total entropy of all the black holes in the system. The *generalized second law* is the statement that the total entropy cannot decrease:

$$S_{tot} = S_{black\ holes} + S_{stuff} \ , \qquad \Delta S_{tot} \geq 0 \ . \tag{2.36}$$

If you throw a cup of hot tea into a black hole, then this entropy seems to vanish. This is puzzling, because if we didn't know about black hole entropy, we might conclude that the ordinary 2nd law (applied to the tea) had been violated by destroying entropy. However, the generalized second law guarantees that in this process the area of the horizon will increase, and this will (at least) make up for lost entropy of the tea.

Counting microstates

Finally, we know that in quantum mechanics, entropy is supposed to count the states of a system:

$$S(E) = \log\left(\#\text{ states with energy } E\right) \ . \tag{2.37}$$

For a supermassive astrophysical black hole like the one at the center of the Milky Way, this is an enormous number, or order $\left(\frac{10^6 km}{\ell_P}\right)^2 \sim 10^{88}$. For comparison, the entropy of all baryons in the observable universe is around 10^{82}, and the entropy of the CMB is about 10^{89}. So black holes must have enormous number of states!

Classical black holes have no microstates. They are completely specified by M, J, Q (this statement is called the *no hair theorem*). How, then, can they have entropy? The answer should be that in the UV completion of quantum gravity, black holes have many microstates. This is exactly what happens in certain examples in string theory, and in AdS/CFT, as we'll see later. Unlike Hawking radiation, to understand the microscopic origin of black hole entropy requires the UV completion of quantum gravity. Turning this around, this means that black hole entropy is a rare and important gift from nature: an infrared constraint on the ultraviolet completion, that we should take very seriously in trying to quantize gravity.

3 Rindler Space and Hawking Radiation

The next couple of lectures are on Hawking radiation. There are many good references to learn this subject, for example: Carroll's GR book Chapter 9; Townsend gr-qc/9707012; Jacobson gr-qc/0308048. Therefore in these notes I will go quickly through some of the standard material, but slower through the material that is hard to find elsewhere. I strongly recommend reading Carroll's chapter too.

Hawking radiation is a feature of QFT in curved spacetime. It does not require that we quantize gravity – it just requires that we quantize the perturbative fields on the black hole background. In fact we can see very similar physics in flat spacetime.

3.1 Rindler space

2d Rindler space is a patch of Minkowski space. In 2D, the metric is

$$ds^2 = dR^2 - R^2 d\eta^2 \ . \tag{3.1}$$

There is a horizon at $R = 0$ so these coordinates are good for $R > 0, \eta =$anything.

Notice the similarity to polar coordinates on R^2, if we take $\eta \to i\phi$. This suggests the following coordinate change from 'polar-like' coordinates to 'Cartesian-like' coordinates,

$$x = R\cosh\eta, \quad t = R\sinh\eta \ . \tag{3.2}$$

The new metric is just Minkowski space $R^{1,1}$,

$$ds^2 = -dt^2 + dx^2 \ . \tag{3.3}$$

Looking at (3.2), we see $x^2 - t^2 = R^2 > 0$, so the Rindler coordinates only cover the patch of Minkowski space with

$$x > 0, \quad |t| < x \ . \tag{3.4}$$

This is the 'right wedge', which covers one quarter of the Penrose diagram.

Higher dimensional Rindler space is

$$ds^2 = dR^2 - R^2 d\eta^2 + d\vec{y}^{\,2} \ , \tag{3.5}$$

so we can map this to a patch of $R^{1,D-1}$ by the same coordinate change. The other coordinates just come along for the ride.

A 'Rindler observer' is an observer sitting at fixed R. This is not a geodesic — it is a uniformly accelerating trajectory. You can check this by mapping back to Minkowski space. Rindler observers are effectively 'confined' to a piece of Minkowski space, and they see a horizon at $R = 0$. This horizon is in many ways very similar to a black hole horizon.

Exercise: Rindler time translations are Minkowski boosts
Difficulty level: a few lines
$\zeta_{(\eta)} = \partial_\eta$ is an obvious Killing vector of Rindler space, since the metric is independent of η. By explicitly transforming this vector to Minkowski coordinates, show that it is a Lorentz boost.*

3.2 Near the black hole horizon

Black holes have an approximate Rindler region near the horizon. For example, start with the Schwarzschild solution

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega_2^2, \qquad f(r) = 1 - \frac{2M}{r} \ . \tag{3.6}$$

Make the coordinate change

$$r = 2M(1 + \epsilon^2), \quad \text{so} \quad f(r) \approx \epsilon^2 \tag{3.7}$$

*Reminder: The notation $\zeta = \partial_\eta$ means, in components, $\zeta^\mu \partial_\mu = \partial_\eta$, *i.e.*, $\zeta = \hat{\eta}$.

and expand the metric at small ϵ,

$$ds^2 = -\epsilon^2 dt^2 + 16M^2 d\epsilon^2 + 4M^2 d\Omega_2^2 + \cdots \tag{3.8}$$

The (t, ϵ) piece of this metric is Rindler space (we can rescale t and ϵ to make it look exactly like (3.1)).

Although 2d Rindler is a solution of the Einstein equations, the metric written in (3.8) (excluding the dots) is $R^{1,1} \times S^2$. This is not a solution of the Einstein equations. It is only an approximate solution for small ϵ.

3.3 Periodicity trick for Hawking Temperature

Now we will give a 'trick' to derive the Hawking temperature. The trick is to argue that that, in the black hole metric, the time coordinate must be periodic in the imaginary direction, and this imaginary periodicity implies that the black hole has a temperature. Actually, this trick is completely correct, and we will justify it later from the path integral, but don't expect this subsection to be very convincing yet!

First we want to argue that QFT at finite temperature is periodic in imaginary time, with periodicity

$$t \sim t + i\beta, \qquad \beta = 1/T \ . \tag{3.9}$$

We will return this in detail later, but for now one way to see it is by looking at the thermal Green's function*

$$G_\beta(\tau, x) \equiv -\ \text{Tr}\ \rho_{thermal} T_E \left[O(\tau, x) O(0,0)\right]\ = -\frac{1}{Z}\ \text{Tr}\ e^{-\beta H} T_E \left[O(\tau, x) O(0,0)\right], \tag{3.10}$$

where $\tau = it$ is Euclidean time, and T_E means Euclidean-time ordering (*i.e.,* put the

*This definition holds for $-\beta < \tau < \beta$. A good reference for the many types of thermal Green's functions is Fetter and Walecka, *Quantum Theory of Many-particle Systems*, 1971, in particular chapters 7 and 9. The function G_β defined here is equal to the Euclidean Green's function on a cylinder that we will discuss later (up to normalization).

larger value of τ on the left). This is periodic in imaginary time,*

$$\begin{aligned}
G_\beta(\tau, x) &= -\frac{1}{Z} \text{ Tr } e^{-\beta H} O(\tau, x) O(0,0) && (3.11)\\
&= -\frac{1}{Z} \text{ Tr } O(0,0) e^{-\beta H} O(\tau, x) && (3.12)\\
&= -\frac{1}{Z} \text{ Tr } e^{-\beta H} O(\beta, 0) O(\tau, x) && (3.13)\\
&= G_\beta(\tau - \beta, x) && (3.14)
\end{aligned}$$

Now returning to black holes, Rindler space (3.1) is related to polar coordinates, $dR^2 + R^2 d\phi^2$ with $\eta = i\phi$. Polar coordinates on R^2 are singular at the origin unless ϕ is a periodic variable, $\phi \sim \phi + 2\pi$. Therefore η is periodic in the imaginary direction,

$$\eta \sim \eta + 2\pi i \ . \tag{3.15}$$

Going back through all the coordinate transformations relating the near horizon black hole to Rindler space, this implies that the Schwarzschild coordinate t has an imaginary periodicity

$$t \sim t + i\beta \ , \qquad \beta \equiv 8\pi M \ . \tag{3.16}$$

Comparing to finite temperature QFT,

$$T = \frac{1}{\beta} = \frac{1}{8\pi M} \ . \tag{3.17}$$

This agrees with the Hawking temperature derived from the first law, (2.19) (setting $Q = 0$). As mentioned above, this derivation probably is not very convincing yet, but it is often the easiest way to calculate T given a black hole metric.

Exercise: Schwarzschild periodicity

Difficulty level: easy

Work through the coordinate transformations in section (3.2) to relate the Schwarzschild near-horizon to Rindler space, and show that (3.15) implies (3.16).

*1st line: definition (we assume $0 < \tau < \beta$, so this is τ-ordered). 2nd line: cyclicity of trace. 3rd line: definition of time translation, $O(\tau, x) = e^{\tau H} O(0, x) e^{-\tau H}$. 4th line: definition.

Exercise: Kerr periodicity

Difficulty level: moderate – a few pages

Field theory at finite temperature and and angular potential is periodic in imaginary time, but with an extra shift in the angular direction:

$$(t, \phi) \sim (t + i\beta, \phi - i\beta\Omega) \ . \tag{3.18}$$

(a) Derive (3.18) by an argument similar to (3.11). The density matrix for QFT at finite temperature and angular potential is $\rho = e^{-\beta(H-\Omega J)}$ where J is the angular momentum.

(b) Starting from the Kerr metric (11.22), we will follow steps similar to section 3.2 to relate the near horizon to Rindler space. This will be easiest if you write $\Delta(r) = (r - r_+)(r - r_-)$ and work in terms of $r_\pm$ instead of plugging in all the M's and a's. Also, you can safely ignore that θ-direction by setting $\theta = \pi/2$ (you should come back at the end of the problem and convince yourself that this was reasonable).

Plug in $r = r_+(1 + \epsilon^2)$ and expanding in ϵ. You should find something of the form

$$ad\epsilon^2 + (bd\phi - cdt)^2 + \epsilon^2(ed\phi + fdt + \cdots) + \cdots \ . \tag{3.19}$$

where a, b, c, e, f are some constants.

(c) To find the correct periodicity in this situation, define the corotating angular coordinate $\tilde{\phi} = b\phi - ct$ (this coordinate rotates with the horizon). Now the usual Rindler argument implies that t has an imaginary identification with $\tilde{\phi}$ held fixed. Translate this identification back into t, ϕ coordinates and compare to (3.18) to read off β and Ω. Check that your answers agree with the ones you derived from the first law in the exercise around (11.22).

3.4 Unruh radiation

Suppose our spacetime is Minkowski space, in the vacuum state. An observer on a worldline of fixed x will not observe any excitations. So, if the observer carries a thermometer, then the thermometer will read 'temperature = 0' for all time. More generally, if the observer carries any device with internal energy levels that can be excited by interacting with whatever matter fields exist in the theory (an *Unruh detector*), then this device will forever remain in its ground state.

However, now consider a uniformly accelerating observer with acceleration a. This corresponds to a Rindler observer sitting at fixed $R = 1/a$. We will show that in the same quantum state – the Minkowski vacuum – this observer feels a heat bath at temperature $T = \frac{a}{2\pi}$.

No unique vacuum

Thus the Minkowski vacuum is not the same as the Rindler vacuum. This is a general feature of quantum field theory in curved space (although in this example spacetime is flat!). In general, there is no such thing as *the* vacuum state, only the vacuum state according to some particular observer. The reason for this is the following. In QFT, we expand quantum fields in energy modes,

$$\hat{\phi} = \sum_{\omega>0,k} \left(a_{\omega,k} e^{i\omega t - ikx} + a^\dagger_{\omega,k} e^{-i\omega t + ikx} \right) . \tag{3.20}$$

The 'vacuum' is defined as the state annihilated by the negative energy modes:

$$a_\omega |0\rangle = 0 , \tag{3.21}$$

and excitations are created by the positive-energy modes $a^\dagger$.

The ambiguity comes from the fact that energy is observer dependent. The energy is the expectation value of the Hamiltonian; and the Hamiltonian is the operator that generates time evolution

$$\frac{i}{\hbar}[H, O] = \partial_t O . \tag{3.22}$$

Therefore the Hamiltonian depends on a choice of time t. In GR, we are free to call

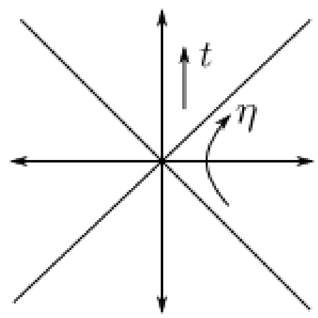


Figure 1: Different choices of time in Minkowski vs Rindler space.

any timelike direction t. Different choices of this coordinate correspond to different choices of Hamiltonian, and therefore different notions of positive energy, and therefore different notions of vacuum state. See figure.

The Unruh temperature

The quick way to the Unruh temperature is the imaginary-time periodicity trick. Since the Rindler time coordinate η is periodic under $\eta \sim \eta + 2\pi i$, this looks like a temperature $T = \frac{1}{2\pi}$. This is the temperature associated to the time translation ∂_η. In other words, if H_η is the Hamiltonian that generates η-translations, then the periodicity trick tells us that the density matrix for fields in Rindler space is

$$\rho_{Rindler} = e^{-2\pi H_\eta} \ . \tag{3.23}$$

The proper time of a Rindler observer at fixed $R = R_0$ is

$$d\tau = R_0 d\eta \ . \tag{3.24}$$

Therefore R_0 is the redshift factor, and the temperature actually observed (say by a thermometer) is

$$T_{thermometer} = \frac{1}{\sqrt{g_{\eta\eta}}}\frac{1}{2\pi} = \frac{1}{2\pi R_0} = \frac{a}{2\pi} \ . \tag{3.25}$$

Where in this argument did we actually decide which state we are in? There are excitations of Rindler/Minkowski space – say, a herd of elephants running by – where an observer will certainly not measure a uniform heat bath of thermal radiation! The answer is that by applying the periodicity trick, we have actually selected one particular very special state, (3.23). What we still need to show is that this state is in fact the Minkowski vacuum.

First, some FAQ:*

- *Does a Minkowski observer see that the Rindler observer is detecting particles?* Yes, the Minkowski observer can see that the Rindler observer's particle detector is clicking. *How does that make any sense if the Minkowski observer doesn't see the particles being absorbed?* The Minkowski observer actually sees the Rindler observer's detector *emit* a particle when it clicks. This can be worked out in detail, but the easy way to see this is that the absorption of a Rindler mode changes the quantum state of the fields; the only way to change the vacuum state is to excite something.

- *Doesn't this violate conservation of energy?* No, the Rindler observer is uniformly accelerating. So this observer must be carrying a rocket booster. From the point of view of a Minkowski observer, the rocket is providing the energy that excites the Rindler observer's thermometer, and causes Minkowski-particle emission.

- *How hot is Unruh radiation?* Not very hot. In Kelvin, $T = \hbar a/(2\pi c k_B)$. To reach $1K$, you need to accelerate at $2.5 \times 10^{20} m/s^2$.

- *According to the equivalence principle, sitting still in a uniform gravitational field is the same as accelerating. So do observers sitting near a massive object see Unruh radiation?* Only if it's a black hole (as discussed below). Observers sitting on Earth do not see Unruh radiation because the quantum fields near the Earth are not in the state (3.23). This is similar to the answer to the question 'why doesn't an electron sitting on the Earth's surface radiate?'.

Back to Minkowski

One very explicit method to show that the thermal Rindler state is in fact the Minkowski vacuum is to compare the Rindler modes to the Minkowski modes, and check that imposing $a_w^{Minkowski}|0\rangle = 0$ leads to the state (3.23) in the Rindler half-space. This is already nicely written in Carroll's GR book and many other places so I will not bother

*See http://www.scholarpedia.org/article/Unruh_effect for further discussion and references.

writing it here. Read the calculation there. But don't be fooled – that method gives the impression that the Rindler temperature has something to do with the modes of a free field in Minkowski space. This is not the case. The Rindler temperature is fixed by symmetries, and holds even for strongly interacting field theories, for example QCD. The general argument will be given in detail below, using Euclidean path integrals.

Exercise: Rindler Modes
Work through the details of section 9.5 in Carroll's GR textbook *Spacetime and Geometry*.

3.5 Hawking radiation

We have shown that Unruh observers see a heat bath set by the periodicity in imaginary time. We have also seen that the near-horizon region of a black hole is Rindler space. Putting these two facts together we get *Hawking radiation*: black holes radiate like a blackbody at temperature T. Some comments are in order:

1. This section has argued intuitively for Hawking radiation but don't be disturbed if you find the argument unconvincing. There are two ways to give a more explicit and more convincing derivation. One is to match in modes to out modes and calculate Bogoliubov coefficients; I recommend you read this calculation in Carroll's book. The other method is using Euclidean path integrals to put the imaginary-time trick on a solid footing. We will follow this latter method in the next section.

2. This is the same T that appeared in the 1st law of thermodynamics. Historically, the 1st and 2nd law were discovered before Hawking radiation. Since T and S show up together in the 1st law, it was only possible to fix each of them up to a proportionality factor. This ambiguity was removed when Hawking discovered (to everyone's surprise) that black holes actually radiate.

3. In this derivation of Hawking radiation we have chosen a particular state by applying the imaginary-time periodicity trick. In the Unruh discussion, this trick selected the Minkowski vacuum. Another way to say it is that the imaginary-time trick picks a state in which the stress tensor is regular on the past and future Rindler horizon. Therefore, by applying this trick to black holes, we have selected a particular quantum state which is regular on the past and future event horizon. This state is called the *Hartle-Hawking vacuum*. Physically, it should be interpreted as a black hole in thermodynamic equilibrium with its surroundings.

   Others commonly discussed are:

   (a) The *Boulware* vacuum is a state with no radiation. In Rindler space, it corresponds to the Rindler vacuum, $\rho_{Rindler} = |0\rangle\langle 0|$. This state is singular on the past and future Rindler/event horizons, so is not usually physically relevant.

   (b) The *Unruh* vacuum is a state in which the black hole radiates at temperature T, but the surroundings have zero temperature. This is the state of an astrophysical black hole formed by gravitational collapse. It is regular on the future event horizon, but singular on the past even horizon (which is OK because black holes formed by collapse do not have a past event horizon). If we put reflecting boundary conditions far from the black hole to confine the radiation in a box, or if we work in asymptotically anti-de Sitter space, then the Unruh state eventually equilibrates so at late times is identical to the Hartle-Hawking state.

4. If you stand far from a black hole, you will actually not quite see a blackbody. A Rindler observer sees an exact blackbody spectrum, as does an observer hovering near a black hole horizon. But far from the black hole, the spectrum is modified by a 'greybody factor' which accounts for absorption and re-emission of radiation by the intervening geometry. Far from a black hole, the occupation number of a mode with frequency ω is

$$\langle n_\omega \rangle = \frac{1}{e^{\beta\omega} - 1} \times \sigma_{abs}(\omega) \ . \tag{3.26}$$

   The first term is the blackbody formula and the second term is the greybody factor. The greybody factor is equal to the absorption cross-section of a mode with frequency ω hitting the black hole, since transmission *into* the black hole is equal to transmission *out of* the black hole.

5. In Rindler space, an observer on a geodesic (*i.e.,* Minkowski observer) falls through the Rindler horizon, and this observer does not see the Unruh radiation. Similarly, you might expect that freely falling observers jumping into a black hole will not see Hawking radiation. This is almost correct, as long as the infalling observer is near the horizon in the approximately-Rindler region, but not entirely — the potential barrier between the horizon and $r = \infty$ causes some of the radiation to bounce back into the black hole, and this can be visible to an infalling observer.

4 Path integrals, states, and operators in QFT

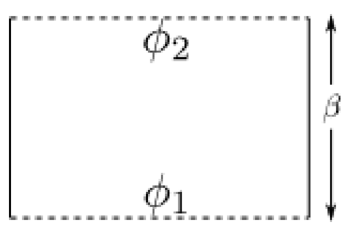
To put our derivation of Hawking radiation on a solid footing, and for other applications to gravity later on, we will now take a slight detour to explain the relationship between path integrals and states in quantum field theory. (This is material not normally covered in detail in QFT courses or books; it is assumed that the reader is already familiar with path integrals at the level of Peskin and Schröder.)

4.1 Transition amplitudes

Path integrals define transition amplitudes. A Euclidean path integral defines a transition amplitude under evolution by $e^{-\beta H}$:

$$\langle \phi_2 | e^{-\beta H} | \phi_1 \rangle = \int_{\phi(\tau=0)=\phi_1}^{\phi(\tau=\beta)=\phi_2} D\phi \; e^{-S_E[\phi]} \; . \tag{4.1}$$

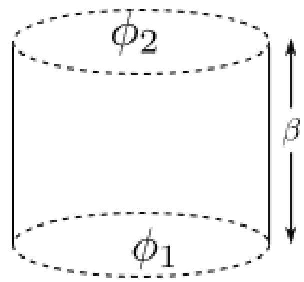
This involves a split into space and (Euclidean) time; $\phi_{1,2}$ is a boundary condition that specifies data at a fixed time. Exactly what this path integral means depends on the topology of space. If space is a plane (or line in 2d), then we depict this by

$\langle \phi_2 | e^{-\beta H} | \phi_1 \rangle =$  (4.2)

meaning it's a Euclidean path integral over an infinite strip $R^{d-1}\times$interval, with the boundary conditions shown and the interval has length β.

If space is a sphere (or circle in 2d), then the appropriate path integral is

$$\langle \phi_2 | e^{-\beta H} | \phi_1 \rangle = \tag{4.3}$$



ie it is a path integral over a cylinder $S^{d-1}\times$interval, of length β.

4.2 Wavefunctions

The transition amplitude defines the wavefunction, in the Schroedinger picture. For example the wavefunction for the state

$$|\Psi\rangle = |\phi_1(\tau)\rangle = e^{-\tau H}|\phi_1\rangle \tag{4.4}$$

is the overlap

$$\Psi[\phi_2] \equiv \langle \phi_2 | \Psi \rangle \ . \tag{4.5}$$

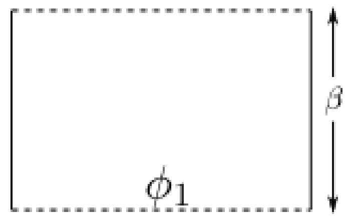
4.3 Cutting the path integral

A 'cut' is a spatial slice of the Euclidean manifold. It is a codimension-1 surface Σ. To define the transition amplitude, we specified data on two cuts, at $\tau = 0$ and $\tau = \beta$. We can formally think of a path integral with one set of boundary conditions and one open cut as a quantum state. That is, the state

$$|\Psi\rangle = e^{-\beta H}|\phi_1\rangle \tag{4.6}$$

is the path integral

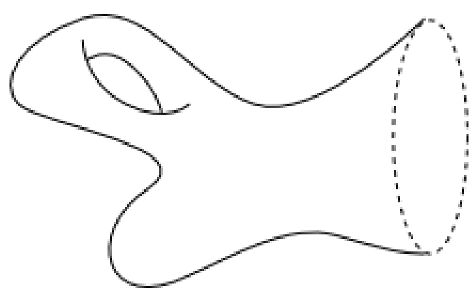
$$|\Psi\rangle = \int_{\phi(\tau=0)=\phi_1}^{\phi(\tau=\beta)=??} D\phi\ e^{S[\phi]} = \tag{4.7}$$



This is a formal object where the data on the top cut is left unspecified. It is a functional $|\Psi\rangle$ that turns field data $\langle\phi_2|$ into complex numbers $\langle\phi_2|\Psi\rangle$.

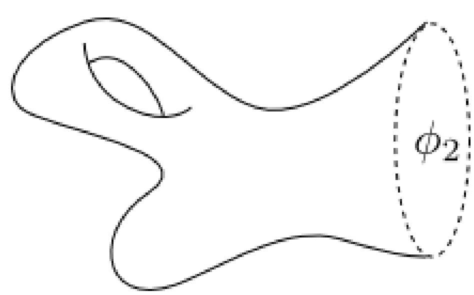
More generally, any path integral with an open cut Σ defines a quantum state on Σ. For example, this Euclidean path integral in a 2D QFT defines some particular state on a circle, $\Sigma = S^1$:

$$|X\rangle = \tag{4.8}$$



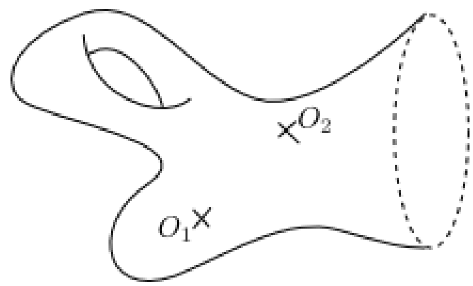
The wavefunction of this state is computed by the path integral

$$X[\phi_2] = \langle\phi_2|X\rangle = \tag{4.9}$$



We could also insert some operators into this path integral to get a different state:

$$|X'\rangle = \tag{4.10}$$

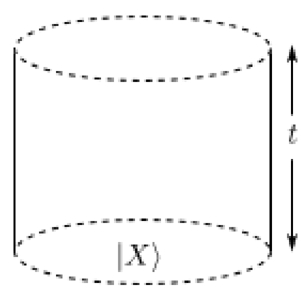


This means a Euclidedan path integral weighted by $O_1(x_1)O_2(x_2)e^{-S_E[\phi]}$, instead of just the usual $e^{-S_E[\phi]}$.

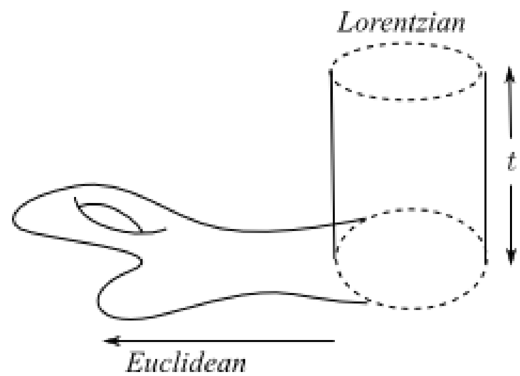
4.4 Euclidean vs. Lorentzian

So far we have discussed Euclidean path integrals. But states are states: they are defined on a spatial surface and do not care about Lorentzian vs Euclidean. The state $|X\rangle$, defined above by a Euclidean path integral, is a state in the Hilbert space of the Lorentzian theory. It is defined at a particular Lorentzian time, call it $t = 0$. It can be evolved forward in Lorentzian time by acting with e^{-iHt}, or equivalently by performing the Lorentzian path integral:

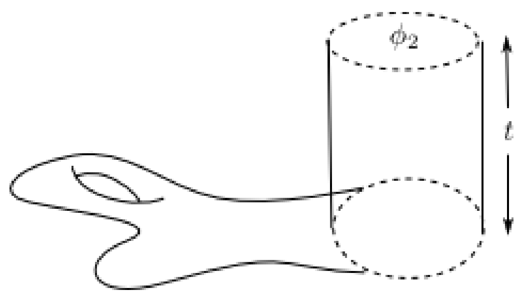
$$|X(t)\rangle = e^{-iHt}|X\rangle = \tag{4.11}$$



Since $|X\rangle \equiv |X(0)\rangle$ was defined by a Euclidean path integral, the state $|X(t)\rangle$ is a path integral that is part Euclidean, part Lorentzian:

$|X(t)\rangle =$  (4.12)

Again, this equation should be read as a formal definition of the state that tells you what path integral to perform to compute transition amplitudes:

$\langle \phi_2 | X(t)\rangle =$  (4.13)

4.5 The ground state

Evolution in Euclidean time damps excitations. Suppose we start in some state $|Y\rangle$ and expand in energy eigenstates:

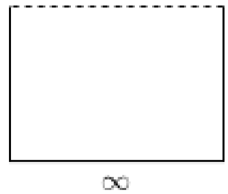
$$|Y\rangle = \sum_n y_n |n\rangle, \qquad H|n\rangle = E_n|n\rangle \ . \tag{4.14}$$

Then by evolving over a long Euclidean time we can project onto the lowest energy state,

$$e^{-\tau H}|Y\rangle \approx e^{-\tau E_0} y_0 |0\rangle \ . \qquad (\tau \to \infty) \tag{4.15}$$

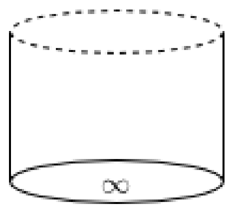
It follows that we can define the (unnormalized) ground state by doing a path integral that extends all the way to infinity in one direction. For example the ground state on the line is produced by the Euclidean path integral

$$|0\rangle_{line} = \tag{4.16}$$



This means a path integral on the semi-infinite plane, with an open cut at the edge. The ground state on a circle is produced by the path integral on a semi-infinite Euclidean cylinder,

$$|0\rangle_{circle} = \tag{4.17}$$

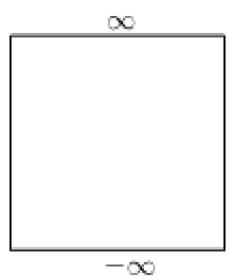


These states are unnormalized.

4.6 Vacuum correlation functions

Path integrals with cuts can be glued together to make transition amplitudes. For example, for a theory on a line, the vacuum-to-vacuum amplitude is

$$\langle 0|0\rangle = \int D\phi e^{-S_E[\phi]} = \tag{4.18}$$



The lower half-plane produces $|0\rangle$, the upper half-plane produces $\langle 0|$, and glueing them together along the cuts at $\tau = 0$ produces the transition amplitude. One way to see that we should glue is to insert the identity:

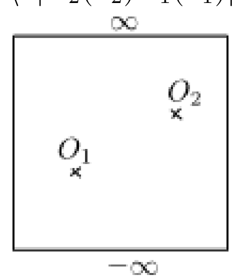
$$\langle 0|0\rangle = \sum_{\phi_1} \langle 0|\phi_1\rangle\langle\phi_1|0\rangle \ . \tag{4.19}$$

The first term is a path integral on the upper half plane; the second term is a path integral on the lower half plane; and summing over all possible boundary conditions ϕ_1 in the middle just says that fields should be continuous across $\tau = 0$ and therefore glues the half-planes together.

Expectation values of local operators are computed by similar path integrals, but with extra operator insertions. For example, correlation functions are expectation values in the vacuum state. In Euclidean signature these are computed by the path integral

$$\langle O_1(x_1) O_2(x_2) \rangle \equiv \langle 0 | O_2(x_2) O_1(x_1) | 0 \rangle \tag{4.20}$$

$$= \tag{4.21}$$

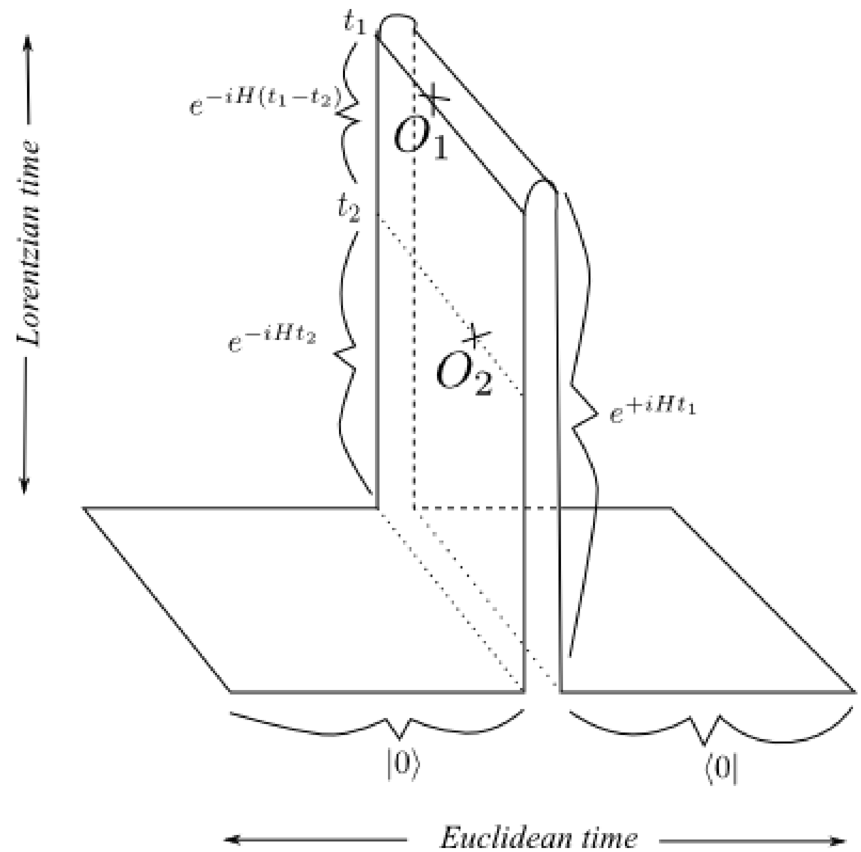


This picture means the path integral of $O_1(x_1)O_2(x_2)e^{-S_E[\phi]}$ over fields on R^d. (The ordering of operators does not matter on the lhs of (4.20), but is important on the rhs; more in this below.)

Time-ordered Lorentzian vacuum correlation functions are computed by a more complicated, 'folded' path integral that is part Euclidean and part Lorentzian. For example (assuming $t_1 > t_2$),

$$\langle O_1(t_1, \vec{x}_1) O_2(t_2, \vec{x}_2) \cdots \rangle = \langle 0 | \left(e^{iHt_1} O_1(0, \vec{x}_1) e^{-iHt_1} \right) \left(e^{iHt_2} O_2(0, \vec{x}_2) e^{-iHt_2} \right) | 0 \rangle \tag{4.22}$$

is computed by the following path integral:



(4.23)

This path integral starts at $t = -i\infty$ on the left; evolves to $t = 0$ to prepare the vacuum state; evolves in Lorentzian time to $t = t_2$, where O_2 is inserted; then evolves to t_1 where O_1 is inserted; then evolves backwards in Lorentzian time to $t = 0$; then evolves to $t = +i\infty$ for the vacuum 'bra'. Again, this picture means you should do the path integral

$$\int D\phi\ O_1(t_1, \vec{x}_1) O_2(t_2, \vec{x}_2) e^{iS[\phi]} \tag{4.24}$$

where we integrate over all fields ϕ defined on the mixed-signature manifold in the picture. The Lorentzian action appears in this expression; when you integrate over the Euclidean part of the manifold, the fact that t is imaginary will automatically change this into $e^{-S_E[\phi]}$.

We rarely need to think about doing folded path integrals like (4.23). Instead, we do one of two equivalent things: (1) We compute the Euclidean path integral with arbitrary values of the insertion points, then analytically continue to Lorentzian time, or (2) We use an $i\epsilon$ prescription to compute the Lorentzian path integral. Actually the usual $i\epsilon$ prescription is just a deformation of the integration contour (that is, integration contour in field space) shown in figure (4.23), and computes exactly the same quantity. So if you're ever wondered what you were doing with that $i\epsilon$, the answer is the figure in (4.23)!

Aside: Some comments on time ordering

In a sense, time ordering does not really exist in Euclidean signature: fields commute,

$$\langle \cdots O_1(x_1) O_2(x_2) \cdots \rangle = \langle \cdots O_2(x_2) O_1(x_1) \cdots \rangle \ . \tag{4.25}$$

One way to see this is to note that correlators computed by the path integral

$$\int D\phi \ O_1(x_1) O_2(x_2) \cdots e^{-S_E} \tag{4.26}$$

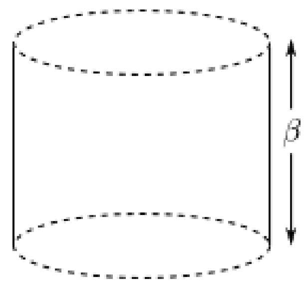
are just statistical averages, so they commute just like observables in stat-mech. Put differently, the *reason* that fields don't commute in Lorentzian signature is because the correlator is not an analytic function of the coordinates. It has branch cuts when O_2 hits the light-cone of O_1, and requires an $i\epsilon$ prescription to define the function. Different choices of $i\epsilon$ prescription give different types of correlation functions, and we denote these different choices by writing the fields in a different order. In Euclidean signature, correlators are analytic, there are not branch cuts, and there are no $i\epsilon$'s, so we don't have to worry about how fields are ordered.

However, when we cut the path integral to translate to operator language, the field operators don't commute, even in Euclidean signature. They are 'time'-ordered according to whatever slicing we choose for the path integral. So if states are defined on constant-Euclidean-time slices, the path integral translates into an operator expression with fields ordered according to their Euclidean time. If states are defined on constant-r slices (as we often do in conformal field theory), then the corresponding operator expression has radially-ordered fields.

4.7 Density matrices

A density matrix is an operator; it takes a bra and a ket, and produces a complex number. Thus any path integral with two open cuts defines a density matrix (unnormalized). For example, the density matrix $\rho = e^{-\beta H}$, for a theory on a circle, is formally the doubly-cut Euclidean path integral

$$\rho \equiv e^{-\beta H} = \tag{4.27}$$



This is just a picture representing the statement that matrix elements $\langle\phi_2|\rho|\phi_1\rangle$ are computed by the path integral with boundary conditions $\phi_{1,2}$ on the cuts.

4.8 Thermal partition function

The density matrix $\rho = e^{-\beta H}$ is the density matrix in a thermal ensemble at temperature $T = 1/\beta$. The thermal partition function is

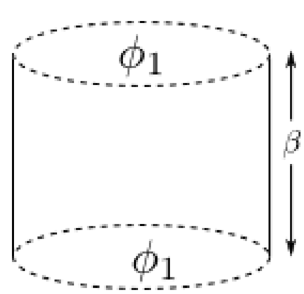
$$Z(\beta) = \mathrm{tr}\ e^{-\beta H}\ . \tag{4.28}$$

This can be represented by a Euclidean path integral as follows:

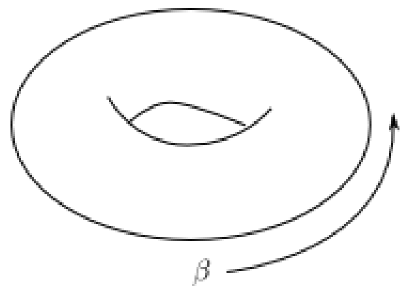
$$Z(\beta) = \mathrm{tr}\ e^{-\beta H} \tag{4.29}$$

$$= \sum_{\phi_1} \langle\phi_1|e^{-\beta H}|\phi_1\rangle \tag{4.30}$$

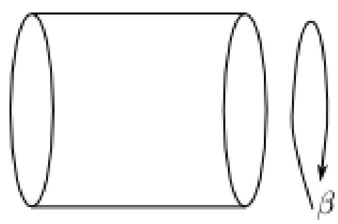
$$= \sum_{\phi_1} \tag{4.31}$$



In the last line, by summing over ϕ_1 we are really just imposing periodic boundary conditions on the cylinder. This glues together the two ends of the cylinder, producing a torus. So the thermal partition function for a 2D theory on a circle is equal to a path integral on a torus:

on circle: $Z(\beta) =$  (4.32)

Similarly, the thermal partition function for a 2D theory on a line is computed by a path integral on an infinitely long cylinder of period β:

on line: $Z(\beta) =$  (4.33)

The trace 'glues together' parts of the Euclidean manifold that computes ρ.

The same thing works in higher-dimensional QFT at finite temperature: If space is a plane R^{d-1}, the thermal partition function is the path integral on $R^{d-1} \times S^1$, and for a theory on S^{d-1}, the thermal partition function is a path integral on $S^{d-1} \times S^1$.

4.9 Thermal correlators

Equal-time correlators at finite temperature are defined (up to normalization) by

$$\langle O_1(t=0,\vec{x}_1) O_2(t=0,\vec{x}_2) \cdots \rangle_\beta \equiv \ \mathrm{Tr}\ e^{-\beta H} O_1(0,\vec{x}_1) O_2(0,\vec{x}_2) \cdots \tag{4.34}$$

By the same logic, this is computed by a path integral on a cylinder $R^{d-1} \times S^1$ (if space is a plane) or on $S^{d-1} \times S^1$ (if space is a sphere).

To compute different-time Lorentzian correlators at finite temperature, the easist method is usually to compute the Euclidean correlators first, as functions of arbitrary insertion points on the Euclidean cylinder, then analyatically continue.

Finite-temperature correlators in 2d CFT

Difficulty level: moderate, a couple pages

In a 2d conformal field theory, the 2pt function on the Euclidean cylinder of size β is fixed entirely by conformal invariance. Let $w = x + it_E$ be a complex coordinate on the Euclidean cylinder (t_E is Euclidean time and x is space). Then the 2pt function on the cylinder is

$$\langle O(w_1, \bar{w}_1) O(w_2, \bar{w}_2)\rangle_\beta = \left(\frac{1}{\sinh(2\pi(w_1 - w_2)/\beta)\sinh(2\pi(\bar{w}_1 - \bar{w}_2)/\beta)}\right)^\Delta \tag{4.35}$$

where Δ is called the scaling dimension of the operator O.

(a) Draw a picture of the path integral on the cylinder that computes (4.35).

(b) Translate your picture into operator language. Compare to (3.10). (Don't worry about the overall sign, this is a convention.)

(c) Check that the 2pt function written in (4.35) indeed has the periodicity of a thermal correlator (see discussion around (3.10)).

(d) Analytically continue to find the finite-temperature 2pt function at real (Lorentzian) times $\langle O(t_1, x_1) O(t_2, x_2)\rangle$, where t is Lorentzian time. Don't worry about which Lorentzian ordering you are computing, just pick one. (The most obvious continuation will compute the time-ordered Lorentzian correlator.)

(e) Fix $t_1 = x_1 = 0$. Draw a picture of the complex-t_2 plane showing the singularities of (4.35). When you analytically continued in part (d), you implicitly chose a contour in this plane to define the analytic continuation. Check that if (t_2, x_2) lies inside the future light-cone of (t_1, x_1), then the analytic continuation is ambiguous, due to one of the poles in the complex-t_2 plane. This ambiguity is why timelike separated fields in Lorentzian signature do not commute.

5 Path integral approach to Hawking radiation

5.1 Rindler Space and Reduced Density Matrices

We will use the Euclidean path integral to justify the claim in (3.23) that the Minkowski vacuum corresponds to the Rindler state $\rho_{Rindler} = e^{-2\pi H_\eta}$. Consider a 2d QFT on a line, and let the state of the full system by the Minkowski vacuum,

$$\rho = |0\rangle\langle 0| \ . \tag{5.1}$$

As argued above, this state is prepared by a path integral on a half-plane, cut on the line $t = 0$. Let us divide the line into $x > 0$ (region A) and $x < 0$ (region B). The reduced density matrix in region A is

$$\rho_A \equiv \mathrm{tr}_B\, \rho \ . \tag{5.2}$$

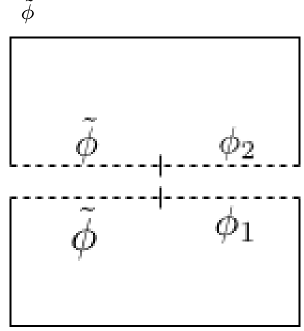
This has the nice property that all observables restricted to region A (or to the Rindler wedge that is the causal evolution of region A) can be computed from ρ_A alone:

$$\mathrm{Tr}\ \rho O(x_1)\cdots O(x_n) = \ \mathrm{Tr}\ \rho_A O(x_1)\cdots O(x_n) \ , \quad \text{for} \quad x_i > 0, |t| < x_i \ . \tag{5.3}$$

The path integral representation of ρ_A is

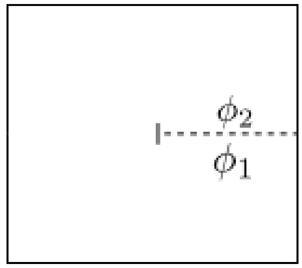
$$\langle\phi_2|\rho_A|\phi_1\rangle \ = \ \sum_{\tilde\phi}\langle\tilde\phi,\phi_2|0\rangle\langle 0|\phi_1,\tilde\phi\rangle \tag{5.4}$$

$=$ (5.5)



The upper half of this diagram corresponds to the transition amplitude $\sum_{\tilde\phi}\langle\tilde\phi,\phi_2|0\rangle$ and the lower half to the transition amplitude $\langle 0|\phi_1,\tilde\phi\rangle$. The trace sums over fields in the left Rindler wedge, which glues together these slits in the path integral, so in fact

$$\langle \phi_2 | \rho_A | \phi_1 \rangle = \qquad (5.6)$$



Now comes the key observation: we can re-slice this path integral by going to polar coordinates $dR^2 + R^2 d\phi^2$, and calling ϕ 'time'. Let $H_{Rindler}$ be the operator that generates ϕ-evolution. That is,

$$\frac{1}{\hbar}[H, O] = \partial_\phi O \qquad (5.7)$$

for any operator O. Then we can translate this *same* path integral back into operator language in a different way. That is, the path integral in (5.6) is equal to $\langle \phi_2 | e^{-2\pi H_{Rindler}} | \phi_1 \rangle$. Therefore

$$\rho_A = e^{-2\pi H_{Rindler}} \ . \qquad (5.8)$$

This looks just like a thermal state at temperature $1/2\pi$, but it is thermal with respect to the rotation generator. When we go back to Minkowski space $\phi \to i\eta$, this becomes the boost generator corresponding to the causal development of the Rindler wedge. Therefore $H_{Rindler}$ is exactly what we called H_η above.

This is a complete path-integral derivation of the statement that the Minkowski vacuum leads to a thermal state in Rindler space. As mentioned above, this can also be shown by explicit comparison of modes, but the path integral derivation can be more useful for intuition. Another big advantage is that in the path integral derivation, we did not assume anywhere that the matter fields were free, or even necessarily weakly coupled—it is completely general.

Modular Hamiltonian

The Hamiltonian that appears in the relation $\rho_{Rindler} = e^{-2\pi H_{Rindler}}$ is a special case of a *modular Hamiltonian*. A modular Hamiltonian is simply defined as the log of a density matrix (up to normalization). It is very useful for characterizing entanglement, both in quantum gravity and in condensed matter physics.

5.2 Example: Free fields

So far, we have answered the question "What is the quantum state of fields on Rindler space?" The complete answer is equation (5.8), and does not require any mention of "particles" (which only make sense at weak coupling), or any particular observer.

However to gain a more concrete intuition for the physics it is very useful to think in terms of particles. So in this subsection we will apply to result (5.8) to free (or weakly interacting) fields, and discuss what an accelerating observer capable of detecting these particles would actually experience.

A massless free field in 2D Rindler space (in Lorentz signature) obeys the wave equation

$$\Box\Phi = \nabla_\mu \nabla^\mu \Phi = 0 \ . \tag{5.9}$$

Since η is our 'time' coordinate, we take the ansatz $\Phi = e^{-i\omega\eta} f(R)$, and find the solution

$$\Phi = e^{-i\omega\eta + ik\log R}, \qquad \omega^2 = k^2 \ , \qquad \omega > 0 \ . \tag{5.10}$$

As usual in QFT, we expand the field operator in terms of creation and annihilation operators,

$$\hat{\Phi}(\eta, R) = \int dk \left(b_k \Phi_k + b_k^\dagger \Phi_k^* \right) \tag{5.11}$$

The creation operators $b^\dagger$ create positive-energy modes ϕ_k. The b's annihilate positive-energy modes. The Rindler vacuum state is defined by

$$|0\rangle_R = b_k |0\rangle_R = 0 \ , \qquad \forall k \ . \tag{5.12}$$

It is clear that this is *not* the Minkowski vacuum state: Minkowski modes are expanded in Minkowski plane waves, and Minkowski creation and annihilation operators $a_k^\dagger, a_k$ are not the same as the Rindler ones. The fact that Rindler space has a different choice of 'time' means it has a different choice of 'energy' and therefore a different notion of 'particle' and 'vacuum':

$$\text{time coordinate} \leftrightarrow \text{energy} \leftrightarrow \text{particle} \leftrightarrow \text{vacuum} \ . \tag{5.13}$$

How does this relate to our more abstract path integral discussion above? ω in the mode-expansion (5.11) is exactly the eigenvalue of the Rindler Hamiltonian $H_{Rindler}$. That is, the 1-Rindler-particle state,

$$|k\rangle_R = b_k^\dagger |0\rangle_R \tag{5.14}$$

satisfies*

$$H_{Rindler}|k\rangle_R = \omega |k\rangle \ . \tag{5.15}$$

Just like in flat space there are also multiparticle states.

In the Minkowski vacuum, the quantum state of these fields is $\rho_{Rindler} = e^{-2\pi H_{Rindler}}$. We can use this to calculate observables. For example, what is the occupation number of a mode with Rindler energy $\omega = |k|$? The calculation is identical to the usual blackbody calculation:

$$\langle n_k \rangle = \frac{1}{Z} \text{ Tr } e^{-2\pi H_{Rindler}} b_k^\dagger b_k, \qquad Z \equiv \text{ Tr } e^{-2\pi H_{Rindler}} \tag{5.16}$$

The number operator $b_k^\dagger b_k$ counts the number of quanta in the mode, so it ranges from $n = 0 \dots \infty$, so

$$\begin{aligned} \langle n_k \rangle &= \left(\sum_{n \geq 0} n e^{-2\pi n |k|} \right) / \left(\sum_{n \geq 0} e^{-2\pi n|k|} \right) \\ &= \frac{1}{e^{2\pi |k|} - 1} \ . \end{aligned} \tag{5.17}$$

This is of course the Planck blackbody spectrum.

What does an observer actually *see*?

An observer who has can detect the Φ-particle will see the blackbody spectrum (5.17). However there is one last subtlety: an observer carrying a thermometer, or a calorimeter

*For the same reason that, in flat spacetime, if we write modes $\propto e^{-i\omega t}$ then ω is the energy.

that measures energy in, say, joules, does not measure the energy ω. In fact, ω is dimensionless, since the Rindler time coordinate is dimensionless, so this wouldn't even make sense.* What an observer actually calls 'energy' is the quantity conjugate to the observer's proper time. That is, the observer will consider a mode $\sim e^{-i\omega_{obs}\tau_{obs}}$ to have energy ω_{obs}, in joules or similar energy units. The proper time of a uniformly accelerating observer with acceleration a (and therefore Rindler position $R_{obs} = 1/a$) is

$$d\tau_{obs} = R_{obs} d\eta = \frac{1}{a} d\eta \ , \tag{5.18}$$

so the observer will see a mode $e^{-i\omega\eta}$ to have energy

$$\omega_{obs} = a\omega \ . \tag{5.19}$$

Accordingly, the temperature shown on an accelerating thermometer will be

$$T_{obs} = \frac{a}{2\pi} \ . \tag{5.20}$$

Aside: Transient acceleration

Strictly speaking, our discussion of accelerating observers assumes that the observer has always been accelerating, and will continue accelerating forever. Only observers who will continue accelerating forever actually have a Rindler horizon.

However, a temporarily accelerating observer will also see Unruh radiation. It does not quite make sense to talk about a 'temperature' in this case because the observer's thermometer will not reach exact equilibrium in any finite time. When the observer starts accelerating, there will be some transient effects, and then the observer will feel thermal radiation; the thermometer will start to heat up, asymptotically approaching temperature $a/2\pi$; and when the observer stops accelerating the thermometer will again experience some transient effects, then radiate and cool back down to zero temperature.

So, as long as the acceleration lasts a long time compared to the equilibration timescale $t_{equil} \sim 1/T \sim R_0$, the Unruh temperature is still meaningful in this situation. On the other hand, for short bursts of acceleration, our analysis does not apply. Instead we would need to solve a time-dependent problem. This can be done using Feynman

*ie $ds^2 = dR^2 - R^2 d\eta^2$, so all the dimensions are carried by R; η is like an angular coordinate.

diagrams that describe emission/absorption of particles from an arbitrary worldline. (There are many wrong papers on this topic. A correct, clear, and short paper that also has a nice derivation of the Unruh effect from Green's functions is: "Transient phenomena in the Unruh effect," Bauerle and Koning.)

5.3 Importance of entanglement

What does physics look like in the Rindler vacuum, $|0\rangle_R$? To an accelerating observer, it would look ordinary: this observer would detect no particles. A geodesic observer, however, would detect observers, since this observer must notice that fields are not in the Minkowski vacuum. As long as the geodesic observer is in the Rindler wedge, this just looks like some particular excited state. However, timelike geodesics cannot stay in the Rindler wedge forever — eventually they go through the Rindler horizon. The Rindler vacuum state is singular at the horizon. That is, the energy density measured by a geodesic observer diverges at the Rindler horizon. There is no 'beyond' the horizon in this state.

This makes sense. In the Rindler vacuum, there are *no* correlation between fields in the left and right Rindler wedges:*

$$ {}_R\langle 0|\phi(x_1)\phi(x_2)|0\rangle_R = 0 \quad \text{for} \quad x_1 \in R_{left}, \quad x_2 \in R_{right} \ . \tag{5.21} $$

If there are no correlations, who's to say that these wedges are actually 'next to each other'? In a sense they are not. Thus in the vacuum state, the Rindler wedge does not extend beyond the horizon.

The key to obtaining a finite energy density on the Rindler horizon is to have a lot of entanglement between the left and right Rindler wedges. In the exercise below you will show explicitly how, in the Minkowski vacuum, the left and right Rindler wedges are maximally entangled, much like the two spins in Bell's thought experiment.† Any state with smooth horizon must be highly entangled across the horizon.

---

*In this equation $|0\rangle_R$ means the product vacuum where each Rindler wedge is in its vacuum.
†http://en.wikipedia.org/wiki/Bell's_theorem

Exercise: Entanglement warm-up

Difficulty level: easy

Consider a quantum system consisting of two particles A and B, each with two states $|0\rangle$ and $|1\rangle$ (which you can think of as spin-up and spin-down). Suppose the full system is in the maximally entangled pure state

$$|\psi\rangle = |0\rangle_A |1\rangle_B + |1\rangle_A |0\rangle_B \ . \tag{5.22}$$

(This is sometimes called a Bell pair). Find the reduced density matrix ρ_A for particle A. You will find a mixed state. Compute the *entanglement entropy* of this mixed state, defined as

$$S_A = -\operatorname{tr}_A \rho_A \log \rho_A \ . \tag{5.23}$$

Exercise: Thermofield Double

Difficulty level: conceptual

Consider a quantum system with Hilbert space $\mathcal{H}$. Any mixed state ρ can be thought of as a pure state in an enlarged system. That is, we can always add an auxiliary Hilbert space $\tilde{\mathcal{H}}$ and find a pure state

$$|\Psi\rangle \in \tilde{\mathcal{H}} \otimes \mathcal{H} \tag{5.24}$$

such that

$$\rho = \operatorname{tr}_{\tilde{H}} |\Psi\rangle\langle\Psi| \ . \tag{5.25}$$

This is called *purifying* the mixed state. In this problem you will show that Minkowski space is a purification of Rindler space.

The Minkowski Hilbert space* factorizes† into two copies of the Rindler Hilbert space,

$$\mathcal{H}_M = \tilde{\mathcal{H}}_R \otimes \mathcal{H}_R \ , \tag{5.26}$$

which are the Hilbert spaces associated to the left Rindler $x < 0$ and right Rindler

*By 'Minkowski Hilbert space' we really mean Hilbert space of the theory on an infinite plane, since Hilbert spaces are defined by the *space* on which a theory lives, not the spacetime. Similarly by 'Rindler Hilbert space' we mean the Hilbert space of the theory quantized on a half-plane.

†This is not quite true due to UV divergences, but this doesn't matter for this problem

$x > 0$.* In terms of the field data, this just means that a field in Minkowski space at $t = 0$, $\phi_M(x)$, can instead be written as the pair $(\tilde{\phi}_R, \phi_R)$ where $\tilde{\phi}_R$ is a field on the left Rindler half-plane, and ϕ_R is a field on the right Rindler half-plane.

(a) The Minkowski groundstate $|0\rangle$ is formally a functional that turns field data at $t = 0$ into complex numbers. That is, the ground state wavefunction is

$$\Psi_0[\tilde{\phi}_R, \phi_R] = {}_M\langle\tilde{\phi}_R, \phi_R|0\rangle_M \ . \tag{5.27}$$

The subscript M means 'Minkowski', ie a state on the full space. Write down the path integral that computes this wavefunction, and draw the corresponding picture along the lines of the path integral pictures above.

(b) Now re-slice this same path integral using the Rindler Hamiltonian $H_{Rindler}$, which generates Euclidean rotations ∂_ϕ. That is, write an operator expression of the form

$$\Psi_0[\tilde{\phi}_R, \phi_R] = {}_R\langle\tilde{\phi}_R| \cdots |\phi_R\rangle_R \tag{5.28}$$

and fill in the dots with an expression involving $H_{Rindler}$.

(c) We want to show that the Minkowski state is the same as the doubled Rindler state

$$|TFD\rangle_{R\otimes R} \equiv \sum_n e^{-\beta E_n/2}|n\rangle_R|n\rangle_R^* \tag{5.29}$$

where this is a sum over Rindler energy eigenstates, E_n is the Rindler energy, $\beta = 2\pi$ is the Rindler temperature, and $*$ means CPT conjugate. This is called the thermofield double state.

To demonstrate this, check that the matrix elements of the state defined in (5.29) agree with the ones you wrote above,

$${}_M\langle\tilde{\phi}_R, \phi_R|TFD\rangle_{R\otimes R} = \Psi_0[\tilde{\phi}_R, \phi_R] \ . \tag{5.30}$$

To do you this you will need to note that the mapping from Minkowski states to Rindler

*Clearly $\mathcal{H}_R = \tilde{\mathcal{H}}_R$ but the tilde will be useful to keep track of things.

$\otimes$ Rindler states is

$$|\tilde{\phi}_R, \phi_R\rangle_M \to |\tilde{\phi}\rangle_R^* |\phi_R\rangle_R \ , \tag{5.31}$$

where the conjugation is needed because time runs 'backward' in the left Rindler wedge.

What you have just shown is that the full Minkowski vacuum can be reinterpreted as the thermofield double in two copies of Rindler space.

(d) Finally, check that tracing over the left Rindler Hilbert B space produces a thermal state in the right Rindler Hilbert space A,

$$\rho_A \equiv \mathrm{tr}_B \, |TFD\rangle\langle TFD| = e^{-2\pi H_{Rindler}} = \sum_n e^{-2\pi E_n} |n\rangle_R \, {}_R\langle n| \tag{5.32}$$

This is an alternative derivation of the Unruh thermal state.

Reference: This problem is based on Maldacena's thermofield double interpretation of black holes in AdS/CFT [hep-th/0106112], which we will hopefully discuss later in the course.

Comment: Another way to approach this problem is to think of the infinite-temperature state $\sum_n |n\rangle_R |n\rangle_R^*$ as produced by a path integral over an infinitessimal strip in the Euclidean plain, positioned along the negative τ-axis. Then the thermofield double state is produced by evolving this by $\beta/4$ to the left and $\beta/4$ to the right, producing a state on the $\tau = 0$ line.

Exercise: State on an interval in 2d CFT

Difficulty level: a couple pages

In this exercise you will work out the state of a 2d conformal field theory in the vacuum, when restricted to a finite interval. This is similar to the Unruh state, but on a finite region instead of a semi-infinite plane (which in 2d would be a half-line).

At $t = 0$, we define region A to be the interval $x \in (0, \ell)$, and region B is everything else. Let z be a complex coordinate on 2d Euclidean space. The conformal mapping

$$z = -\frac{w}{w - \ell} \tag{5.33}$$

maps the half-line in the z-coordinate to the interval $x \in (0, \ell)$, where $w = x + it_E$. In the z coordinate, evolution of the half-line is generated by the rotational vector field ζ.

(a) Write ζ in the z coordinate,* then do the coordinate change (5.33) to write it in the x, t_E coordinates.

(b) The modular Hamiltonian, which generates the time evolution of the interval, is then

$$H_A = \int_A dx \; T_{t\mu}\zeta^\mu|_{t_E=0} \ , \tag{5.34}$$

where $T_{\mu\nu}$ is the usual stress tensor. Write the integrand explicitly in terms of T_{tt}, x, and ℓ.

It follows that the quantum state on the interval is

$$\rho_A = e^{-2\pi H_A} \ . \tag{5.35}$$

(c) Sketch the vector field ζ on the Euclidean (x, t_E) plane.

(d) Continue $t_E \to it$, and sketch the vector field ζ in 2d Minkowski space (x, t).

5.4 Hartle-Hawking state

We have seen there is no unique vacuum state in quantum field theory. The same is true on a black hole background. A natural state to consider, which is analogous to the vacuum state we defined in Minkowski space, is a state prepared by a path integral on the analytically continued Euclidean spacetime,

$$ds^2 = (1 - 2M/r)d\tau^2 + \frac{dr^2}{1 - 2M/r} + r^2 d\Omega_2^2 \ . \tag{5.36}$$

with the imaginary-time identification $\tau \sim \tau + \beta$. This spacetime only has $r > 0$, there is no interior. The $t = 0$ slice of the Lorentzian spacetime is the $\tau = 0$ slice of the Euclidean spacetime, see figure 2.

* *i.e.*, define $z = z_1 + iz_2$ and write ζ in terms of the two real coordinates z_1 and z_2.

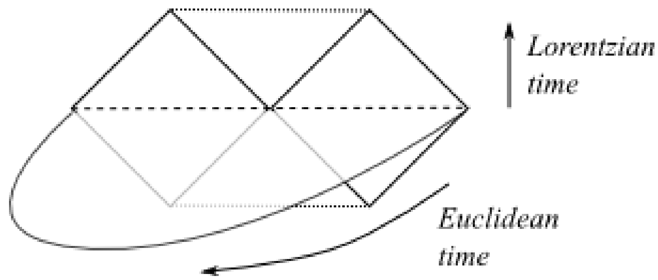


Figure 2: Schwarzschild spacetime. The Euclidean path integral produces a pure, highly entangled state on the two-sided Lorentzian spacetime. The quantum state on the right half of the Penrose diagram, where we live, is therefore mixed. This reduced state is the Hartle-Hawking thermal state.

Sending $\tau \to \tau + \beta/2$ takes us to the other side of the Penrose diagram in the maximal analytic extension of Schwarzschild. This can be shown in detail using Kruskal coordinates. A simpler way to see this is to go to Rindler coordinates near the horizon. By changing these Rindler coordinates to Minkowski-like coordinates good near the horizon, we can continue through the horizon to the other side of the Penrose diagram. So, just like in Rindler space, we get to the other side of the horizon by going half way around the Euclidean circle.

This path integral prepares an entangled state on $\tilde{M} \times M$, the product of the left and right Minkowski spaces. Just as in Rindler space, the reduced density matrix on our spacetime M will be a mixed state,

$$\rho_{HH} = e^{-\beta H} \tag{5.37}$$

where H is the ordinary Minkowski Hamiltonian associated to time translations ∂_t. This is called the 'Hartle-Hawking state.' It describes a black hole in equilibrium with a bath of radiation outside the black hole.

This is not the only state we could consider. See note 3 for a discussion of other possibilities. Hawking showed that a black hole formed by collapse will end up in the 'Unruh state', which is a state where the black hole radiates into a cold outside region.

Greybody factors

The Hartle-Hawking vacuum (5.37) is time-independent. This means that, in each mode, the flux of outgoing Hawking radiation is equal to the flux of ingoing radiation.

A mode ϕ_k outside the black hole does not necessarily fall in; it is absorbed with probability given by the absorption cross-section $\sigma_{abs}(k)$. Therefore, the only way a black hole can be in thermal equilibrium with a bath at temperature T is if the Hawking emission measured at infinite is actually

$$\langle n_k \rangle = \frac{1}{e^{\beta w} - 1} \sigma_{abs}(k) \ . \tag{5.38}$$

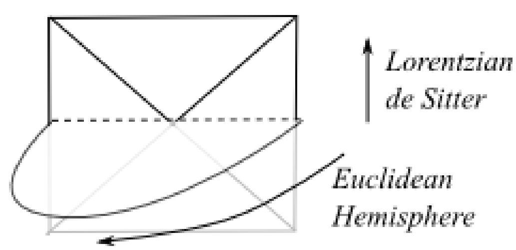
The extra factor is the 'greybody factor'. We will probably calculate some greybody factors later.

Aside: Cosmology

If the early universe is described by inflation, then it is the story of a slowly evolving de Sitter spacetime. De Sitter spacetime is the Lorentzian continuation of a sphere. That is, the metric of Euclidean de Sitter is just

$$ds^2 = d\Omega_D^2 \ . \tag{5.39}$$

The equator of the S^D is the $t = 0$ slice of global de Sitter space:



The state of quantum fields during inflation is responsible for present-day observables including the primordial temperature fluctuations in the CMB, observed by experiments like COBE, WMAP, and Planck. Since there is no unique vacuum, we must pick a state of the quantum fields in de Sitter. For various reasons,* we usually assume this state is the so-called 'Euclidean vacuum', also called the 'Bunch-Davies vacuum'

*Here are some reasons: (1) This state respects the symmetries of de Sitter; (2) At short distances, this vacuum is the one in which comoving observers see no particles (ie it coincides locally with the Minkowski vacuum); (3) at late times, due to the cosmological expansion, any state will dilute into this state.

or various other things. This state is prepared by a Euclidean path integral on the hemisphere, cut along the equator. Therefore this quantum state, unlike the Hartle-Hawking state, has quite possibly already been observed experimentally.

Exercise: Decay of Schwarzschild

Difficulty: a couple pages

(a) A black hole in asymptotically flat spacetime loses energy via Hawking radiation. If the initial mass is M, how long before the black hole radiates away completely?

(b) How heavy, in solar masses, would a black hole need to be for its lifetime to be the age of the universe $t \sim 13$ billion years?

(If such black holes exist, we might be able to observe the final moments of decay, when a large burst of energy is released in Hawking radiation. Unfortunately there is no particularly good reason to think they should exist, since black holes formed by stellar collapse must have $M_{initial} \gtrsim$ a fewM_{sun}.)

(c) What is the typical energy (in eV) of a particle emitted from a solar mass black hole via Hawking radiation?

Exercise: Superradiance

Difficulty: a few lines

Rotating (Kerr) black holes are labeled by mass M and angular momentum J, or equivalently by a temperature T and angular potential Ω. The spacetime is rotationally invariant and stationary, so modes of a scalar field can be written $\phi \sim e^{-i\omega t + im\phi} S_n(r,\theta)$, where n labels the solutions of given ω, m.* The Hawking decay rate of a rotating black hole is

$$\Gamma_{\omega,m,n} = \frac{1}{e^{\beta(\omega - m\Omega)} - 1} \sigma_{abs}(\omega, m, n) \tag{5.40}$$

(a) Take the zero-temperature limit of (5.40). (*Hint:* $\omega > 0$ and m is any integer. The

*Actually, the wave equation fully separates, so in fact $S(r,\theta) = R(r)F(\theta)$. This is surprising and nontrivial, since the background has only two Killing vectors. Similarly, the geodesic equation on Kerr has an 'extra' conserved quantity.

answer should not be trivial.)

(b) For this decay rate to make any sense, what can you conclude about σ_{abs}?

Your conclusion is a phenomenon called 'superradiance.' It is a wave analogue of the Penrose process discussed previously. In this exercise we took the Hawking formula as our starting point, but the result is entirely classical – you would reach the same conclusion by solving the wave equation on the Kerr background, and treating the black hole scattering experiment as a 1d quantum mechanics barrier transmission problem.

Superradiance very efficiently converts rest mass to radiation energy. It is believed to be responsible for the absurdly high luminosity of *quasars*: a single quasar consisting of a highly rotating black hole surrounded by infalling matter has roughly the luminosity of the entire Milky Way (10^{11} stars!).[*]

[*] More accurately, a close cousin of superradiance involving magnetic fields. The details of how this works are still largely unknown.

6 The Gravitational Path Integral

6.1 Interpretation of the classical action

In ordinary QFT, to do a path integral we first fix the spacetime manifold M, then integrate over fields defined on M. We did the same thing in our discussion of Hawking radiation. In quantum gravity, however, we must integrate over the geometry itself. We are only allowed to specify the boundary conditions on the geometry as $r \to \infty$, just like for other fields. The gravitational path integral (in Euclidean signature) is

$$Z = \int DgD\phi \, e^{-S_E[g,\phi]}, \qquad S_E[g] = -\frac{1}{16\pi G_N} \int \sqrt{g} \, (R + \cdots) + \text{boundary terms} \ , \tag{6.1}$$

where ϕ denotes all the matter fields.

The meaning of this path integral depends on the boundary conditions, as usual. In analogy to the QFT case, we define the thermal partition function $Z(\beta)$ as the path integral on a Euclidean manifold with the boundary condition that Euclidean time is a circle of proper size β,

$$t_E \sim t_E + \beta \ , \qquad g_{tt} \to 1, \qquad \text{at infinity} \ . \tag{6.2}$$

Of course we cannot actually do the path integral. In fact, we don't even really know how to define it.* The best we can do is to approximate it by expanding around a classical saddlepoint, *i.e.,* a solution of the classical equations of motion:

$$Z(\beta) \approx \exp \left(-S_E[\bar{g}, \bar{\phi}] + S^{(1)} + \cdots \right) \ . \tag{6.3}$$

The leading term, in which $\bar{g}, \bar{\phi}$ is a solution of the classical equations of motion, is the semiclassical approximation to the path integral. This solution must of course obey the correct boundary condition. The next term is the 1-loop term and is $O(G_N^0)$, and the dots indicate higher-loop contributions.

We already know a solution with the correct boundary conditions: the Euclidean

*The situation in gravity is even worse than in ordinary QFTs, since the Euclidean action is not bounded below.

Schwarzschild black hole. This is a classical saddlepoint with a Euclidean time circle of size β. Therefore, to leading approximation, the thermal free energy is the Euclidean on-shell action:

$$\log Z(\beta) \approx -S_E[\bar{g}] \ , \tag{6.4}$$

with $\bar{g}$ the Schwarzschild metric. (We have dropped $\bar{\phi}$ because no matter fields are non-zero in the Schwarzschild background.)

This partition function can be used in all of the same ways as an ordinary thermodynamic partition function. For example, recall that $\log Z = S - \beta E$, so the entropy and energy are

$$S = (1 - \beta\partial_\beta)\log Z(\beta), \qquad E = -\partial_\beta \log Z \ . \tag{6.5}$$

We will see that these agree with the area law and the black hole mass.

A similar discussion applies with an angular potential and electric potential, but we will stick to the Schwarzschild black hole to keep things simple.

Exercise: RN free energy
Difficulty: 2 lines
Using our previously calculated results for S and T (from the 1st law), and assuming energy $E = M$, find the free energy of the Reissner-Nordstrom black hole.

Exercise: RN specific heat
Difficulty: 2 lines
Compute the specific heat of Reissner-Nordstrom.

6.2 Evaluating the Euclidean action

We will now do this explicitly, in Einstein gravity (*i.e.,* no higher curvature corrections) with zero cosmological constant. It is not as simple as computing R (which vanishes for Schwarzschild!) and integrating over spacetime, since there are boundary terms to worry about and infinities to regulate.* Although this is entirely classical, the procedure to regulate divergences involves counterterms much like those in QFT; in fact we will see later there is a direct link between these two apparently different divergences.

Gibbons-Hawking-York boundary term

The Euclidean action is computed by first cutting off the spacetime at some large but fixed $r = r_0$. In the presence of a boundary we must add to the bulk Einstein action a boundary term, called the Gibbons-Hawking-York term (once again setting $G_N = 1$),

$$S_E[g] = -\frac{1}{16\pi}\int_M \sqrt{g}R - \frac{1}{8\pi}\int_{\partial M}\sqrt{h}K \ . \tag{6.6}$$

Here h_{ij} is the induced metric on the boundary ∂M, and the *extrinsic curvature* of ∂M is

$$K_{ij} \equiv \tfrac{1}{2}\mathcal{L}_n h_{ij} = \nabla_{(i}n_{j)} \ , \qquad K = h^{ij}K_{ij} \ , \tag{6.7}$$

with n is the inward-pointing unit normal to ∂M.†

The Gibbons-Hawking-York term is needed for the action to be stationary around classical solutions. The variation of the Einstein term has the schematic form

$$\delta\int_M \sqrt{g}R \sim \int_M (\text{eom})\delta g + \int_{\partial M}\left[A(g,\partial g)\delta g + B(g,\partial g)\partial\delta g\right] \ , \tag{6.9}$$

where 'eom' essentially means the Einstein tensor‡ and the boundary terms come from integrating by parts. On a classical solution, the bulk term vanishes. If we impose boundary conditions that fix the metric at $r = r_0$, then $\delta g|_{\partial M} = 0$, so the first boundary term vanishes, but the boundary term involving $\partial\delta g$ does not. The Gibbons-Hawking-York term fixes this problem. It is chosen so that the variation of the full action (6.6)

*This subsection follows Hawking's chapter in *General Relativity, an Einstein Centenary Survey*, Hawking and Ellis *eds.*

† In the simple case that the boundary is, say, at fixed r, the induced metric $h_{ij} = g_{ij}$ where i runs over the transverse directions. That is all we will need. But more generally, the projector onto ∂M is

$$h_{\mu\nu} = g_{\mu\nu} - n_\mu n_\nu \tag{6.8}$$

(as you can see by noting $n^\mu h_{\mu\nu} = 0$), and then you must define intrinsic coordinate x^i on ∂M.

‡ $(eom)\delta g \propto \sqrt{g}G^{\mu\nu}\delta g_{\mu\nu}$

has the form

$$\delta S_E[g] = \int_M (\text{eom})\delta g + \tfrac{1}{2}\int_{\partial M} \sqrt{h} T^{\mu\nu}\delta g_{\mu\nu} \ . \tag{6.10}$$

We will return to this 'stress tensor' later, but for now the important thing is just that the boundary term has been chosen to eliminate $\partial\delta g$. Thus $\delta S_E[\bar{g}] = 0$ for variations satisfying the boundary condition and $\bar{g}$ satisfying the equations of motion.

Euclidean Schwarzschild Black Hole

The Euclidean Schwarzschild solution is obtained from the ordinary Schwarzschild metric by sending $t \to -i\tau$,

$$ds^2 = \left(1 - \frac{2M}{r}\right) d\tau^2 + \frac{dr^2}{1 - \frac{2M}{r}} + r^2 d\Omega_2^2 \ . \tag{6.11}$$

What was the horizon $r = 2M$ in Lorentzian signature is now just the origin of a polar coordinate system, with angular coordinate τ identified as required for regularity at the origin,

$$\tau \sim \tau + 8\pi M \ . \tag{6.12}$$

Euclidean black holes are completely smooth solutions; they do not have an interior or a singularity.

Now we want to evaluate the action. The bulk term vanishes, since the vacuum Einstein equations set $R = 0$. The boundary term, evaluated on the surface $r = r_0$, is

$$\int_{\partial M} \sqrt{h} K = \beta(8\pi r_0 - 12\pi M) \ . \tag{6.13}$$

This is infinite as we take $r_0 \to \infty$. The procedure to regulate this divergence* is to add a 'counterterm' to the action,

$$S_E[g] = -\frac{1}{16\pi}\int_M \sqrt{g} R - \frac{1}{8\pi}\int_{\partial M} \sqrt{h} K + \frac{1}{8\pi}\int_{\partial M} \sqrt{h} K_0 \ , \tag{6.14}$$

where K_0 is the extrinsic curvature of the *same* boundary manifold ∂M, embedded in

*Caveat: this version of the procedure does not always work in asymptotically flat spacetime. As far as I know there is no entirely satisfactory and unique prescription with zero cosmological constant. Things are understood better in de Sitter space, which has finite volume, or in asymptotically anti-de Sitter space, where a similar procedure always works and plays an important role in AdS/CFT.

flat spacetime. This is very similar to what we do in quantum field theory, but this calculation is entirely classical. (We will see later that in anti-de Sitter space, there is a direct connection between the two ideas). Note that the counterterm depends only on data intrinsic to the boundary surface – it is not allowed to depend on $\partial_n h$.

To compute the counterterm, we embed the boundary metric

$$ds^2_{bdry} = (1 - 2M/r_0)d\tau^2 + r_0^2 d\Omega_2^2 \tag{6.15}$$

in flat space, by repeating the calculation for the flat geometry

$$ds^2_{subtraction} = (1 - 2M/r_0)d\tau^2 + dr^2 + r^2 d\Omega_2^2 \ . \tag{6.16}$$

This gives*

$$\int_{\partial M} \sqrt{h} K_0 = \beta(8\pi r_0 - 8\pi M + O(1/r_0)) \ . \tag{6.17}$$

This eliminates the divergence (and changes the finite term!), giving our final answer

$$S_E = \frac{\beta M}{2} = 4\pi M^2 \tag{6.18}$$

Thus the thermal partition function, or leading approximation to the path integral, is

$$Z(\beta) = \exp\left(-4\pi M^2\right) = \exp\left(-\frac{\beta^2}{16\pi}\right) \ . \tag{6.19}$$

From this we can rederive the entropy and energy using standard thermodynamics,

$$\begin{aligned} S &= (1 - \beta\partial_\beta)\log Z = 4\pi M^2 \\ E &= -\partial_\beta \log Z = M \end{aligned} \tag{6.20}$$

The entropy agrees with the area law $S = \text{Area}/4$.

Entropy and conical defects

We have just checked this for a special case, the Schwarzschild black hole, but this

*In this case you can get the same answer by just evaluating K with $M = 0$. However this does not always work. The correct procedure is to subtract the curvature of a boundary surface of identical intrinsic geometry, embedded in flat spacetime.

always works and agrees with the area law. Roughly, the reason it is proportional to area is that we can think of the equation $(1-\beta\partial_\beta)\log Z$ as calculating the change in the classical action produced by changing the imaginary-time identification. If you smoothly deform a solution, then $\delta S_E = 0$ by the equations of motion; but if you introduce a defect, this contributes $\delta S_E = \int_{defect}(\text{something})$. Going through the details, you can derive Area/4.* This is also the easiest way to derive Wald's formula, which includes the corrections to the entropy from higher curvature terms in the action.

Exercise: Schwarzschild action

Difficulty: a page or two

Derive equations (6.13) and (6.18) using (6.7).

Exercise: Euclidean methods for the BTZ black hole

Difficulty: difficult, I suspect

Evaluate the on-shell action of the Euclidean BTZ black hole obtained by Wick-rotating the metric (2.23). Check that you reproduce the correct entropy and energy. (*Caveat!* What I called M in the metric is not the energy. The energy is $E = M^2/8$.)

This calculation is similar to what we just did for asymptotically-flat Schwarzschild black holes. Note that the bulk term no longer vanishes, $R - 2\Lambda \neq 0$. The full action, including the counterterm, is

$$S_E[g] = -\frac{1}{16\pi}\int_M \sqrt{g}(R-2\Lambda) - \frac{1}{8\pi}\int_{\partial M}\sqrt{h}K + \frac{a}{8\pi}\int_{\partial M}\sqrt{h}\ . \tag{6.21}$$

Choose a to cancel the divergence; the remaining finite expression is the correct S_E.

Reference: [Balasubramanian and Kraus, hep-th/9902121].

Comment: The counterterm depends on the dimensionality of spacetime. The simple counterterm in (6.21) only works in AdS_3. In higher-dimensional AdS, there are more available counterterms, for example $\int_{\partial M} R[h]$ (the intrinsic boundary curvature), and these are also required to cancel all divergences.

*The clearest reference I know for this is section 3.1 of Lewkowycz and Maldacena, 1304.4926.]

7 Thermodynamics of de Sitter space

Inflation is the idea that the very early universe, at $t \lesssim 10^{-32}s$, is approximately de Sitter spacetime. We will now take a detour to apply our methods to de Sitter space, since in this situation they produce observable effects via the imprint of primordial density perturbations on the CMB.

de Sitter is the maximally symmetric solution of the Einstein equations with a positive cosmological constant. The metric is*

$$ds^2 = -dT^2 + \ell^2 \cosh^2(T/\ell) d\Omega_{D-1}^2 \qquad \text{(global coordinates)} \ . \tag{7.1}$$

These coordinates, which cover all of de Sitter space, describe a sphere S^{D-1} that is very large at $T \to -\infty$, contracts to a minimum radius ℓ at $T = 0$, then expands as $T \to \infty$. The length scale is set by the cosmological constant, $\Lambda \sim 1/\ell^2$.

To draw the Penrose diagram, we need to make the range of T finite. This can be done by defining

$$\tan(\eta/2) = \tanh(T/2\ell) \ . \tag{7.2}$$

In these coordinates, which run over $\eta \in (-\frac{\pi}{2}, \frac{\pi}{2})$, the metric is

$$ds^2 = \frac{\ell^2}{\cos^2 \eta} \left(-d\eta^2 + d\Omega_{D-1}^2 \right) \ . \tag{7.3}$$

For the Penrose diagram (which captures the causal structure but ignores distances) we can ignore the overall factor, so this gives the diagram in figure 3. Note that de Sitter space has an initial and final conformal boundary. (Although the diagram also appears to have left and right boundaries, these are not really boundaries – at each value of η space is a sphere, so those lines are just the north and south poles of the sphere S^{D-1}.)

Vacuum

As usual, there is no unique vacuum. However we can follow our usual prescription,

*I will call these global coordinates, but in the context of cosmology this is usually called the 'closed slicing.'

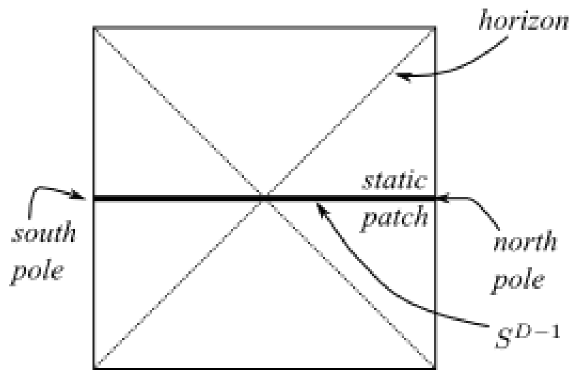
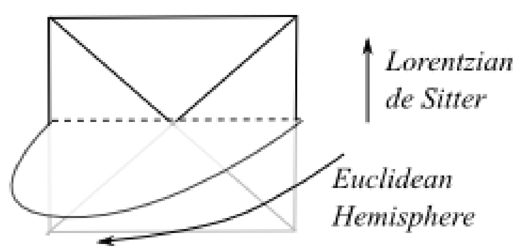


Figure 3: Penrose diagram of de Sitter space.

as we did for Minkowski and Schwarzschild, and define a vacuum which is prepared by a path integral on the Euclidean continuation of de Sitter. Euclidean de Sitter space is just a sphere: send $T \to i\ell\theta$ in (7.1), and you see that $\theta \in [-\pi/2, \pi/2]$ is the polar coordinate on a sphere S^D of radius ℓ. It is shifted by $\pi/2$ from the usual definition, so the surface $T = \theta = 0$ is the equator of the S^D. This equator is the minimal-size spatial section of de Sitter space, an S^{D-1} of radius ℓ. This is the $T = 0$ surface at the middle of the Penrose diagram.

The *Euclidean vacuum*[*] is the state prepared by a path integral on the hemisphere. This path integral prepares a quantum state on the equator, which can then be evolved in Lorentzian time. Here is a cartoon for this path integral:



(7.4)

Roughly speaking, 'most' states will eventually end up close to the Euclidean vacuum, since the de Sitter expansion dilutes any excitations.

This vacuum state has in a sense been observed experimentally, since it leaves an imprint in the CMB. I say 'in a sense' because it is not a very fine-grained or direct test of the details of de Sitter space, but it is nonetheless the only good explanation

[*]Also called the 'Bunch-Davies vacuum' and various other things.

we have for those fluctuations.

7.1 Vacuum correlators

Now we will calculate the Gaussian fluctuations in de Sitter that are seen in the CMB, using path integral methods.* We consider a free massless scalar field, in the 'conformal coordinates' (7.3) in $D = 4$. (It turns out that scalar fluctuations of the metric reduce to this problem.) The action is

$$I \sim \int_M \sqrt{-g}\nabla_\mu\phi\nabla^\mu\phi \ . \tag{7.5}$$

We want to compute the wavefunction for this scalar field, in the Euclidean vacuum. This wavefunction is defined to be the transition amplitude

$$\Psi[\phi_0; t_0] = \langle\phi_0|e^{-iHt_0}|0\rangle \ . \tag{7.6}$$

This is a path integral over field configurations on the mixed spacetime (7.4). The past boundary condition is regularity on the Euclidean sphere; the future boundary condition is $\phi(\Omega, t_0) = \phi_0(\Omega)$.

In the WKB approximation the wavefunction is simply the on-shell action of a classical solution satisfying these boundary conditions,

$$\Psi[\phi_0] \sim e^{iI[\phi]} \ . \tag{7.7}$$

In fact, since this is a free field, this expression is exact. The on-shell action is a pure boundary term, since we can integrate by parts in (7.5) then use the equations of motion. Thus

$$I_{on-shell}(t_0) \sim \int_{\Sigma(t_0)} \sqrt{h}\phi\partial_n\phi \tag{7.8}$$

where $\Sigma(t_0)$ is the spatial slice at time t_0, ∂_n is the derivative normal to this slice, and h is the spatial metric.

At late times, the S^{D-1} spatial sphere in (7.3) is very large. When we look back at the

*This is based on Maldacena astro-ph/0210603.

CMB, we are looking at a tiny patch of this sphere. Therefore at late times we can think of the spatial sphere as being essentially flat. That is, we can only see modes with wavelength much smaller than ℓ. This simplifies the calculation of the wavefunction because at late times we can approximate the metric by

$$ds^2 = \frac{\ell^2}{\eta^2}(-d\eta^2 + d\vec{x}^2) \ . \tag{7.9}$$

This is enough to compute the wavefunction, since (7.8) is a pure boundary term at $t = t_0$ ($\eta = \eta_0$).

For a plane wave $\phi_0 = h_k(\eta)e^{i\vec{k}\vec{x}}$, the wave equation at late times (*i.e.*, using the metric (7.9)) is

$$\Box\phi = 0 \qquad \Rightarrow \qquad \partial_\eta^2 h_k(\eta) - \frac{2}{\eta}\partial_\eta h_k(\eta) + \vec{k}^2 h_k(\eta) = 0 \ . \tag{7.10}$$

In order to find the wavefunction, we need to find the classical solution satisfying the boundary conditions stated above. This is

$$\phi = \phi_k^0 \frac{(1 - ik\eta)e^{ik\eta}}{(1 - ik\eta_0)e^{ik\eta_0}} \ . \tag{7.11}$$

This satisfies all of our criteria: it solves the wave equation, and it is equal to ϕ_k^0 at time $\eta = \eta_0$. The last criterion was regularity on the Euclidean sphere – this is what picks the solution going as $e^{+ik\eta}$, rather than the other solution $\sim e^{-ik\eta}$. To see this, note the pole of the Euclidean sphere is $t = i\ell\pi/2$, which corresponds to

$$\eta = 2\tan^{-1}(i\pi/4) = i\infty \ . \tag{7.12}$$

Only the $e^{+ik\eta}$ solution is regular in this limit, so we've picked the correct solution, corresponding to the Euclidean vacuum state. (This is very similar to the path-integral calculation of the groundstate for the harmonic oscillator.) Although ϕ is a real field, (7.11) is complex. This is fine: even though the path integral is over real field configurations, the stationary phase approximation can pick out a complex saddlepoint. It is just a trick to compute the path integral.

Plugging (7.11) into (7.8),

$$iI = i \int \frac{d^3k}{(2\pi)^3} \frac{\ell^2}{2\eta_0^2} \phi^0_{-k} \partial_\eta \phi^0_k |_{\eta=\eta_0} \sim \int \frac{d^3k}{(2\pi)^3} \frac{\ell^2}{2} \left[i\frac{k^2}{\eta_0} - k^3 + \cdots \right] \phi^0_{-k} \phi^0_k \tag{7.13}$$

This gives the wavefunction (exactly in a free theory) via (7.7). Knowing the wavefunction, we can calculate correlators using $\langle \phi^2 \rangle \sim \int D\phi \phi^2 |\Psi(\phi)|^2$. In detail, accounting for the normalization of the wavefunction:

$$\begin{aligned} \langle \phi_k \phi_{-k} \rangle &= \frac{\int D\phi\, \phi_k \phi_{-k} \exp\left(- \int \frac{d^3p}{(2\pi)^3} \ell^2 p^3 \phi_p \phi_{-p} \right)}{\int D\phi \, \exp\left(- \int \frac{d^3p}{(2\pi)^3} \ell^2 p^3 \phi_p \phi_{-p} \right)} \\ &= (2\pi)^3 \frac{1}{2\ell^2} k^{-3} \ , \end{aligned} \tag{7.14}$$

where we used the Gaussian integral $\int dz z^2 e^{-az^2} / \int dz e^{-az^2} = \frac{1}{2a}$ (and we've dropped an overall momentum-conserving delta function).

The power law in (7.14) is what is meant by the statement that 'inflation predicts a scale-invariant spectrum of primordial scalar perturbations.'

7.2 The Static Patch

A single inertial observer travels on a geodesic which we might as well call the north pole of S^{D-1}. Thus the worldline of an observer is the solid line on the right side of the Penrose diagram. It is clear from the diagram that this observer is in causal contact with only a subregion of de Sitter. This is because the universe is expanding very rapidly, so you cannot communicate with someone beyond a certain critical distance: the cosmological horizon. This is labeled 'horizon' in figure 3. Like Rindler space, the position of the horizon depends on the observer.

The *static patch* is the region of de Sitter in causal contact with an observer sitting at the north pole. This is the analogue of the Rindler patch. The coordinates on the static patch are

$$ds^2 = -(1 - r^2/\ell^2) dt^2 + \frac{dr^2}{1 - r^2/\ell^2} dr^2 + r^2 d\Omega^2_{D-2} \ . \tag{7.15}$$

This looks like a black hole except, the static patch where an observer lives is the *inside*, $r < \ell$, with a cosmological horizon at $r = \ell$.

Temperature

In the Euclidean vacuum, an observer in the static patch will see a temperature.* This is the same reason we discussed above for Rindler space and then for the Schwarzschild spacetime in the Hartle-Hawking state. The Euclidean path integral prepares a state entangled between the left and right static patches, and when we trace over the hidden region, the resulting density matrix is thermal. To find the temperature we can apply the imaginary time periodicity trick. Starting from the static patch coordinates, define $r = \ell(1 - \epsilon^2)$ and expand in ϵ to find:

$$ds^2 \approx 2\ell^2(d\epsilon^2 - \frac{\epsilon^2}{\ell^2}dt^2) + \cdots \tag{7.16}$$

This looks like Rindler, so the Euclidean continuation is regular only if $t \sim t + 2\pi i\ell$. Therefore the de Sitter temperature is

$$T_{dS} = \frac{1}{2\pi\ell} \ . \tag{7.17}$$

This is the temperature that you will read on your thermometer if you are an inertial observer in de Sitter. (Due to dark energy we are now at the start of a new de Sitter epoch. The present-day de Sitter temperature is the Hubble scale, $T \sim 10^{-33} eV$.)

Entropy

The area of the cosmological horizon is

$$\text{Area} = \ell^2 Vol(S_{D-2}) \tag{7.18}$$

So in $D = 4$, S =area/4 gives

$$S_{D=4} = \pi\ell^2 \ . \tag{7.19}$$

*See Gibbons and Hawking, "Cosmological event horizons, thermodynamics, and particle creation," 1977.

7.3 Action

The Euclidean action of dS_D is the action of S^D. This is finite and there are no boundary terms to worry about on a sphere, so it's straightforward to calculate. In $D = 4$, the Euclidean action is*

$$I_E = -\frac{1}{16\pi}\int_{S^D}\sqrt{g}(R-2\Lambda) = -\frac{\ell^4}{16\pi}\left(\frac{12}{\ell^2}-\frac{6}{\ell^2}\right)Vol(S_4) = -\pi\ell^2 \ . \tag{7.20}$$

Curiously, this is minus the entropy, *i.e.,* the entropy is

$$S = \log Z \ . \tag{7.21}$$

Here's why: Recall the thermodynamic identity $\log Z = -\beta F = -\beta E + S$. The energy in GR is a pure boundary term (see next lecture!), so a compact space has $E = 0$. Thus thermodynamics predicts $\log Z = S$, and that's exactly what we found in de Sitter.†

de Sitter is mysterious

We will not say much more about de Sitter space in this course. A big reason for this is that we don't have any UV-complete theory of gravity in de Sitter, like we do in anti-de Sitter. We also have no clear answer to the question 'What is the de Sitter entropy?' (Does it count the microstates of something?) Since we live in de Sitter, this seems like a very important question.

Exercise: de Sitter in general dimensions

Difficulty level: straightforward, if using Mathematica to compute curvatures

*Given the proliferation of things called 'S' I will start using I for action

† *Caveat:* In this calculation $T = 1/(2\pi\ell)$, with ℓ a fixed parameter, so the temperature is not actually a tunable variable. We can't really make sense of thermodynamics if the temperature is not a free variable. To really make sense of the discussion in this section, we need to add some additional matter to de Sitter, which allows us to tune the temperature. Then there is a true 1st law, with the temperature a tunable variable. This done in for example the nice de Sitter review hep-th/0110007. You can think of what we've described here as the limit of that calculation where the extra energy is set to zero.

Compute the entropy and the on-shell action of D-dimensional de Sitter space, and verify the relation $S = \log Z$.

---

8 Symmetries and the Hamiltonian

Throughout the discussion of black hole thermodynamics, we have always assumed energy $= M$. Now we will introduce the Hamiltonian formulation of GR and show how to define conserved charges associated to spacetime symmetries. The energy is a special case, associated to time-translation symmetry. There are quicker ways to reach the conclusion energy $= M$ (see Carroll's book), but we will take the more careful route because it's useful later.

8.1 Parameterized Systems

[*References:* The original paper is very nice and still worth reading, especially sections 1-3: "The Dynamics of General Relativity" by Arnowitt, Deser, Misner (ADM), 1962 (but available on arXiv at gr-qc/0405109). See also appendix E of Wald's textbook, and for full detail see Poisson's *Relativist's Toolkit* chapter 4.]

Time plays a special role in the canonical formulation of quantum mechanics, and in the Hamiltonian approach to classical mechanics, since it is the independent variable. In GR, time t is just an arbitrary parameter, and the dynamics are reparameterization-invariant under $t \to t'(t)$, since this is just a special case of diffeomorphisms. To see how this fits into Hamiltonian mechanics we first consider a simple analog in quantum mechanics.

Suppose we have a system with a single degree of freedom $q(t)$ with conjugate momentum p, and action

$$I = \int dt L. \tag{8.1}$$

The Hamiltonian is the Legendre transform

$$H(p,q) = p\dot{q} - L(q,\dot{q})|_{p=\partial L/\partial \dot{q}} \ . \tag{8.2}$$

Hamilton's equations of motion are

$$\frac{dp}{dt} = -\frac{\partial H}{\partial q} \ , \qquad \frac{dq}{dt} = \frac{\partial H}{\partial p} \ . \tag{8.3}$$

The independent variable t is special. It labels the dynamics but does not participate as a degree of freedom. In GR, time is just an arbitrary label – it is not special, and the theory is invariant under time reparameterizations. To mimic this in our simple system with 1 dof, we will introduce a fake time-reparameterizations symmetry. To do this we label the dynamics by an arbitrary parameter τ, and introduce a physical 'clock' variable T, treated as a dynamical degree of freedom. So instead consider the system of variables and conjugate momenta

$$q(\tau), \quad p(\tau), \quad T(\tau), \quad \Pi(\tau) \tag{8.4}$$

where Π is the momentum conjugate to T. This is equivalent to the original 1 dof if we use the 'parameterized' action

$$I' = \int d\tau (pq' + \Pi T' - NR) \ , \qquad R \equiv \Pi + H(p,q) \ , \tag{8.5}$$

where prime $= d/d\tau$. Here $N(\tau)$ is a Lagrange multiplier, which enforces the 'constraint equation'

$$\Pi + H(p,q) = 0 \ . \tag{8.6}$$

The action (8.5) is reparameterization invariant under $\bar{\tau} = \bar{\tau}(\tau)$, since after all τ is just a label that we invented. The Hamiltonian of the enlarged system is simply

$$H' = N(\Pi + H(p,q)) \ , \tag{8.7}$$

so it vanishes on-shell due to the constraint equation!

To recap: we introduced time-covariance by adding a fake degree of freedom, and imposing a constraint. The resulting Hamiltonian vanishes on-shell, because it generates τ-translations, which is just part of the reparameterization symmetry.

To reverse the procedure, *i.e.,* to go from the parameterized action back to the ordinary action with 1 dof, we plug in the constraint

$$I' = \int d\tau \, [pq' - H(p,q)T'] \tag{8.8}$$

and then rewrite the dynamics in terms of the clock variable:

$$I' = \int dT \, [p\dot{q} - H(p,q)] \tag{8.9}$$

where dot $= d/dT$. So we see that T is just the original physical time t.

The equation of motion for T is

$$T' = N \frac{\partial}{\partial T}(\Pi + H) \tag{8.10}$$

But both T' and N are unspecified by the dynamics. For example we are free to pick the 'gauge condition' $T = \tau$, which corresponds to some particular choice of $N(\tau)$.

8.2 The ADM Hamiltonian

GR is already a 'parameterized system:' the t coordinate is like our τ coordinate above, and we will see that the Hamiltonian is very much like (8.7).

The canonical variables are

$$h_{ij}(\vec{x},t), \qquad \pi_{ij}(\vec{x},t) \tag{8.11}$$

where h_{ij} are the space components of the metric, and π_{ij} are their canonical conjugates. The full spacetime metric is parameterized as

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \ . \tag{8.12}$$

$N = \sqrt{-1/g^{tt}}$ is called the 'lapse' and $N^i = N^2 g^{ti}$ is the 'shift'. These are Lagrange multipliers, just like N in our discussion above. They are not fixed by the dynamics, but a choice of parameterization. In other words, *any geometry can be sliced into 'time' and 'space' in a such a way that N and N^i can be set to any functions you like.* They are called the lapse and shift because they correspond to our choice of how our coordinates on a time-slice of fixed $t = t_0$ are related to the coordinates on a time-slice of fixed $t = t_0 + \delta t$. The flow vector, which tells you the arrow of time from one slice to the next, is*

$$\zeta^\mu = N u^\mu + N^\mu \ . \tag{8.13}$$

In the coodinates (8.12), $\zeta = \partial_t$, but we will treat N and N^a as arbitrary parameters.

The action of GR, discussed above but now in Lorentzian signature, is

$$I = \frac{1}{16\pi}\int_M d^4x \sqrt{-g}R - \frac{1}{8\pi}\int_{\partial M} d^3x \sqrt{-\gamma}(K - K_0) \tag{8.14}$$

where K_0 is the subtraction term (the extrinsic curvature of the boundary embedded in flat spacetime). Recall that the on-shell variation is

$$\delta I_{on-shell} = \tfrac{1}{2}\int_{\partial M} d^3x \sqrt{-\gamma} T^{ij}\delta g_{ij}, \tag{8.15}$$

where the 'boundary stress tensor' (aka Brown-York stress tensor) is

$$T^{ij} = \frac{1}{8\pi}\left(K^{ij} - \gamma^{ij}K\right) - \text{background subtraction} \ . \tag{8.16}$$

After quite a bit of work†, the full off-shell action (8.14) can be written

$$I = \int_M d^4x \left[\pi^{ij}\dot{h}_{ij} - N\mathcal{H} - N^i\mathcal{H}_i\right] - \int_{\partial M} d^3x \sqrt{\sigma} u^\mu T_{\mu\nu}\zeta^\mu \ , \tag{8.17}$$

From here we can read off the Hamiltonian‡

$$H[\zeta] = \int_\Sigma d^3x \left(N\mathcal{H} + N^i\mathcal{H}_i\right) + \int_{\partial\Sigma} d^2x \sqrt{\sigma} u^\mu T_{\mu\nu}\zeta^\nu \ , \tag{8.18}$$

which is an integral over a spatial slice Σ.

Now to explain all these terms: $\mathcal{H}$ and $\mathcal{H}_i$ are called the Hamiltonian and momentum constraints, which are essentially the G_{00} and G_{0i} components of the Einstein equations (see Wald for explicit formulae). These components of the equations of motion involve only 1st time derivatives. They are called 'constraints' because if we think of GR as an initial value problem – specify initial data, then evolve in time according to the

*Here $N^\mu = h^\mu_a N^a$, where $h_{\mu\nu} = g_{\mu\nu} + u_\mu u_\nu$, *i.e.,* h^μ_a is the projector onto a spatial slice.

†See Brown and York, "Quasilocal energy and conserved charges derived from the gravitational action, " 1993, and also Poisson's *Relativist's Toolkit*, Chapter 4.

‡The bulk term is called the 'ADM Hamiltonian'. As far as I know, the boundary terms were first derived by Brown and York, and by Hawking and Horowitz.

dynamical equations – these are constraints on the allowed initial data $h_{ij}, \dot{h}_{ij}$ at $t = 0$. This is in contrast to other, dynamical equations of motion, which tell you how that data evolves in time.* Finally u^μ is the timelike unit normal on the boundary, with $u^2 = -1$.

A few remarks about our final answer (8.18):

- The bulk term vanishes on-shell due to the constraint equations. The boundary term does not vanish in general. This is related to the fact that diffeomorphisms acting on the boundary are 'real' dynamics, whereas diffeomorphisms away from the boundary are just redundancies.

- We have written the Hamiltonian as a functional of the lapse and shift, since the dynamics leaves ζ unspecified. This corresponds to a choice of time evolution. That is, the Dirac bracket† of the Hamiltonian with any function X of the canonical variables is

$$\{H[\zeta],\ X\} = \mathcal{L}_\zeta X \ . \tag{8.19}$$

  If we choose, for example, $\zeta^\mu = (1, 0)$, then this Hamiltonian generates time evolution in the t-direction.

- The on-shell Hamiltonian looks just like the Hamiltonian of a 3-dimensional theory living on the boundary with a 3-dimensional stress tensor $T_{\mu\nu}$. We will see that at least in AdS this is actually literally the case.

Exercise: Constraints in electrodynamics

Difficulty level: medium

Derive the Hamiltonian of electrodynamics. Start from the action $I = -\frac{1}{4}\int d^4x F_{\mu\nu}^2$,

*In electrodynamics, the action involves only first derivatives of A_t, so this is a Lagrange multiplier like the lapse in GR. The Hamiltonian of electrodynamics has a term $A_t C$ where $C = \nabla \cdot E - \rho_{matter}$ is the Gauss constraint.

†The Dirac bracket is the Poisson bracket, but accounting for gauge symmetries which modify the bracket acting on physical fields. The Dirac bracket is what becomes a commutator in the quantum theory.

identify the canonical coordinates and conjugate momenta, and rewrite it like we did for gravity in (8.17). Identify the Lagrange multiplier(s) and constraint(s).

In gravity, we gave a physical interpretation of the lapse and shift Lagrange multipliers as a choice of foliation of spacetime. What is the analogous interpretation of A_t in electrodynamics? (It might be useful to couple to a matter field to answer this.)

Reference: Appendix E of Wald. But write your answers in terms of the vector potential, not $\vec{E}$ and $\vec{B}$.

8.3 Energy

As usual, the numerical value of the Hamiltonian, evaluated on a solution, is the energy. In GR we must specify a lapse and shift to define the Hamiltonian. The energy is associated to time translations, so we identify the energy as the Hamiltonian with $N \to 1$ and $N^a \to 0$ at the boundary. For this choice, the surface deformation vector $\zeta_{(t)}$ has components $\zeta^\mu_{(t)} = (1, 0, 0, 0)$, so

$$E \equiv H[\zeta_{(t)}]|_{on-shell} = \int_{\partial\Sigma} d^2x \sqrt{\sigma} u^i T_{it} \ . \tag{8.20}$$

This is the usual (covariantized) expression for the energy in terms of the stress tensor.*

We must impose boundary conditions to ensure that energy is conserved. It can be shown that the $G_{r\mu}$ components of the Einstein equations† are

$$\nabla_i T^{ij} = -n_\alpha T^{\alpha j}_{matter} \ , \tag{8.21}$$

where n is the spacelike unit normal to the boundary. Therefore, if we impose the boundary condition that matter fields go to zero fast enough as $r \to \infty$, then the

*This equation agrees with other definitions of energy you may have seen, like the Komar formula, whenever those definitions apply.

†*i.e.,* the 'constraints' in a radial slicing of the spacetime, which contain only first order r-derivatives.

boundary stress tensor is conserved,

$$\nabla_i T^{ij} = 0 \ . \tag{8.22}$$

If in addition

$$\nabla_{(i}\zeta_{j)} = 0 \qquad \text{as} \qquad r \to \infty \tag{8.23}$$

then the energy current $j^i = T^i{}_j \zeta^j$ is conserved, $\nabla_i j^i = 0$. In this case the energy is independent of what slice Σ we choose to evaluate (8.20):

$$E(\Sigma) - E(\Sigma') = \int_{\partial\Sigma} d^2x\sqrt{\sigma}u^i T_{it} - \int_{\partial\Sigma'} d^2x\sqrt{\sigma}u^i T_{it} = \int d^3x\sqrt{-\gamma}\nabla_i \left(T^{ij}\zeta_j\right) = 0 \ . \tag{8.24}$$

The equation (8.23) is the Killing equation, so the conclusion is that energy is conserved as long as (i) matter fields fall off fast enough near infinity, and (ii) $\zeta = \partial_t$ is an *asymptotic Killing vector*.

What about matter?

The expression (8.20) includes the contribution from matter. The constraints ensure that the metric at infinity knows about any matter localized in the interior: the matter backreacts on the metric, and therefore contributes at infinity. This is just like the Gauss law in E&M.

8.4 Other conserved charges

Other asymptotic Killing vectors will similarly lead to conserved quantities. For example, if $\zeta = \partial_\phi$ satisfies (8.23), then we can define the conserved charge

$$J = \int_{\partial\Sigma} d^2x\sqrt{\sigma}u^i T_{i\phi} \ . \tag{8.25}$$

This is in fact the angular momentum, and agrees with all the usual formulae for computing the angular momentum of a spacetime.

We could also define boost charges, and get the full Poincare group. This requires some modifications, since in this discussion ζ_i was a vector within the fixed ∂M, whereas boosts act on ∂M. The results are similar.

8.5 Asymptotic Symmetry Group

We have seen that the bulk Hamiltonian vanishes, but there are boundary terms that compute conserved charges. Now I will try to explain physically what is going on here.

Local diffs are fake. Global diffs are real.
GR is locally diff invariant, but it is not invariant under diffs that reach the boundary. To see this from the action, just vary it under a general diff ζ. The Lie derivative for a density is

$$\delta_\zeta(\sqrt{g}f) \equiv \mathcal{L}_\zeta(\sqrt{g}f) = \nabla_\mu(f\zeta^\mu) \ . \tag{8.26}$$

Applying this the Lagrangian density of GR we see that it is only diff-invariant up to a boundary term,

$$\int_M \delta_\zeta(\sqrt{g}\mathcal{L}) = \int_{\partial M} dA^\mu \zeta_\mu \mathcal{L} \ . \tag{8.27}$$

This is important, so I'll rephrase: General relativity is invariant under local diffeomorphisms. These are like gauge symmetries: fake symmetries, redundancies, that do not change the physics and are just a convenient human invention to describe massless particles. However it is *not* invariant under diffeomorphisms that reach the boundary. The coordinates as $r \to \infty$ are actually important and meaningful, like the coordinates in a non-gravitational theory. A time reparameterization with compact support, *i.e.,* $t \to t'(t, x)$ such that $t' \to t$ as $r \to \infty$, is a local diff and involves no physics. A global time shift $t \to t + 1$ acts at infinity and is true time evolution.

The bulk terms in the Hamiltonian, *i.e.,* the constraints, correspond to local diffs, and the boundary terms correspond to diffs that reach the boundary. That is why the bulk term vanishes on shell and the boundary term does not.

Certain diffs that reach infinity are actual symmetries. By 'actual' symmetries, I mean symmetries that act on the space of states in the theory: they take one state to a *distinct but related* state with similar properties, as opposed to gauge symmetries which physically do nothing.

Asymptotic symmetries in $U(1)$ gauge theory

The precise version of all these statements is the formalism of *asymptotic symmetries*. The definition of the asymptotic symmetry group is the group of symmetry transformations modded out by trivial symmetries,

$$ASG = \frac{\text{symmetries}}{\text{trivial symmetries}} \ . \tag{8.28}$$

The definition of a 'trivial symmetry' is one whose associated conserved charge vanishes.

Let's consider electromagnetism as an example. The action $I = -\frac{1}{4}\int d^4x(F_{\mu\nu}F^{\mu\nu} + A_\mu J^\mu_{matter})$ is invariant under an infinite number of transformations,

$$\delta A_\mu = \partial_\mu \Lambda(x) \ , \qquad \delta\phi = i\Lambda(x)\phi \ , \tag{8.29}$$

where the second term indicates the usual phase rotation on charged matter. These are gauge symmetries. A local gauge symmetry, ie a transformation for which $\Lambda(x)$ has compact support, does not have any conserved charge associated to it. In fact despite this infinite number of symmetries we know electromagnetism has only one conserved quantity, the total charge

$$Q \sim \int_\Sigma d^3x J^0_{matter} \sim \int_{\partial\Sigma} d^2x\, F_{tr} \ . \tag{8.30}$$

This is the conserved charge associated to the global $U(1)$ rotation – it exists and is conserved even in the un-gauged theory. Thus the global rotation is physical, while local phase rotations are just redundancies.

The definition (8.28) of the asypmtotic symmetry group is the group of all transformations, mod gauge transformations with zero associated charge. Therefore in electromagnetism,

$$ASG = U(1)_{global} \ . \tag{8.31}$$

Asymptotic symmetries in gravity

We will not go into depth on the ASG in gravity right now, but I will just mention some facts. The ASG in gravity is generated by the conserved charges, which we argued above are the charges associated to some special vector fields, including those for which $\nabla_{(i}\zeta_{j)} \to 0$ at infinity. In asymptotically flat spacetimes, ie spacetimes approaching Minkowski space fast enough as $r \to \infty$, these are simply the Killing vectors of Minkowski space. Thus the asymptotic symmetry group of asymptotically flat spacetimes is the Poincare group.*

This notion is important, because general spacetimes have no isometries, and therefore no local conserved charges. (For example, there is no 'energy' conserved along the geodesic of a probe particle.) Asymptotic symmetries allow us to define global conserved quantities in this situation.

The Poincare algebra in 4D has 10 generators: 4 translations P_μ and 6 Lorentz generators $M_{\mu\nu}$. The generators obey the Poincare Lie algebra

$$[P_\mu, P_\nu] = 0 \tag{8.32}$$

$$\frac{1}{i}[M_{\mu\nu}, P_\rho] = \eta_{\mu\rho}P_\nu - \eta_{\nu\rho}P_\mu \tag{8.33}$$

$$\frac{1}{i}[M_{\mu\nu}, M_{\rho\sigma}] = \eta_{\mu\rho}M_{\nu\sigma} - \eta_{\mu\sigma}M_{\nu\rho} - \eta_{\nu\rho}M_{\mu\sigma} + \eta_{\nu\sigma}M_{\mu\rho}\ . \tag{8.34}$$

If we just label the generators as V^A for $A = 1 \dots 10$, then this is just a Lie algebra

$$i[V^A, V^B] = f^{AB}_{\ \ C}V^C \tag{8.35}$$

with some structure constants $f^{AB}_{\ \ C}$. Each of these generators is associated to a Killing vector of Minkowski space:

$$V^A \leftrightarrow \zeta^{(A)\mu}, \qquad A = 1 \dots 10\ . \tag{8.36}$$

For example $P^\mu \leftrightarrow \partial_\mu$, $M_{tx} \leftrightarrow t\partial_x + x\partial_t$, etc. The Killing vectors obey the same algebra, under the Lie bracket:

$$[\zeta^A\ ,\ \zeta^B]^\mu_{LB} \equiv \zeta^{A\nu}\partial_\nu\zeta^{B\mu} - \zeta^{B\nu}\partial_\nu\zeta^{A\mu} = f^{AB}_{\ \ C}\zeta^{C\mu}\ . \tag{8.37}$$

(Here A is a label of which vector, and μ is a spacetime index.)

*This is true at spacelike infinity. The story at null infinity is much more subtle since, in non-static spacetimes, gravitational radiation reaches null infinity and distorts the asymptotics. This leads to what is called the BMS group, which is an active area of research.

Recall that conserved charges generate the action of the diffeomorphism under Dirac brackets. That is, the charge

$$Q^A = H[\zeta^A] \tag{8.38}$$

generates

$$\{Q^A, X\}_{DB} = \mathcal{L}_{\zeta^A} X \ . \tag{8.39}$$

For this to be consistent with the algebra, the charges themselves must obey the same algebra:

$$\{Q^A \ , \ Q^B\}_{DB} = f^{AB}{}_C Q^C \ . \tag{8.40}$$

In other words,

$$\{H[\zeta], H[\chi]\}_{DB} = H\big[[\zeta, \chi]_{LB}\big] + \text{constant} \ , \tag{8.41}$$

where we have allowed a constant 'central charge' term in the algebra of charges, since this would still be consistent with the action of the generators on X (and actually does appear in important examples).

Sometimes the ASG leads to surprises. A famous example is in anti-de Sitter space. The isometry group of AdS_D is $SO(D-1,2)$. So a natural guess is that the asymptotic symmetry group of asymptotically-AdS spacetimes is also $SO(D-1,2)$. This is true for $D > 3$ but wrong in $D = 3$, as shown by Brown and Henneaux. We will talk about this more later.

8.6 Example: conserved charges of a rotating body

The linearized solution of GR that carries both energy and angular momentum is

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 + \frac{2M}{r}\right)(dr^2 + r^2 d\Omega^2) - \frac{4j\sin^2\theta}{r} dt d\phi \ . \tag{8.42}$$

This is, for example, the metric far away from a Kerr black hole, or a rotating planet.

We will compute the energy and angular momentum using the on-shell Hamiltonian (8.18). Here it is again, after enforcing the constraints:

$$H[\zeta] = \int_{\partial\Sigma} d^2x \, \sqrt{\sigma} u^i T_{ij} \zeta^j \ . \tag{8.43}$$

The energy is associated to $\zeta = \partial_t$ and the angular momentum to $\zeta = \partial_\phi$.

Kinematics

We want to compute T^{ij}. This is a tensor living on ∂M, which is the surface $r = r_0$. To define tensors on ∂M, we first compute the unit normal to the ∂M,

$$n_\mu dx^\mu = \sqrt{1 + \frac{2M}{r}} dr \ . \tag{8.44}$$

The full metric can be split into the normal and tangential parts as

$$g_{\mu\nu} = \gamma_{\mu\nu} + n_\mu n_\nu \ . \tag{8.45}$$

γ^μ_ν projects onto the boundary, since $n_\mu \gamma^\mu_\nu = 0$. The components γ^μ_i for $\mu = t, r, \theta, \phi$ and $i = t, \theta, \phi$ can be used to turn spacetime tensors into boundary tensors, and vice-versa:

$$V_i \equiv \gamma^\mu_i V_\mu \ . \tag{8.46}$$

The induced metric on ∂M is

$$\gamma_{ij} dx^i dx^j = -\left(1 - \frac{2M}{r_0}\right) dt^2 + r_0^2 \left(1 + \frac{2M}{r_0}\right)(d\theta^2 + \sin^2\theta d\phi^2) - \frac{4j}{r_0}\sin^2\theta d\phi dt \ . \tag{8.47}$$

The timelike unit normal to a constant-t hypersurface is

$$u_\mu dx^\mu = (-1 + \frac{M}{r} + O(1/r^2))dt \ . \tag{8.48}$$

(This could be used to define the induced metric from $h_{\mu\nu} = g_{\mu\nu} + u_\mu u_\nu$ and corresponding projector but we won't need these to compute the charges.) Projecting the timelike normal onto the boundary doesn't change anything, we still have

$$u_i dx^i = (-1 + \frac{M}{r_0} + O(1/r_0^2))dt \ , \tag{8.49}$$

where remember i runs over the boundary directions $x^i = (t, \theta, \phi)$.

Finally, we need the volume element of the boundary at fixed time. The induced metric on $\partial\Sigma$ is

$$\sigma_{AB}dx^A dx^B = r_0^2(1+\frac{2M}{r_0})(d\theta^2+\sin^2\theta d\phi^2) \ , \tag{8.50}$$

with volume element

$$\sqrt{\sigma} = r_0^2\left(1+\frac{2M}{r_0}\right)\sin\theta \ . \tag{8.51}$$

Stress tensor

The extrinsic curvature of ∂M is

$$K_{\mu\nu} = -\nabla_{(\mu}n_{\nu)} \ . \tag{8.52}$$

As a boundary tensor,

$$K_{ij} = h_i^\mu h_j^\nu K_{\mu\nu} \ . \tag{8.53}$$

The trace of K is the same whether we use $K_{\mu\nu}$ or K_{ij} (check this!). It is

$$K = -\frac{2}{r_0} - \frac{3M}{r_0^2} + O(r_0^{-3}) \ . \tag{8.54}$$

Now we compute the stress tensor from its definition (ignoring the background subtraction for now), $T_{ij} = K_{ij} - \gamma_{ij}K$. (I've rescaled it by 8π to unclutter notation, but will put the 8π back in the Hamiltonian below.) It has components

$$T_{tt} = -\frac{2}{r_0} + \frac{8M}{r_0^2}, \quad T_{t\phi} = -\frac{5j\sin^2\theta}{r_0^2}, \quad T_{\theta\theta} = r_0 + M, \quad T_{\phi\phi} = \sin^2\theta(r_0+M) \tag{8.55}$$

plus higher order terms $O(M^2/r_0^2)$. (In this equation we are still ignoring the background subtraction, we will deal with that below.)

Energy

The energy is the on-shell Hamiltonian for $\zeta = \partial_t$. Putting it all together, we have so far for the energy

$$E_{unsub} = \frac{1}{8\pi}\int_{\partial M} 2r_0\sin\theta = -r_0 \ , \tag{8.56}$$

where 'unsub' means we have not dealt with the background subtraction yet.

To do the background subtraction, we repeat the whole calculation on the flat spacetime

$$ds^2_{sub} = -\left(1 - \frac{2M}{r_0}\right) dt^2 + \left(1 + \frac{2M}{r_0}\right)(dr^2 + r^2 d\Omega^2) \ . \tag{8.57}$$

This is a flat spacetime with the same intrinsic geometry on ∂M.*

Going through all the steps again, the subtraction term is $E_{sub} = -r_0 - M$. Therefore the final answer is

$$E = M \ , \tag{8.58}$$

as expected.

Angular momentum

The angular momentum is the on-shell Hamiltonian for $\zeta = -\partial_\phi$.† There is no background subtraction necessary. We find

$$J = \frac{1}{8\pi} 3j \int_{\partial\Sigma} d\theta d\phi \sin^3\theta = j \ . \tag{8.59}$$

*We could include the angular momentum term, but we can shift the time coordinate to make it $O(1/r_0^2)$ and it does not contribute. Put differently, we really only need to embed $\partial\Sigma$ into flat spacetime, not all of ∂M, so this is only important for the energy and we can ignore the angular momentum.

†The minus sign here is the standard convention. It is related to the fact that a mode $e^{-i\omega t + im\phi}$ carries energy $E = +\omega$ and angular momentum $J = +m$.

9 Symmetries of AdS_3

This section consists entirely of exercises. If you are not doing the exercises, then read through them anyway, since this material will be used later in the course. The main goal of this section is derive the famous result of Brown and Henneaux on the central charge of AdS_3. This was done in the 80s, using slightly different techniques from what we'll use here, and later came to play an important role in AdS/CFT, as we'll see later.

9.1 Exercise: Metric of AdS_3

Anti-de Sitter space is a constant-negative-curvature spacetime. It is the maximally symmetric solution of Einstein's equation with a negative cosmological constant. AdS_D can be realized as a hyperboloid embedded in a $D+1$-dimensional geometry. In this section we will talk about AdS_3, which is the hyperboloid

$$X_A X^A = -\ell^2 \tag{9.1}$$

where $A = 0, 1, 2, 3$ is an index in the space Minkowski$_2\times$ Minkowski$_2$, with metric

$$H_{AB} dX^A dX^B = -dX_0^2 + dX_1^2 + dX_2^2 - dX_3^2 \ . \tag{9.2}$$

To find intrinsic coordinates on AdS_3, we just need to solve (9.1). One way to solve this equation is by

$$X_0 = \ell \cosh\rho \cos t, \quad X_1 = \ell \sinh\rho \sin\phi, \quad X_2 = \ell \sinh\rho \cos\phi, \quad X_3 = \ell \cosh\rho \sin t \ . \tag{9.3}$$

(a) Check that this solves (9.1), and use (9.2) to find the induced metric on the hyperboloid.

Answer:

$$ds^2 = \ell^2(-\cosh^2\rho \, dt^2 + d\rho^2 + \sinh^2\rho \, d\phi^2) \ . \tag{9.4}$$

These are global coordinates on AdS_3. Although on the hyperboloid (9.1) we can see from (9.3) that t is a periodic coordinate, when we say 'AdS_3' we will always mean the space in which t is 'unwrapped', $t \in (-\infty, \infty)$ (the *universal covering space* of the hyperboloid).

(b) Find the cosmological constant in terms of the AdS radius ℓ.

9.2 Exercise: Isometries

AdS_3 inherits the isometries of the embedding space that preserve the hyperboloid. (For the same reason that the isometries of S^2 are inherited from rotations in R^3.) The group of rotations+boosts in a 4d geometry with signature $(+,+,-,-)$ is $SO(2,2)$, so we expect this to be the isometry group of AdS_3. In this problem we'll confirm this.

(a) As an example, consider the boost vector

$$V = X^1 \partial_{X^0} + X^0 \partial_{X^1} \tag{9.5}$$

in the embedding space (9.2). This preserves the hyperboloid, since under $X^A \to X^A - V^A$, the lhs side of (9.1) is unchanged to linear order (check this).

Write V as an isometry of AdS_3, in the coordinates (9.4). To do this, first define the projection tensor

$$P_\mu^A = \frac{\partial X^A}{\partial x^\mu} \tag{9.6}$$

where x^μ are the coordinates of AdS_3. This can be used to convert the 4-vector V^A into a tensor living on the hyperboloid,

$$\chi_\mu = P_\mu^A V_A \ . \tag{9.7}$$

Find χ^μ, and check that it is a Killing vector of the metric (9.4).

This same procedure can be used to find all of the Killing vectors of AdS, but I will spare you the trouble. The answer, in a convenient basis, is

$$\begin{aligned}
\zeta_{-1} &= \tfrac{1}{2}\left[\tanh(\rho)e^{-i(t+\phi)}\partial_t + \coth(\rho)e^{-i(t+\phi)}\partial_\phi + ie^{-i(t+\phi)}\partial_\rho\right] \\
\zeta_0 &= \tfrac{1}{2}(\partial_t + \partial_\phi) \\
\zeta_1 &= \tfrac{1}{2}\left[\tanh(\rho)e^{i(t+\phi)}\partial_t + \coth(\rho)e^{i(t+\phi)}\partial_\phi - ie^{i(t+\phi)}\partial_\rho\right] \\
\bar\zeta_{-1} &= \tfrac{1}{2}\left[\tanh(\rho)e^{-i(t-\phi)}\partial_t - \coth(\rho)e^{-i(t-\phi)}\partial_\phi + ie^{-i(t-\phi)}\partial_\rho\right] \\
\bar\zeta_0 &= \tfrac{1}{2}(\partial_t - \partial_\phi) \\
\bar\zeta_1 &= \tfrac{1}{2}\left[\tanh(\rho)e^{i(t-\phi)}\partial_t - \coth(\rho)e^{i(t-\phi)}\partial_\phi - ie^{i(t-\phi)}\partial_\rho\right]
\end{aligned} \tag{9.8}$$

Note that the subscripts here are just labels, not spacetime indices.

(b) Check that the vectors ζ_{-1}, ζ_0, ζ_1 are Killing vectors.

(c) Now check that they obey the $SL(2,R)$ algebra:

$$[L_1, L_{-1}] = 2L_0, \quad [L_1, L_0] = L_1, \quad [L_{-1}, L_0] = -L_{-1}\ . \tag{9.9}$$

That is, the Killing vectors obey this algebra under Lie brackets, with an additional i, for example

$$i\{\zeta_1, \zeta_{-1}\}_{LB} = 2\zeta_0\ , \qquad \textit{etc.} \tag{9.10}$$

The barred zetas in (9.8) commute with the unbarred zetas, and form another $SL(2,R)$ algebra. Therefore the isometries of AdS_3 form the algebra

$$SL(2,R)_L \times SL(2,R)_R. \tag{9.11}$$

The subscripts mean 'left' and 'right', since the ζ's involve only the 'left-moving' combination $t+\phi$ and the $\bar\zeta$'s involve the 'right-moving' combination $t-\phi$.

Note that as a Lie algebra, of $SO(2,2) = SL(2,R)\times SL(2,R)$. This is a special feature of AdS_3. In general the AdS_D isometry group is $SO(D-1,2)$, which does not split into two factors.

9.3 Exercise: Conserved charges

(a) Do the coordinate change

$$t^{\pm} = t \pm \phi \ , \qquad \rho = \log(2r) \ , \tag{9.12}$$

and expand the metric (9.4) at large r. Show that to leading order

$$ds^2 = \ell^2 \left(\frac{dr^2}{r^2} - r^2 dt^+ dt^- \right) \ . \tag{9.13}$$

These are called *Poincaré coordinates*, and in fact this metric is an exact solution of Einstein's equation – it covers a subregion of AdS_3 called the Poincaré patch.

A spacetime is called *asymptotically AdS* if it approaches (9.13) as $r \to \infty$.*

(b) Consider the asymptotically AdS spacetime

$$ds^2 = \ell^2 \left(\frac{dr^2}{r^2} - r^2 dt^+ dt^- \right) + h_{++}(dt^+)^2 + h_{--}(dt^-)^2 + 2h_{+-}dt^+ dt^- \ , \tag{9.14}$$

where the h's are arbitrary functions of t^+ and t^- but independent of r. We will compute the boundary stress tensor (Brown-York tensor) and use it to define the energy and other conserved charges in AdS_3.

The boundary stress tensor is defined as the variation of the on-shell action

$$T^{ij} \equiv \frac{2}{\sqrt{-\gamma}} \frac{\delta S_{on-shell}}{\delta \gamma_{ij}} \tag{9.15}$$

where the action is

$$S[g] = \frac{1}{16\pi} \int_M \sqrt{-g}(R - 2\Lambda) + \frac{1}{8\pi} \int_{\partial M} \sqrt{-\gamma} K + \frac{a}{8\pi} \int_{\partial M} \sqrt{-\gamma} \ . \tag{9.16}$$

For the bulk term and Gibbons-Hawking term, we can use our formulae from flat space given in previous lectures. The last term is a counterterm, which takes the place of the

*Specifically, the subleading components of the metric must have a certain fall-off at large r. These conditions are basically chosen so that the Hamiltonian can be defined.

'background subtraction' we did in flat space. This leads to

$$T^{ij} = \frac{1}{8\pi}\left[K^{ij} - K\gamma^{ij} + \tilde{a}\gamma^{ij}\right] \ . \tag{9.17}$$

Choose the counterterm coefficient $\tilde{a}$ so that T_{ij} is finite as the cutoff surface $r_0 \to \infty$. Compute T_{++}, T_{--}, and T_{+-} to first order in the perturbation h_{ij} in the limit $r_0 \to \infty$.

Reference: Balasubramanian and Kraus, hep-th/9902121.

(c) Compute the energy of the spacetime (9.14). It is defined as

$$E = \frac{1}{\ell}\int_0^{2\pi} d\phi \sqrt{\sigma} u^i T_{ij} \zeta^j \tag{9.18}$$

where u^i is the timelike normal to a fixed-t slice, and $\zeta = \partial_t$. (The overall $1/\ell$ is a convention, necessary due to the fact we are using dimensionless coordinates.)

(d) An example of an asymptotically AdS spacetime is the BTZ black hole,

$$ds^2 = \ell^2\left[-(r^2 - 8M)dt^2 + \frac{dr^2}{r^2 - 8M} + r^2 d\phi^2\right] \ . \tag{9.19}$$

Check that for this spacetime

$$E = M \ . \tag{9.20}$$

To use your results of the previous problem in this calculation you must first change coordinates to put it in the form (9.14). In particular you will need to redefine $r' = r'(r)$ to eliminate the perturbation to g_{rr}.

(e) Compute the energy of global AdS, by keeping the subleading terms in the coordinate transformation (9.12) and plugging them into your formula for the stress tensor. (*Hint:* the answer is negative. That's OK, this is just a choice of zero for energy.)

Comment: We've focused on the energy, but we could compute conserved charges corresponding to all the other Killing vectors in exactly the same way.

10 Interlude: Preview of the AdS/CFT correspondence

The rest of this course is, roughly speaking, on the AdS/CFT correspondence, also known as 'holography' or 'gauge/gravity duality' or various permutations of these words. AdS/CFT was conjectured by Maldacena in a famous paper in 1997. A full understanding of Maldacena's motivations and results, and the huge body of work to follow, requires some string theory, but AdS/CFT itself is independent of string theory and we will not follow this route. Instead we will 'discover' AdS/CFT by throwing stuff at black holes. In fact, this parallels the historical discovery of AdS/CFT in 1996-1997, though we will obviously take a shorter path. Our starting point will be a black-hole-like solution in 6 dimensions, which might seem umotivated, so the purpose of this interlude is to describe where we are headed, so you know we are doing this for a good reason.

10.1 AdS geometry

Anti-de Sitter space is the maximally symmetric solution of the Einstein equations with negative cosmological constant. We worked out the metric of AdS_3 in global and Poincaré coordinates in the previous section. For general dimension AdS_{d+1}, the metric in global coordinates is

$$ds^2 = \ell^2 \left(-\cosh^2 \rho dt^2 + d\rho^2 + \sinh^2 \rho d\Omega_{d-1}^2\right) \ . \tag{10.1}$$

To find the Penrose diagram, we can extract a factor of $\cosh^2 \rho$ and then define a new coordinate by

$$d\sigma = \frac{d\rho}{\cosh \rho} \Rightarrow \sigma = 2 \tan^{-1} \tanh(\rho/2) \ . \tag{10.2}$$

As ρ runs from 0 to ∞, ρ runs from 0 to $\pi/2$. Each value of t, σ is a sphere S^{d-1}. Therefore the Penrose diagram looks like a solid cylinder, where ρ is the radial coordinate of the cylinder, and t, Ω are the coordinates on the surface of the cylinder.

Unlike flat space, the conformal boundary (usually just called 'the boundary') of AdS is *timelike*. From the Penrose diagram, we can see that massless particles reach the boundary in finite time t. (Massive particles cannot reach the boundary; they feel an $e^{2\rho}$ potential if they try to head to large ρ.)

The metric of the Poincaré patch is

$$ds^2 = \frac{\ell^2}{z^2}\left(dz^2 - dt^2 + d\vec{x}^2\right) \ , \tag{10.3}$$

where $\vec{x} = (x^1, \dots, x^{d-1})$. In these coordinates, the boundary is at $z = 0$. These coordinates cover a wedge of the global cylinder. You can check this for AdS_3 using the coordinate transformations derived in the previous section.

10.2 Conformal field theory

A conformal field theory (CFT) is a QFT with a particular spacetime symmetry, *conformal invariance*. Conformal invariance is a symmetry under local scale transformations. We will discuss this in detail later. For now I will just mention that one consequence of conformal symmetry is that correlation functions behave nicely under coordinate rescalings $x \to \lambda x$. Correlation functions of primary operators (which are lowest weight states of a conformal representation) obey

$$\langle O_1(x_1)O_2(x_2)\cdots O_n(x_n)\rangle = \lambda^{\Delta_1+\Delta_2+\cdots+\Delta_n}\langle O_1(\lambda x_1)\cdots O_n(\lambda x_n)\rangle \tag{10.4}$$

where Δ_i is called the *scaling dimension* of the operator O_i. This (together with rotation and translation invariance) implies for 2pt functions

$$\langle O(x)O(y)\rangle \propto \frac{1}{|x-y|^{2\Delta}} \ . \tag{10.5}$$

The simplest example of a CFT is a free massless scalar field, where for example in 4d $\langle\phi(x)\phi(y)\rangle = (x-y)^{-2}$. A massive free field is not conformal, since m shows up in correlation functions and spoils the simple power behavior. This is generally true – CFTs do not have any dimensionful parameters, so there can be no mass terms in the Lagrangian. However the converse is not true, since there are theories with no mass terms in the classical theory are not necessarily conformal. For example in massless QCD, scale symmetry is broken in the the quantum theory so the theory acquires a dimensionful parameter via dimensional transmutation.

There are also very nontrivial interacting conformal field theories. We will discuss a couple of examples later.

10.3 Statement of the AdS/CFT correspondence

The AdS/CFT correspondence is the an exact relationship between any* theory of quantum gravity in asymptotically AdS_{d+1} spacetime and an ordinary CFT_d, without gravity. This relationship is called a *duality*. It is *holographic* since the gravitational theory lives in (at least) one extra dimension. The theories are believed to be entirely equivalent: any physical (gauge-invariant) quantity that can be computed in one theory can also be computed in the dual. However, the mapping between the two theories can be highly nontrivial. For example, easy calculations on one side often map to strongly coupled, incalculable quantities on the other side.

It is often useful to think of the CFT as 'living at the conformal boundary' of AdS. Indeed, the CFT lives in a spacetime parameterized by $x = (t, \vec{x})$, whereas gravity fields are functions of x and the radial coordinate ρ. And when we discuss correlation functions of local operators we will see that a CFT point x corresponds to a point on the conformal boundary of AdS. But it is not quite accurate to say that the CFT lives on the boundary, for two reasons. First, we should not think about having both theories at once; we either do CFT or we have an AdS spacetime, never both at the same time. Second, the CFT is dual to the entire gravity theory, so in a sense it lives everywhere.

The two theories are commonly referred to as 'the bulk' (*i.e.,* the gravity theory) and 'the boundary' (ie the CFT).

In this course we will mostly restrict our attention to two types of observables in AdS/CFT: thermodynamic quantities and correlation functions.

*Some people might obect to the word 'any' here. To be safe, we could say 'any theory that we know how to define in the UV and acts like ordinary gravity+QFT in the IR.'

Thermodynamics

The mapping between thermodynamic quantities on the two sides of the duality is simply that they should be equal, for example the thermal partition functions obey

$$Z_{cft}(\beta) = Z_{gravity}(\beta) \ . \tag{10.6}$$

Here $Z_{cft} = \text{ Tr } e^{-\beta H_{cft}}$ is the ordinary thermodynamic partition function of a QFT. $Z_{gravity}$ is the quantity whose semiclassical limit we discussed above, related to the on-shell action of a black hole,

$$Z_{gravity}(\beta) = e^{-S_E[g]} + \cdots \ . \tag{10.7}$$

For the exact relation (10.6) we must in principle include all the quantum corrections to this semiclassical formula.

Correlation functions

The goal of the next couple lectures is to derive the *dictionary* that relates CFT correlators to a gravity calculation. We will give the exact prescription later, but here is the general idea. Each field $\phi_i(\rho, x)$ in the gravitational theory there is a corresponding operator $O_i(x)$ in the CFT.* The mass of ϕ determines the dimension of O. CFT correlation functions can be computed on the gravity side by computing a gravity correlator of ϕ, with the points inserted at the boundary:

$$\langle O_1(x_1) \cdots O_n(x_n) \rangle_{cft} \leftrightarrow \text{``} \lim_{\rho \to \infty} \text{''} \langle \phi_1(\rho, x_1) \cdots \phi_n(\rho, x_n) \rangle_{gravity} \tag{10.8}$$

The limit is in quotes because actually we need to rescale by some divergent factors that we'll come to later.

Top down, bottom up, and somewhere in between

AdS/CFT is general, we do not need to refer to a particular theory of a gravity or a particular CFT. However it is often useful to have specific theories in mind, with

*Here $x = (t, \vec{x})$ denotes all d dimensions of the CFT and ρ is the radial coordinate.

detailed microscopic definitions. For example: *Type IIB string theory on* $AdS_5 \times S^5$ *is dual to* $\mathcal{N} = 4$ *supersymmetric Yang-Mills in 4d.*

Super-Yang-Mills is a particular CFT with a known Lagrangian. Although IIB string theory is not defined non-perturbatively (except via this duality), it has many known microscopic ingredients. Calculations in these two specific theories can be compared in great detail.

There are other microscopic examples — deformations of this one, and different versions in different dimensions, with different types of dual CFTs. All of them (as far as I know) come from brane constructions in string theory. This is often called the 'top down' approach to AdS/CFT.

Another approach is to simply assume that we have a CFT with some low-dimension primaries with a particular pattern, and perhaps with some assumptions about the symmetries and conserved charges of the theory. This is more in the spirit of effective field theory and is often called 'bottom up.' In many cases we can also include information about the UV completion of the CFT (*i.e.,* very high dimension operators) in this approach so it actually goes beyond effective field theory, but without every specifying the actual Lagrangian of the CFT.

Both approaches are important. Often calculations that can be done in one approach are impossible in the other, or calculations first done microscopically turn out to have more general and possibly more intuitive explanations via effective field theory.

11 AdS from Near Horizon Limits

Anti-de Sitter space appears in the near horizon region of extremal black holes. In this section we will describe how the near-horizon limit of extremal Reissner-Nordstrom is $\mathrm{AdS}_2 \times S^2$. Since the case $d = 1$ (*i.e.*, AdS_2) is a special case of AdS/CFT that we would like to mostly avoid, we then discuss the 6d black string. This solution has a near-horizon AdS_3 which will serve as our main example for AdS/CFT.

11.1 Near horizon limit of Reissner-Nordstrom

Here is the 4d Reissner-Nordstrom black hole again, from (2.10):

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega_2^2 \ , \quad f(r) = 1 - \frac{2M}{r} + \frac{Q^2}{r^2} \ . \tag{11.1}$$

Recall that this solution is restricted to $M > Q$ (we assume $Q > 0$) by cosmic censorship. In general, the near horizon geometry is approximately Rindler $\times S^2$, as discussed above. But something special happens in the *extremal limit*,

$$M = Q \ . \tag{11.2}$$

In this limit the horizon is a double zero, $f(r) = (1 - Q/r)^2$. So the inner and outer horizons coincide, $r_+ = r_- = Q$, and the Hawking temperature (2.19) is zero.

To take the near horizon limit of the extremal black hole, we define

$$r = Q(1 + \lambda/z), \quad t = QT/\lambda \tag{11.3}$$

where λ is an arbitrary parameter. Plugging this into the extremal metric and taking the limit $\lambda \to 0$ with z, T, θ, ϕ held fixed gives the spacetime

$$ds^2 = \frac{Q^2}{z^2}\left(-dT^2 + dz^2\right) + Q^2 d\Omega_2^2 \ . \tag{11.4}$$

The metric (11.4) is $AdS_2 \times S^2$. This is the near horizon region of the original black hole, since if $\lambda \to 0$ with z held fixed, $r \to Q$. Recall that this spacetime is supported by some nontrivial electric field. Applying the same procedure to the field strength gives a constant electric field in the near horizon region.*

λ has disappeared entirely from the solution. This means that (11.4) (together with the constant electric field) is actually by itself a solution to the Einstein-Maxwell equations. This did not happen for the non-extremal case: there, the spacetime was only approximately Rindler near the horizon, and Rindler $\times S^2$ does not solve the Einstein-Maxwell equations.

Geometry of the Near Horizon Region

The key difference between the extremal and non-extremal near horizon limits is that *the near horizon region of an extremal black hole is infinitely long.* To see this let us calculate the distance to the horizon along a fixed-t slice in the general metric (11.1), from some arbitrary point $r_0 > r_+$:

$$D = \int_{r_+}^{r_0} dr \sqrt{f(r)} \sim -M \log(r_+ - r_-) \sim M \log \frac{1}{QT_H} \tag{11.5}$$

where T_H is the Hawking temperature (2.19). This diverges as $T_H \to 0$. (The $\sim$ means we are dropping constants and the contribution of the r_0 limit to the integral, which doesn't matter.) The long region for small T_H

Global coordinates

The near-horizon metric we found in (11.4) is AdS_2 in Poincaré coordinates. This covers only a patch of the full AdS spacetime. Similarly, we considered only one patch of the Reissner-Nordstrom spacetime. The full global AdS_2 includes many Poincaré patches, and each patch gives the near-horizon region of a different patch of the global Reissner-Nordstrom. This is illustrated in the Penrose diagrams in figure 4.

What about the sphere?

We are only drawing the Penrose diagrams for AdS_2, but the geometry is in fact $\mathrm{AdS}_2 \times S^2$. Actually the sphere does not affect the conformal boundary. Since the

*In the original coordinates, $F_{rt} = Q/r^2$. In the new coordinates, after scaling $\lambda \to 0$, $F_{zT} = -Q/z^2$. This does not look constant in these coordinates, but if we define $\sigma = 1/z$ then $F_{\sigma T} = Q$ is a constant electric field.

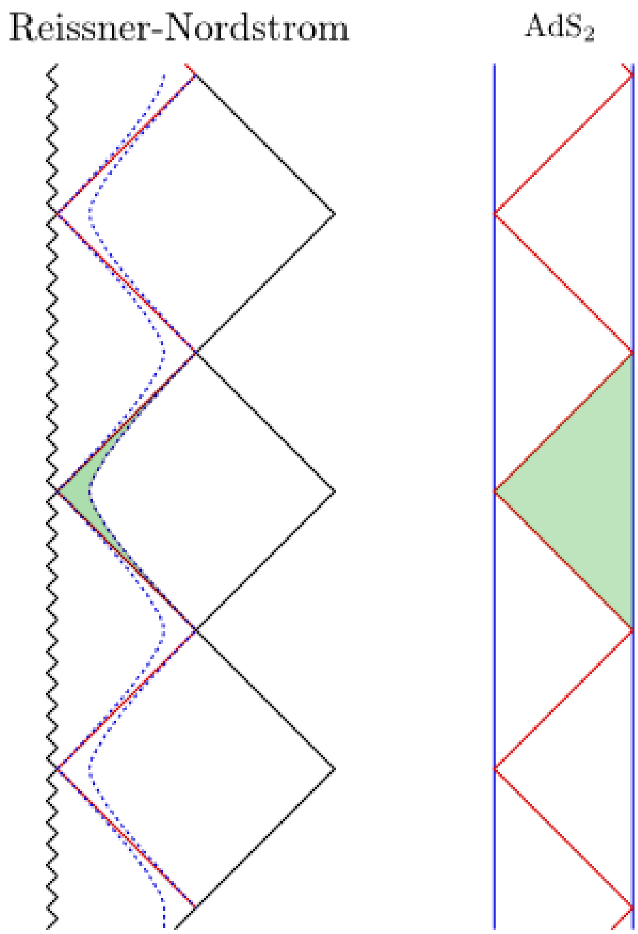


Figure 4: Penrose diagrams for extremal Reissner-Nordstrom and AdS_2. The AdS_2 Penrose diagram is a 'zoomed in' version of the RN diagram which includes only the near horizon region. Unlike higher-dimensional AdS, AdS_2 has two conformal boundaries, which are the blue lines on the left and right. In the RN diagram, the coordinates are degenerate near the horizon so these boundaries are drawn as dashed blue lines slightly away from $r = r_+$. The black hole horizon in RN (red) is the same as the Poincaré horizon in AdS_2. The Poincaré patch of AdS_2 is shaded green in both diagrams.

prefactor in front of AdS_2 blow up near the boundary, after a conformal rescaling the sphere just drops out. So points on the conformal boundary are labeled only by T, not by T, θ, ϕ.

Near horizon as a low-energy limit

A particle on a geodesic has a conserved energy-at-infinity

$$E \equiv -p_t = f(r)\frac{dt}{d\tau} \ . \tag{11.6}$$

For a particle in the near-horizon region, this is strictly zero as $\lambda \to 0$. So from the point of view of an observer at infinity, everything in the near-horizon region is infinitely redshifted. Similarly, a wave in the near-horizon region $\propto e^{-i\omega_{near}T}$ has zero frequency as measured from infinity, since

$$\omega_{near}T \sim \lambda\omega t \sim 0 \cdot t \ . \tag{11.7}$$

11.2 6d black string

Unfortunately AdS_2 is the runt of the AdS/CFT litter. It is very interesting in its own right but quite different from other dimensions (since CFT_1 does not really make sense) so not suitable for our purposes. We will focus on AdS_3 instead, which appears in the near horizon limit of a 6d black string (among other things). In the exercises you will treat the other most popular example of AdS/CFT which involves AdS_5.

The 6d black string is similar to a black hole, but with horizon topology $S^3 \times S^1$. The 6d metric is*

$$\begin{aligned} ds^2 = (f_1 f_5)^{-1/2}\left(-dt^2 + d\phi^2 + \frac{r_0^2}{r^2}(\cosh\sigma dt + \sinh\sigma d\phi)^2\right) \\ + (f_1 f_5)^{1/2}\left(\frac{dr^2}{1 - r_0^2/r^2} + r^2 d\Omega_3^2\right) \ . \end{aligned} \tag{11.8}$$

*My favorite references on this solution and the discussion to follow are Kiritsis's textbook, section 12.7, and MAGOO hep-th/9905111. Note that I will not distinguish between the Einstein-frame metric and string-frame metric in this discussion; they differ by just a constant in the near-horizon region which can be absorbed into the definition of ℓ.

where

$$f_1 = 1 + \frac{r_1^2}{r^2}, \qquad f_5 = 1 + \frac{r_5^2}{r^2} \ , \tag{11.9}$$

and $\phi \sim \phi + 2\pi R$ is compact. I will not write the other fields but there are some nontrivial scalars and gauge fields that can be found in the references. This solution carries two charges, called Q_1 and Q_5, related to r_1 and r_5. (In string theory these count the number of D1 branes and D5 branes.) It also carries a momentum proportional to $r_0 \sinh\sigma$ along the ϕ direction, as you can guess from the boost term $\cosh\sigma dt + \sinh\sigma d\phi$. Finally r_0 is the position of the horizon and measures the deviation from extremality. To see this, note the surface gravity is proportional to $\partial_r g_{rr}^{-1}|_{r=r_0}$.

This is called the D1-D5-P black string. (Often in the literature you will find that it is dimensionally reduced to 5d along the ϕ direction, so it becomes a 3-charge black hole.)

AdS_3 in the near horizon of the extremal black string

Far away, *i.e.*, $r \gg 0$, this geometry is just $R^4 \times S^1$. Now we will look at the decoupling near horizon limit. This the same type of near horizon limit that we did for extremal RN, so it produces an exact, infinite-volume solution of the equations of motion.

First we consider the case with zero momentum. The extremal D1-D5 with $P = 0$ is obtained by setting $r_0 = 0$,

$$ds^2 = (f_1 f_5)^{-1/2}(-dt^2 + d\phi^2) + (f_1 f_5)^{1/2}(dr^2 + r^2 d\Omega_3^2) \ . \tag{11.10}$$

The horizon is at $r = 0$. To take the near-horizon limit, we define

$$\ell^2 = r_1 r_5 \tag{11.11}$$

scale

$$r \to \lambda \ell r \ , \quad t \to t\ell/\lambda \ , \quad \phi \to \phi\ell/\lambda \ , \tag{11.12}$$

and send $\lambda \to 0$. This has the effect of just dropping the 1 in $f_i = 1 + r_i^2/r^2$, so

$$ds_{near}^2 = \ell^2 \left(\frac{dr^2}{r^2} + r^2(-dt^2 + d\phi^2) \right) + \ell^2 d\Omega_3^2 \ . \tag{11.13}$$

This is the geometry $AdS_3 \times S^3$, where the curvature radii of AdS and of the sphere are equal.

Near-extremal D1-D5-P

Now let us take a different near-horizon limit of (11.8) where we simultaneously scale $r \to 0$ as we scale the black string towards extremality, $r_0 \to 0$. In this limit, $r_0 \cosh \sigma$ stays finite, so this is an extremal limit with finite momentum.

Starting from (11.8) we scale

$$r \to \lambda \ell r \ , t \to t\ell/\lambda \ , \phi \to \phi \ell/\lambda \ , r_0 \to \lambda \ell r_0 \ . \tag{11.14}$$

The resulting metric is

$$ds^2_{near} = \ell^2 \left[-r^2 dt^2 + \frac{dr^2}{r^2 - r_0^2} + r^2 d\phi^2 + r_0^2 \left(\cosh \sigma dt + \sinh \sigma d\phi \right)^2 \right] + \ell^2 d\Omega_3^2 \ . \tag{11.15}$$

The term in brackets is in fact a 3d black hole, called the BTZ black hole. To see this in more standard BTZ coordinates, define the parameters

$$w_+ = r_0 \cosh \sigma, \qquad w_- = r_0 \sinh \sigma \tag{11.16}$$

and do the coordinate change

$$r^2 = w^2 - w_-^2 \ . \tag{11.17}$$

The resulting metric is

$$\frac{ds^2_{near}}{\ell^2} = -h(w) dt^2 + \frac{dw^2}{h(w)} + w^2 \left(d\phi + \frac{w_+ w_-}{w^2} dt \right)^2 + d\Omega_3^2 \ , \tag{11.18}$$

where

$$h(w) = \frac{(w^2 - w_+^2)(w^2 - w_-^2)}{w^2} \ . \tag{11.19}$$

The w, t, ϕ part of this metric is a 3d black hole carrying mass and angular momentum, with horizons at $w_\pm$. Setting $w_- = 0$ and $w_+ = 8M$ gives the $J = 0$ BTZ that was used in some examples earlier in the course.

Exercise: AdS_5 as near horizon limit

Difficulty: easy

Consider the 10D metric

$$ds^2 = f^{-1/2}(-dt^2 + d\vec{x}^2) + f^{1/2}(dr^2 + r^2 d\Omega_5^2) \ . \tag{11.20}$$

where

$$f = 1 + \frac{r_3^4}{r^4} \ . \tag{11.21}$$

r_3 a constant parameter and $\vec{x}$ a coordinate on 4d space R^4. This metric is an extremal black brane. (A *black brane* is like a black hole, but the horizon is a plane instead of a sphere. In string theory, this solution is the geometry corresponding to a stack of Q_3 D3 branes, where Q_3 is a conserved charge of this solution, related to r_3.)

Show that the near-horizon geometry is $\mathrm{AdS}_5 \times S^5$.

Exercise: Near horizon Kerr

Difficulty: A little messier, use Mathematica!

The metric of the 4D Kerr black hole is

$$ds^2 = -\frac{\Delta(r)}{\rho^2}(dt - a\sin^2\theta d\phi)^2 + \frac{\rho^2}{\Delta(r)}dr^2 + \rho^2 d\theta^2 + \frac{1}{\rho^2}\sin^2\theta(adt - (r^2+a^2)d\phi)^2 \ , \tag{11.22}$$

where

$$\Delta(r) = r^2 + a^2 - 2Mr \ , \qquad \rho^2 = r^2 + a^2\cos^2\theta \ , \tag{11.23}$$

and $-M < a < M$. This describes a rotating black hole with mass M and angular momentum

$$J = aM \ . \tag{11.24}$$

Find the value of M (as a function of J) where this black hole is extremal. Then find the near horizon geometry of the extremal Kerr.

Hint: To find a regular near-horizon metric, you must go to a rotating coordinate system that corotates with the black hole horizon, $\psi = \phi - \Omega t$. After going to this rotating coordinate system the calculation is similar to what we did for RN.

Reference: Bardeen and Horowitz, [hep-th/9905099]. (Also [0809.4266].)

12 Absorption Cross Sections of the D1-D5-P

We will now throw a scalar field at a near-extremal D1-D5-P black string. In the process we will rediscover AdS/CFT. The metric is (11.8), where we now assume

$$r_0 \ll r_1, r_5 \ . \tag{12.1}$$

This leads to a low Hawking temperature T_H. (We also assume $\cosh\sigma, r_1/r_5 \sim O(1)$.) Our goal is to calculate the absorption cross-section of a scalar field with low energy,

$$\omega r_5 \ll 1 \ . \tag{12.2}$$

We will assume the scalar χ has zero momentum around the ϕ direction and on the 3-sphere. It is convenient to define

$$T_L = \frac{1}{2\pi}\frac{r_0 e^{\sigma}}{r_1 r_5} \ , \quad T_R = \frac{1}{2\pi}\frac{r_0 e^{-\sigma}}{r_1 r_5} \ , \tag{12.3}$$

which will turn out to be left and right moving temperatures in the dual CFT. These are related to the Hawking temperature by

$$\frac{2}{T_H} = \frac{1}{T_L} + \frac{1}{T_R} \ . \tag{12.4}$$

12.1 Gravity calculation

The wave equation $\Box\chi = 0$ for a scalar field of the form $\chi = e^{-i\omega t}R(r)$ in the metric (11.8) is

$$\left[\frac{h}{r^3}\frac{d}{dr}hr^3\frac{d}{dr} + \omega^2 f\right] R = 0 \tag{12.5}$$

where

$$f = \left(1+\frac{r_1^2}{r^2}\right)\left(1+\frac{r_5^2}{r^2}\right)\left(1+\frac{r_0^2\sinh^2\sigma}{r^2}\right) \ , \qquad h = 1 - \frac{r_0^2}{r^2} \ . \tag{12.6}$$

This is now basically a 1d quantum mechanics problem. To get some intuition for this scattering process, define

$$\chi(r) = \frac{1}{\sqrt{r(r^2-r_0^2)}}\psi(r) \ . \tag{12.7}$$

In these variables, the wave equation is

$$\left[-\frac{d^2}{dr^2} + V(r)\right] \chi = 0 \tag{12.8}$$

where $V(r)$ is easy to find and plot, but annoying to write down. It looks like a well near the horizon $r = r_0$, falls off at infinity, and has a lump somewhere in between. This looks just like an ordinary Schrodinger equation, so we are just scattering through a potential.

To compute the absorption cross-section, we need to solve the wave equation and compare the coefficients of the incoming, transmitted, and reflected waves. The strategy is to solve the equation approximately in the 'near' and 'far' regions, and match these solutions together somewhere in the middle. The near and far regions are defined by

$$\text{far:} \qquad r \gg r_0 \tag{12.9}$$

$$\text{near:} \qquad r \ll r_{1,5}, \qquad r \ll 1/\omega \tag{12.10}$$

These regions overlap in the 'matching region' $r_0 \ll r_m \ll r_{1,5}$.

The general solution of the wave equation in the far region is a linear combination of Bessel functions,

$$R_{far} = r^{-3/2}\sqrt{\frac{\pi\omega r}{2}}\left[AJ_1(\omega r) + BY_1(\omega r)\right] \tag{12.11}$$

The general solution in the near region is

$$R_{near} = \left[\tilde{A}h^{-i(a+b)/2} + \tilde{B}h^{+i(a+b)/2}\right] {}_2F_1(-ia, -ib, 1 - ia - ib, h) \tag{12.12}$$

with

$$a = \frac{\omega}{4\pi T_R}\ , \quad b = \frac{\omega}{4\pi T_L}\ . \tag{12.13}$$

The boundary condition is that the wave is purely ingoing at the horizon $r = r_0$. This sets $\tilde{B} = 0$. Then we expand both R_{near} and R_{far} in the matching region:

$$R_{near} \approx \tilde{A}\frac{\Gamma(1 - ia - ib)}{\Gamma(1 - ia)\Gamma(1 - ib)} + O(r_0^2/r^2) \tag{12.14}$$

$$R_{far} \approx A\frac{\sqrt{\pi}}{2\sqrt{2}}\omega^{3/2} + B\text{-terms} \tag{12.15}$$

We have not written the B terms because they are messy, but we will use conservation of flux to fix B later. Matching the terms in (12.14) gives

$$\sqrt{\frac{\pi\omega^3}{2}}\frac{A}{2} = \tilde{A}\frac{\Gamma(1-ia-ib)}{\Gamma(1-ia)\Gamma(1-ib)} \ . \tag{12.16}$$

The Wronskian of the 2nd order wave equation is interpreted as the conserved flux,

$$\mathcal{F} \equiv \frac{1}{2i}\left[hr^3 R^* \partial_r R - cc\right] , \qquad \frac{d\mathcal{F}}{dr} = 0 \ . \tag{12.17}$$

We would like to compare the incoming flux at infinity to the transmitted flux entering the horizon. The far solution, expanded near infinity, is

$$R_{far} \approx \frac{1}{2r^{3/2}}\left[e^{i\omega r}\left(Ae^{-3\pi i/4} - Be^{-i\pi/4}\right) + e^{-i\omega r}\left(Ae^{3i\pi/4} - Be^{i\pi/4}\right)\right] \ . \tag{12.18}$$

Thus the incoming flux is

$$\mathcal{F}_{in} = -\omega\left|\frac{A}{2}\right|^2 \ . \tag{12.19}$$

Using the same formula to calculate the flux through the horizon gives the absorbed flux

$$\mathcal{F}_{abs} = -r_0^2(a+b)|\tilde{A}|^2 \ . \tag{12.20}$$

The ratio of absorbed flux is (exercise!)

$$R_{abs} = \frac{\mathcal{F}_{ab}}{\mathcal{F}_{in}} = \frac{\omega^4\pi^2(r_1^2 r_5^2)}{4}\frac{e^{\omega/T_H}-1}{(e^{\omega/2T_L}-1)(e^{\omega/2T_R}-1)} \ . \tag{12.21}$$

Greybody factors

This is the greybody factor that appears in Hawking emission, up to a factor. The factor is required since the relation between spherical waves that we considered and plane waves is

$$e^{-i\omega z} = K\frac{e^{-i\omega r}}{r^{3/2}}Y_{000} + \cdots \tag{12.22}$$

where Y_{000} is the s-wave spherical harmonic on S^3. The constant is $K = \sqrt{4\pi/\omega^3}$. Therefore the absorption cross section for a plane wave is

$$\sigma_{abs} = |K|^2 R_{abs} \ . \tag{12.23}$$

This is the greybody factor.

Exercise: Far and near Hawking temperatures

Difficulty: straightforward

(a) Calculate the Hawking temperature of (11.8) (with $\sigma = 0$, *i.e.,* no rotation).

(b) Now calculate the Hawking temperature of the BTZ black hole that appears in the near horizon, (11.18) (again with $\sigma = 0$, so $w_- = 0$).

Note that when we took the near-horizon limit of the near-extremal string, we sent $r_0 \to 0$. So any finite temperature of the BTZ is actually *zero* temperature as viewed from asymptotically flat infinity. There is an infinite redshift between the near horizon region and infinity.

Exercise: Scattering of a massive scalar

Difficulty: difficult

The wave equation for a massive scalar is

$$\Box\chi = m^2\chi \ . \tag{12.24}$$

In this problem we will derive the absorption cross section of a low-energy massive scalar on the near-extremal black string.

(a) Derive the full radial wave equation from (12.24), in the black string geometry (11.8) (but with $\sigma = 0$).

*(b)*Find the 'near-region' wave equation by starting with your answer to part (a) and assuming $r \ll r_{1,5}$ and $r\omega \ll 1$.

(c) Show that your near-region wave equation is identical to the massive wave equation on the BTZ $\times S^3$ geometry,

$$ds^2_{near} = \ell^2\left[-(\tilde{r}^2 - r_0^2)d\tilde{t}^2 + \frac{d\tilde{r}^2}{\tilde{r}^2 - r_0^2} + \tilde{r}^2 d\tilde{\phi}^2 + d\Omega_3^2\right] \ . \tag{12.25}$$

(Where the tilded coordinates are proportional to the original coordinates.)

(c) Find the ingoing solution of the wave equation in the near region. Do not bother with the far-region solutions, since these are messy and nothing interesting happens in the far region.

Hint: Mathematica cannot solve this wave equation without some coaxing. To simplify it, first change variables so $h = 1 - r_0^2/r^2$ is your independent variable. Then define $R(h) = (1-h)^a h^b \psi(h)$, with $a = \frac{1}{2}(1 + \sqrt{1+\ell^2 m^2})$ and $b = -i\omega\ell^2/2r_0$. Then Mathematica can solve it. This is also the best method to solve it by hand (*i.e.,* first strip out the singular points, then reduce the result to a standard hypergeometric equation).

(d) Show that in the matching region $r_0 \ll r \ll r_{1,5}$, the field behaves as

$$R_{near} \approx S r^{-d+\Delta} + F r^{-\Delta} \tag{12.26}$$

where $d = 2$,* S and F are numbers (possibly functions of ω), and

$$m^2 = \Delta(d - \Delta) \; , \qquad \Delta = \frac{d}{2} + \sqrt{\frac{d^2}{4} + m^2\ell^2} \; . \tag{12.27}$$

(e) If we think of the S term as the source, or ingoing term, and F as the response, or outgoing term,† argue that the absorption cross section (of the near region) is proportional to the imaginary part of the ratio,

$$P_{abs} \equiv \mathrm{Im} \frac{F}{S} \tag{12.28}$$

and compute P_{abs}. (Including the far region too would just contribute some overall uninteresting factors.)

The correct answer looks like

$$P_{abs} = k \sinh(2\pi n) |\Gamma(\frac{\Delta}{2} + in)|^4 \tag{12.29}$$

*I've only written d so that the result is true in higher-dimensional AdS_{d+1}/CFT_d. In this problem always set $d = 2$.

†The words 'ingoing' and 'outgoing' are not quite accurate here since these are power-law solutions, not traveling waves, in the matching region. But they have a similar interpretation.

where $k = k(r_0, \Delta)$ is a simple constant you should find, and $n \equiv \ell^2\omega/(2r_0)$.

(f) Define the *retarded Green's function*

$$G_R = \frac{F}{S} \ . \tag{12.30}$$

This measures the response of the field to adding a source. (The relation (12.28) is a version of the optical theorem for this Green's function.)

Find G_R in the high-frequency limit $\omega/T_H \gg 1$. (This is the correlator in the extremal limit, where temperature goes to zero.) You should find a power law. This power-law behavior at short distances is the hallmark of a conformal field theory.*

(g) The zero-temperature 2pt function of a 2d CFT is

$$\langle O(x^+, x^-)O(0)\rangle = |x|^{-2\Delta} = (x^+x^-)^{-\Delta} \tag{12.31}$$

where Δ is the scaling dimension of the operator. Take the 2d Fourier transform,

$$G(\omega_L, \omega_R) \sim \int dx^+dx^- e^{i\omega_L x^+ + i\omega_R x^-}(x^+x^-)^{-\Delta} \ . \tag{12.32}$$

Don't worry about the coefficient; we only care about the power law, so you can do this Fourier transform by dimensional analysis. Check that for $\omega_L = \omega_R = \omega$, your answer agrees with part (f). Therefore, the quantity Δ that we introduced in the process of the solving the wave equation is equal to a CFT scaling dimension.

Exercise: Quasinormal modes

The scattering modes that we found above are modes that obey a single boundary condition: ingoing at the horizon. Such modes have a continuous spectrum. A *quasi-normal mode* is a mode that obeys two boundary conditions: ingoing at the horizon, and outgoing far away from the black hole. These have a discrete spectrum. They are *quasi*-normal instead of normal because they decay (as flux falls into the black hole) so the the discrete frequencies have imaginary parts. If you perturb a black hole from the vicinity of the horizon, the 'ringdown' is (roughly) described by quasinormal modes.

*Why did we have to take the high-energy limit to see this? The answer is that the temperature introduces a scale; correlators in a CFT are only scale-invariant in the vacuum.

The quasinormal modes of BTZ are modes which are ingoing at the horizon and have $S = 0$ in (12.26).

(a) Find the spectrum of of quasinormal modes ω_n for a massless scalar in BTZ.

(b) If you did the previous exercise, then use your solution of the near-region wave equation to find the spectrum of quasinormal modes ω_n for a massive scalar in BTZ.

13 Absorption cross section from the dual CFT

Now we will reproduce the absorption cross section (12.21) using holography. This requires introducing some elements of conformal field theory. We will be more systematic about CFT and about the AdS/CFT correspondence later, for now we are just going to work this example in full detail as an illustration.

13.1 Brief Introduction to 2d CFT

(*References:* Polchinski's String Theory Ch 2 is a brief introduction. For a more detailed systematic introduction to 2d CFT, see chapters 4-6 (especially chapter 5) of the (highly recommended!) book *Conformal Field Theory* by Di Francesco et al.

Consider a 2d QFT on the Euclidean plane $\mathbf{R}^2$, with coordinates x_1 and x_2. It is very convenient to use the complex coordinates

$$z = x_1 + ix_2, \qquad \bar{z} = x_1 - ix_2 \ . \tag{13.1}$$

We take the flat metric on the plane,

$$ds^2 = (dx_1)^2 + (dx_2)^2 = dzd\bar{z} \ . \tag{13.2}$$

2d Conformal Transformations

A *conformal transformation* is a coordinate transformation that leaves the metric unchanged, up to an overall rescaling:

$$ds^2 = dzd\bar{z} \to e^{\sigma(w,\bar{w})}dwd\bar{w} \ . \tag{13.3}$$

First we want to find what type of coordinate changes have this special property. To this end, consider an arbitrary coordinate change $z = f(w,\bar{w}), \bar{z} = \bar{f}(w,\bar{w})$ where $\bar{f}$ is the complex conjugate of f. The metric in $(w,\bar{w})$ coordinates is

$$ds^2 = \left(\frac{\partial f}{\partial w}dw + \frac{\partial f}{\partial \bar{w}}d\bar{w}\right)\left(\frac{\partial \bar{f}}{\partial w}dw + \frac{\partial \bar{f}}{\partial \bar{w}}d\bar{w}\right) \ . \tag{13.4}$$

For this to have the form (13.3) we must impose

$$\frac{\partial f}{\partial w}\frac{\partial \bar{f}}{\partial w} = \frac{\partial f}{\partial \bar{w}}\frac{\partial \bar{f}}{\partial \bar{w}} = 0 \ . \tag{13.5}$$

This is equivalent to the condition that f is a holomorphic function,

$$f = f(w), \qquad \bar{f} = \bar{f}(\bar{w}) \ . \tag{13.6}$$

Thus conformal transformations in two dimensions are equivalent to holomorphic coordinate changes. The conformal group is the group of holomorphic maps. This is infinite dimensional, since you need an infinite number of parameters to specify a whole function. Note that this is not the case in higher dimensions; the conformal group in $d > 2$ dimensions is the finite-dimensional group $SO(d, 2)$.

Mapping the plane to the cylinder

A very important conformal transformation is the mapping of the z-plane to the w-cylinder. The mapping is

$$z = e^{-iw/R} \ , \qquad \bar{z} = e^{i\bar{w}/R} \ . \tag{13.7}$$

The w coordinate labels a cylinder, since if we take $w \to w + 2\pi R$ we get back to where we started. That is, w is identified,

$$w \sim w + 2\pi R \ . \tag{13.8}$$

This circle is a circle of constant magnitude on the z plane. (Draw the pictures for yourself.) If we split w into real coordinates,

$$w = \sigma_1 + i\sigma_2 \ , \tag{13.9}$$

then $\sigma_1 \sim \sigma_1 + 2\pi R$ is the circle and σ_2 is infinite. Negative values of σ_2 correspond to small circles on the z plane, and larger values of σ_2 correspond to increasingly larger circles on the z plane.

Classical CFT

At the classical level, a QFT has conformal symmetry if the action is invariant under conformal transformations. For example consider the action of a free massless scalar

$$S = \int d^2z \partial\phi\bar{\partial}\phi \tag{13.10}$$

where we use the notation

$$\partial = \partial_z, \qquad \bar{\partial} = \partial_{\bar{z}} \ . \tag{13.11}$$

Perform the infinitesimal coordinate change $z \to w(z)$ and it is easy to check that the Jacobian in the measure cancels the factors that show up from $\partial\phi = \frac{dw}{dz}\partial_w\phi$. On the other hand the free massive scalar is not conformally invariant. This illustrates a general feature of conformal field theories: they do not have any dimensionful parameters. Dimensionful parameters set a scale and therefore are not compatible with the scale transformation $z \to \lambda z$, which is part of the conformal group (in any number of dimensions).

Quantum CFT

Classical conformal invariance does not necessarily imply quantum conformal invariance. This is familiar from QCD (setting all quark masses to zero) — this theory is classically scale invariant, but to define the quantum theory we must introduce a regulator, and this leads to the dimensionful QCD scale Λ_{QCD} with important physical consequence (like confinement), so QCD is certainly not scale-invariant or conformally invariant at the quantum level. From now on when we say 'CFT' we mean at the quantum level.

Primary operators

The local operators of a CFT must transform covariantly. *Primary operators*[*] transform with a simple rescaling,

$$O'(w,\bar{w}) = \left(\frac{dw}{dz}\right)^{-h}\left(\frac{d\bar{w}}{d\bar{z}}\right)^{-\bar{h}} O(z,\bar{z}) \tag{13.12}$$

[*]These are often called primary *fields*. The names are interchangeable. But remember that in CFT, a 'field' is not necessarily a fundamental field that appears in the Lagrangian and is integrated over in the functional integral. For example in the free massless scalar, $\partial\phi$ is a primary field.

where $(h, \bar{h})$ are called the *conformal weights*. Another common notation is

$$\Delta = h + \bar{h}, \qquad s = h - \bar{h} \ , \tag{13.13}$$

where Δ is the *scaling dimension* and s is the *helicity*. Δ is the weight under a constant rescaling $(x_1, x_2) \to (\lambda x_1, \lambda x_2)$, ie under

$$\delta z = \lambda z, \qquad \delta\bar{z} = \lambda\bar{z} \tag{13.14}$$

the operator transforms with a factor of $\lambda^{-\Delta}$. s is the helicity because it is the weight under a rotation $(x_1, x_2) \to (x_1 - \lambda x_2, x_2 + \lambda x_1)$, ie under

$$\delta z = \lambda z, \qquad \delta\bar{z} = -\lambda\bar{z} \tag{13.15}$$

the operator transforms with a factor of λ^{-s}. The absolute value $|s| = |h - \bar{h}|$ is the *spin* of the operator. (This is just the usual definition of spin, so for example in free field theory it corresponds to the number of Lorentz indices on a field.

Descendant operators are operators that you get from primaries by acting with conformal transformations. For example, $\partial O(z, \bar{z})$ is a descendant of $O(z, \bar{z})$. The transformation law for descendants is more complicated than (13.12) but is completely fixed by symmetry.

All local operators in a CFT are either *primary* or *descendant*. This ensures that correlation functions transform covariantly under the conformal group. For example, the 2-point function on the plane must have the form

$$\langle O_1(z_1, \bar{z}_1) O_2(z_2, \bar{z}_2) \rangle = \frac{C_{12}}{(z_1 - z_2)^{2h} (\bar{z}_1 - \bar{z}_2)^{2\bar{h}}} \tag{13.16}$$

where

$$h = h_1 = h_2, \qquad \bar{h} = \bar{h}_1 = \bar{h}_2 \ . \tag{13.17}$$

C_{12} is a constant, related to the normalization of the field. The two-point function vanishes if the conformal weights of the two fields are different.

The path-integral definition of $\langle O_1(z_1, \bar{z}_1) O_2(z_2, \bar{z}_2)\rangle$ in (13.16) is (up to normalization)

$$\langle O_1(z_1, \bar{z}_1) O_2(z_2, \bar{z}_2)\rangle = \int D\Phi O_1(z_1, \bar{z}_1) O_2(z_2, \bar{z}_2) e^{-S[\Phi]} \tag{13.18}$$

where Φ stands for the fundamental fields of the theory.* Recall from our discussion of Euclidean path integrals that the path integral on a half-plane prepares the vacuum state. Therefore in operator langauge,

$$\langle O_1(z_1, \bar{z}_1) O_2(z_2, \bar{z}_2)\rangle = {}_{line}\langle 0|O_1(z_1, \bar{z}_1) O_2(z_2, \bar{z}_2)|0\rangle_{line} \ , \tag{13.19}$$

where $|0\rangle_{line}$ is the vacuum state of the theory on an infinite line (which you can think of as the Im $z = 0$ axis).

13.2 2d CFT at finite temperature

Remember from our discussion of Euclidean path integrals that QFT at finite temperature in Lorentzian signature is related to Euclidean QFT on a cylinder, with periodic imaginary time. Now we will see this relation very explicitly in CFT.

Mapping to the cylinder via $w = iR \log z$, and applying the transformation law (13.12) to (13.16), we can easily find the cylinder correlation function

$$\langle O_{cyl}(w_1, \bar{w}_1) O_{cyl}(w_2, \bar{w}_2)\rangle \sim \frac{R^{-2h}}{\sin\left(\frac{w_1 - w_2}{2R}\right)^{2h}} \frac{R^{-2\bar{h}}}{\sin\left(\frac{\bar{w}_1 - \bar{w}_2}{2R}\right)^{2\bar{h}}} \ . \tag{13.20}$$

(The 'cyl' subscript is usually dropped, so functions of w are just assumed to be cylinder operators.)

Exercise: Conformally invariant 2-point functions

(a) Prove (13.16).

*Often we do not have a Lagrangian for a CFT, and there is no useful notion of the 'fundamental' fields. However, path integral manipulations are still useful. Even in non-Lagrangian theories we never get into trouble by pretending that there are some fundamental fields defining the functional integral.

(b) Derive (13.20), including the missing coefficient.

Note two things about this correlator: First, it is invariant under the cylinder periodicity $w_1 \sim w_1 + 2\pi R$.* Second, it has the same short-distance singularity as the plane correlator (13.20), *i.e.,*

$$\langle O_{cyl}(w_1, \bar{w}_1) O_{cyl}(w_2, \bar{w}_2)\rangle = \frac{C_{12}}{(w_1 - w_2)^{2h}(\bar{w}_1 - \bar{w}_2)^{2\bar{h}}} \quad \text{as} \quad w_1 \to w_2 \ . \tag{13.21}$$

This is always true in QFT: the short-distance behavior is fixed by vacuum correlation functions. (In fact these two conditions fix the function (13.20) uniquely, assuming some behavior at infinity, so we do not even strictly need the exponential mapping to derive (13.20).)

From the Lorentzian point of view, the cylinder correlator (13.20) can be interpreted different ways. To go to Lorentzian signature, write $w = \sigma_1 + i\sigma_2$ where $\sigma_{1,2}$ are real coordinates. If we think of σ_2 as 'time', then the Wick rotation to Lorentzian signature is $\sigma_2 = it$. In this case the circle coordinate σ_1 remains a circle in Lorentzian signature, so this Wick rotation gives the Lorentzian theory on the Lorentzian cylinder $S^1\times$ Time. This Wick rotation has nothing to do with finite temperature.

To get the finite temperature theory, we instead Wick-rotate by setting $\sigma_1 = it$. Thus

$$w \to i(t + x), \qquad \bar{w} = i(t - x) \ . \tag{13.22}$$

(Note that in Lorentzian signature, w and $\bar{w}$ are no longer complex conjugates.) This means that the the theory is periodic in imaginary time $t \sim t + 2\pi i R$. Comparing to the finite-temperature periodicity $t \sim t + i\beta$ with $\beta = T^{-1}$, we see that our Euclidean CFT is related to a finite-temperature CFT at temperature

$$\beta = T^{-1} = 2\pi R \ . \tag{13.23}$$

*There are some subtleties with branch cuts making this statement that we'll ignore for now, and it relies on the fact that $(-1)^{2(h-\bar{h})} = 1$ since operators must have integer or half-integer spin.

From (13.20), this means the finite-temperature Lorentzian correlator in CFT is*

$$\begin{aligned} G_\beta(t - i\epsilon, x) &= \text{Tr}\ e^{-\beta H} O(t-i\epsilon, x) O(0,0) \\ &\sim (-1)^{h+\bar{h}} \frac{(\pi T)^{2h}}{\sinh(\pi T(t+x))^{2h}} \frac{(\pi T)^{2\bar{h}}}{\sinh(\pi T(t-x))^{2\bar{h}}} \ . \end{aligned} \tag{13.24}$$

13.3 Derivation of the absorption cross section

We now return to the derivation of the absorption cross-section (12.21). Recall that we scattered a low-energy quantum from the near-extremal black string. The near horizon region relevant to this calculation was a BTZ black hole in AdS_3 (times S^3). We will set $T_L = T_R = T_H$ for simplicity, which corresponds to setting the parameter $\sigma = 0$ in the black string metric. From the point of view of the near-horizon, this sets the angular momentum of the BTZ black hole to zero.

In the gravity calculation (12.21), we found

$$\sigma_{abs} \sim \coth\left(\frac{\omega}{4T_H}\right) \ . \tag{13.25}$$

Now the claim is as follows:

We can replace the near-horizon geometry by 1+1d CFT at temperature $T = T_H$, living on a fictitious 'membrane' at the boundary of AdS_3.

This boundary was the matching location in our gravity calculation, ie some value of r in the range $r_0 \ll r \ll r_{1,5}$.

Which CFT is it?

We will only match the temperature dependence. The overall factor can also be matched by this method, up to a constant. We will not need to specify *which* CFT we are actually considering, we will just need some general properties of the CFT like the value of the temperature, and the existence of an operator with certain conformal weights. The microscopic definition of the *particular* CFT depends the particular

*Setting a convenient normalization, and introducing an $i\epsilon$ to keep track of operator ordering. Recall that the finite-temperature correlator is defined by ordering in Euclidean-time, so this sets the order of the operators in the trace as written.

theory of quantum gravity. The only known microscopic CFTs are the ones coming from string theory, since that is our only candidate theory of quantum gravity, but in principle there could be other CFTs corresponding to other UV completions of gravity. In the string theory examples, where the microscopic definition of the CFT is known, it is also possible to match the coefficient in the absorption calculation and it comes out correctly.

The interaction term

We want to scatter a scalar field against the CFT. We will assume that the bulk scalar field χ couples to a CFT operator O, thus adding to the CFT an interaction term

$$S_{int} = \int dtdxO(t,x)\chi(t,x,r=0) \ . \tag{13.26}$$

In this expression O is a CFT operator and $\chi(r=0)$ — the value of the bulk field at the fictitious membrane where the CFT lives — is treated as a classical source. We will assume that the space direction in the CFT is unwrapped, so we call it $x \in (-\infty,\infty)$ (previously called ϕ), though the S^1 version can also be done with some extra assumptions about the CFT. We also assume the source couples weakly to the CFT so that the interaction term (13.26) can be treated perturbatively.

Absorption rate

When we computed the absorption cross section, we assume

$$\chi = e^{-i\omega t}R(r) \ , \tag{13.27}$$

so $S_{int} \propto \int dtdxO(t,x)e^{-i\omega t}$. The transition amplitude from an initial state $|i\rangle$ to a final state $|f\rangle$ is given by Fermi's Golden Rule, as the matrix element of the interaction Hamiltonian

$$\mathcal{M}_{i\to f} \sim \langle f|\int dtdxO(t,x)e^{-i\omega t}|i\rangle \ . \tag{13.28}$$

The total absorption rate at temperature β is computed by summing this over final states, and averaging over initial states with a thermal ensemble,

$$\Gamma_{abs} \sim \sum_{i,f} e^{-\beta E_i}\int dt_1dx_1dt_2dx_2e^{-i\omega(t_1-t_2)}\langle i|O(t_2,x_2)|f\rangle\langle f|O(t_1,x_1)|i\rangle \tag{13.29}$$

The sum over $|f\rangle$ is just the identity, so up to an overall factor of volume (which you can think of as the momentum-conserving delta function $\delta(0)$),

$$\Gamma_{abs} \sim \int dtdx e^{-i\omega t} \sum_i e^{-\beta E_i} \langle i|O(t,x)O(0,0)|i\rangle \ . \tag{13.30}$$

This sum is the definition of the thermal 2-point function,

$$\Gamma_{abs} \sim \int dtdx G_\beta(t-i\epsilon, x) e^{-i\omega t} \tag{13.31}$$

This thermal correlator was calculated in (13.24). To take the Fourier transform, use the integral

$$\int dy e^{-i\omega y}(-1)^h \left(\frac{\pi T}{\sinh\left[\pi T(y \pm i\epsilon)\right]}\right)^{2h} = \frac{(2\pi T)^{2h-1}}{\Gamma(2h)} e^{\pm\omega/2T} \left|\Gamma(h + i\frac{\omega}{2\pi T})\right|^2 \ . \tag{13.32}$$

First take the Fourier transform assuming indepdendent left and right momenta

$$G_\beta(\omega_L, \omega_R) = (-1)^{h+\bar{h}} \int dtdx e^{-i\omega_L(t+x) - i\omega_R(t-x)} \frac{(\pi T)^{2h}}{\sinh(\pi T(t+x))^{2h}} \frac{(\pi T)^{2\bar{h}}}{\sinh(\pi T(t-x))^{2\bar{h}}}$$

and then set $\omega_L = \omega_R = \omega$. The absorption rate is given by the difference of absorption and emission. These correspond to two different $i\epsilon$ prescriptions (Exercise: why?). So finally

$$\sigma_{abs} \sim \Gamma_{abs} - \Gamma_{emit} \tag{13.33}$$

$$\sim \int dtdx e^{-i\omega t}\left[G(t-i\epsilon, \phi) - G(t+i\epsilon, \phi)\right] \tag{13.34}$$

$$\sim 2\frac{(2\pi T)^{2(h+\bar{h})-2}}{\Gamma(2h)\Gamma(2\bar{h})} \sinh\left(\frac{\omega}{2T}\right) \left|\Gamma(h + i\frac{\omega}{4\pi T})\Gamma(\bar{h} + i\frac{\omega}{4\pi T})\right|^2 \tag{13.35}$$

This matches the gravity answer (13.25) if we set

$$h = \bar{h} = 1 \tag{13.36}$$

and use the identity $|\Gamma(1+ix)|^2 = \pi x/\sinh(\pi x)$. Why should the weight be (13.36)? For now, we just pick them so the answer works out. In general the weights depend on the mass and spin of the bulk field, and (13.36) is the correct choice for a massless bulk field. We will treat this more systematically below.

13.4 Decoupling

The upshot of the last few sections is that

$$\begin{pmatrix} \text{Far-region gravity} \\ + \\ \text{gravity in } AdS_3 \times S^3 \end{pmatrix} \quad = \quad \begin{pmatrix} \text{Far-region gravity} \\ + \\ CFT_2 \text{ on } \ AdS_3 \text{ boundary} \end{pmatrix} .$$

In the gravity calculation, we assumed near-extremal but not exactly extremal. This retained some coupling between the near-horizon degrees of freedom, and the fields in the asymptotically flat far region. Similarly, in CFT, we assume a weak coupling between gravity fields and CFT fields.

If we take $T_H \to 0$, the far region and near regions decouple. This is Maldacena's decoupling limit. In this limit we can completely drop the asymptotically flat part of the calculation, and we are left with the (3d version of the) AdS/CFT correspondence:

$$\text{gravity in } AdS_3 \times S^3 \qquad = \qquad CFT_2 \text{ on } \ AdS_3 \text{ boundary} . \tag{13.37}$$

CFTs are UV-complete, so this duality defines not only low-energy effective gravity, but a UV-complete theory of gravity on $\text{AdS}_3 \times S^3$.

From now on, we will just forget about the 'far region.' It is not needed, and is rarely used in modern AdS/CFT.

14 The Statement of AdS/CFT

14.1 The Dictionary

Choose coordinates

$$ds^2 = \frac{\ell^2}{z^2}(dz^2 + dx^2) \tag{14.1}$$

on Euclidean AdS_{d+1}, where x is a coordinate on R^d. The boundary is at $z = 0$.

We showed above that scattering problems in gravity map to correlation functions in CFT. In this relation the boundary value of the bulk field acted as a source for a CFT operator. This is generalized by the following statement of the AdS/CFT correspondence:

$$Z_{grav}[\phi_0^i(x); \partial M] = \left\langle \exp\left(-\sum_i \int d^d x \phi_0^i(x) O^i(x) \right) \right\rangle_{CFT\ on\ \partial M} \tag{14.2}$$

This is called the GKPW dictionary.* The index i runs over all the light fields in the bulk effective field theory, and correspondingly over all the low-dimension local operators in CFT.

The left-hand side

The lhs of (14.2) is the gravitational partition function in asymptotically AdS space. It is formally computed by the same path integral that we discussed in the context of black hole thermodynamics. Since AdS has a boundary, we must provide boundary conditions to define this path integral. The boundary conditions on bulk scalars are

$$\phi^i(z, x) = z^{d-\Delta}\phi_0^i(x) + \text{subleading as } \ z \to 0 \ . \tag{14.3}$$

where the mass of the bulk scalar is related to the scaling dimension of the CFT operator by

$$m^2 = \Delta(d - \Delta) \ , \qquad \Delta = \frac{d}{2} + \sqrt{\frac{d^2}{4} + m^2\ell^2} \ . \tag{14.4}$$

*After hep-th/9802109 by Gubser, Klebanov and Polyakov and hep-th/9802150 by Witten. I highly recommend reading Witten's paper.

We will see below that (14.3) is the leading solution of the wave equation for a bulk scalar of mass m.

Similar statements apply to all bulk fields, including the metric, though the boundary condition and formula for the dimension is slightly modified for fields with spin. The boundary conditions on the metric involve a choice of topology as well as the actual metric, which is why we've indicated explicitly that Z_{grav} depends on the boundary manifold ∂M.

The right-hand side

The rhs of (14.2) is the generating functional of correlators in a CFT. In this equation the $\phi_0^i(x)$ are sources, and the $O^i(x)$ are CFT operators. Denoting the rhs of (14.2) by $Z_{cft}[\phi_0]$, correlation functions are computed in the usual way,

$$\langle O_1(x_1)\cdots O_n(x_n)\rangle_{CFT} \sim \frac{\delta^n}{\delta\phi_0^1(x_1)\cdots\delta\phi_0^n(x_n)} Z_{cft}[\phi_0]\big|_{\phi_0^i=0} \ . \tag{14.5}$$

The mapping

Each light field in gravity corresponds to a local operator in CFT. The spin of the bulk field is equal to the spin of the CFT operator; the mass of the bulk field fixes the scaling dimension of the CFT operator. Here are some examples:

Scalar: A bulk scalar field $\chi(z,x)$ is dual to a scalar operator in CFT. The boundary value of χ acts as a source in CFT. This is exactly the relationship we used in our derivation of the absorption cross section of the black string.

Graviton: Every theory of gravity has a massless spin-2 particle, the graviton $g_{\mu\nu}$. This is dual the stress tensor $T_{\mu\nu}$ in CFT. This makes sense since every CFT has a stress tensor. The fact that the graviton is massless corresponds to the fact that the CFT stress tensor is conserved. It also fixes the scaling dimension to $\Delta_T = d$. We will see this in more detail later.

Vector: If our theory of gravity has a spin-1 vector field A_μ, then the dual CFT has a spin-1 operator J_μ. If A_μ is massless, then $\Delta_J = d-1$ and J_μ is a conserved current. Otherwise, $\Delta_J > d-1$ and the current is not conserved.

This illustrated a general and important feature of AdS/CFT: *gauge symmetries in the bulk correspond to global symmetries in the CFT.*

This is UV complete.
Note that CFTs are UV complete. Therefore (14.2) is a non-perturbative formulation of a UV complete theory of quantum gravity. Shockingly, it is a definition of gravity from a QFT without gravity. This is very powerful because we understand QFT relatively well.

14.2 Example: IIB Strings and $\mathcal{N} = 4$ Super-Yang-Mills

In some sense, it is believed that the AdS/CFT correspondence as summarized by (14.2) holds for any theory of gravity and and CFT. That is, given a theory of gravity we can use it to *define* a CFT via (14.2), and (perhaps) vice-versa. But aside from certain examples, the correspondence is well defined and useful only in certain limits. To illustrate this we turn to a specific example where AdS/CFT is understood in great detail. This is the duality between IIB string theory and supersymmetric gauge theory:

$$\text{IIB strings on } \; AdS_5 \times S^5 \quad = \quad \text{Yang-Mills in 4d with } \; \mathcal{N} = 4 \text{ supersymmetry} \; .$$

The gravity side
The string theory has adjustable scales $\ell \equiv \ell_{AdS}$, the Planck scale ℓ_P, and the string scale ℓ_s. We do not need to use any details of string theory except to say that at low energies, the effective action is Einstein + Matter + higher curvature corrections suppressed by the string scale:*

$$S_{IIB} \sim \frac{1}{G_N} \int \sqrt{g} \left(R + L_{matter} + \ell_s^4 R^4 + \cdots \right) \tag{14.6}$$

The stringy states have masses of order $1/\ell_s^2$, so at energies below $1/\ell_s^2$ it is just an ordinary effective field theory like we discussed at the beginning of the course.

*There are no R^2 corrections allowed with this amount of supersymmetry, but there are similar examples with non-zero $\ell_s^2 R^2$ terms.

The CFT side

$\mathcal{N} = 4$ Super-Yang-Mills is a highly supersymmetric gauge theory in 4d. Its matter context is fixed uniquely by supersymmetry. It is just an $SU(N)$ gauge field plus all the matter fields required by supersymmetry, which include matrix-valued scalar fields transforming the adjoint representation of $SU(N)$ (unlike the fundamental representations we usually encounter in, say, QCD).

The gauge theory has two dimensionless parameters, N (ie the size of $SU(N)$) and the Yang-Mills coupling constant g_{YM}. Define the combination

$$\lambda = g_{YM}^2 N \ . \tag{14.7}$$

This is called the 't Hooft coupling. It turns out that gauge theory at large N is most naturally organized as an expansion in λ and $1/N$, rather than g_{YM} and $1/N$. This is roughly because there are N fields running in loops, which changes the expansion parameter from g_{YM}^2 to λ.

The mapping

The mapping from string theory parameters to CFT parameters is

$$\lambda \sim \left(\frac{\ell_{AdS}}{\ell_{string}}\right)^4 \tag{14.8}$$

and

$$\frac{\ell_{AdS}^{d-1}}{G_N} \sim \left(\frac{\ell_{AdS}}{\ell_P}\right)^{d-1} \sim N^2 \ . \tag{14.9}$$

(with known coefficients). We will see where this particular scaling comes from below in more generality. For now we just want to note that this is a *strong/weak duality*: when one side is easy, the other is (usually) hard. For example to have semiclassical Einstein gravity, both loops and higher curvature corrections must be suppressed on the gravity side. This means $N \gg 1$ and $\lambda \gg 1$ so the CFT is very strongly coupled. On the other hand if we consider a weakly coupled CFT, then $\ell_s \gg \ell_{AdS}$ so stringy/higher curvature corrections are not suppressed on the gravity side and this presumably behaves nothing like ordinary gravity. (This is related to so-called 'higher spin gravity' or 'Vasiliev gravity'.)

14.3 General requirements

Returning to AdS/CFT in general, we can make some similar observations about when it produces a nice semiclassical theory of gravity. This requires as least two things:

1. **Strongly coupled CFT.** If the CFT is weakly coupled, then there are too many operators. For example, a free scalar field ψ leads to conserved currents of every integer spin:*

$$\psi\partial_\mu\psi, \qquad \psi\partial_\mu\partial_\nu\psi, \qquad \psi\partial_\mu\partial_\nu\partial_\rho\psi, \quad \text{etc.} \tag{14.10}$$

   On the gravity side, this would require lots of massless or very light high-spin states. This is something we expect in string theory at high enough energies but not in our low energy effective field theory.

   So we must require that the CFT has a sparse spectrum of low-dimension operators. This is sometimes called a large 'gap' in the spectrum, meaning a gap between the low-energy fields and the stringy stuff. This can only happen at strong coupling, although there can also be strongly coupled theories with no gap which therefore do not have nice gravity duals.

2. **Large N_{dof}.** In the super-Yang-Mills example, we said $G_N \sim 1/N^2$ so that the large number of degrees of freedom is required for gravity to be weakly coupled. This is true in general, too. There are two ways to see this, both of which we will discuss in more detail later. I will purposely be a little vague about the definition of N_{dof} since there are several reasonable ways to define it, and they are all different.

   First, note that black hole entropy is $S \propto 1/G_N$, which is very large. Since entropy is the log of the density of states, this means holographic CFTs must have an enormous degeneracy of states at high energy. This means there are lots of degrees of freedom. For example, a 2d CFT consisting of N_b free bosons has $S(E) \propto \sqrt{N_b E}$.

   Second, we can roughly measure the degrees of freedom by looking at the stress-tensor 2pt function. This is fixed by conformal invariance up to a single coeffi-

*This is schematic, you must add corrections to these operators for them to be conserved by the EOM.

cient:

$$\langle T_{\mu\nu}(x)T_{\alpha\beta}(y)\rangle = c \times (\text{known function of } x, y) \ . \tag{14.11}$$

The coefficient c is a measure of degrees of freedom.* Consider again lots of free fields: the stress tensors add, so the total stress tensor will have a very big 2pt function.

On the gravity side, the stress tensor is dual to the graviton. We will see in detail below how to calculate corelators, but for now suffice it to say that $\langle TT\rangle_{cft}$ will be related to a graviton scattering experiment $\langle gg\rangle_{gravity} \sim 1/G_N$. Thus $c \sim 1/G_N$ and we see again that weakly coupled gravity requires an enormous number of degrees of freedom.

Clear there is a tension between requirements (1) and (2). We want lots of degrees of freedom, and lots of states at high energies, but very few states at low energies. Roughly speaking, you can think of this as the requirement that the CFT is *confining*: it has lots of states at high energies, but very few at low energies where quarks are confined. Later we will see a very direct link between black hole thermodynamics and confinement.

14.4 The Holographic Principle

Many years ago Bekenstein conjectured that the maximum entropy you can fit into a region of space is equal to the entropy of the corresponding black hole:

$$S_{max} = \frac{\text{area}}{4G_N} \ . \tag{14.12}$$

This is called the *Bekenstein bound*. The argument is simple. If you have lots of stuff in a region and $S_{stuff} > S_{blackhole}$, then you can throw in some more stuff and form a black hole. In doing so, the entropy of the system decreases! Therefore the second law requires a bound like (14.12).†

This bound inspired 't Hooft (in '93) and later Susskind (in '94) to argue that a theory

*But not an entirely satisfactory one. For example, it can increase under RG flow.

†In the last few years this bound has been understood much better using entanglement entropy. See for example 1404.5635 and references therein.

of quantum gravity must secretly live in fewer dimensions than our observed spacetime. This principle is realized concretely by AdS/CFT.

15 Correlation Functions in AdS/CFT

As a first application of (14.2), we will use the gravity side to derive the correlators of a conformal field theory. First we'll start with a purely QFT discussion of correlators in a theory with conformal invariance, then reproduce these results from gravity.

By the way, you've already seen one example of CFT correlators compute from gravity: the absorption cross section calculation. That was related to a CFT correlator at finite temperature. In this section are deriving correlators in the vacuum state, *i.e.,* empty AdS.

15.1 Vacuum correlation functions in CFT

This will be a brief introduction to CFT. For details, see: Polchinski's String Theory book; Kiritsis's String Theory book; or the big yellow CFT book by Di Francesco et al.

The group of conformal symmetries of R^d is $SO(d+1,1)$. In Lorentz signature, the conformal symmetries of $R^{d-1,1}$ are $SO(d,2)$. The generators of $SO(d,2)$ are

$$\begin{aligned} P_\mu &= -i\partial_\mu \\ L_{\mu\nu} &= -i(x_\mu\partial_\nu - x_\nu\partial_\mu) \\ D &= -ix^\mu\partial_\mu \\ K_\mu &= -i(2x_\mu x^\nu\partial_\nu - x^2\partial_\mu) \end{aligned} \tag{15.1}$$

The the first two lines are translations, rotations, and boosts; these generate the Poincare group (which is 10-dimensional in $d=4$). The 3rd line is the dilatation, or scale generator, since under $x^\mu \to x^\mu + i\epsilon D^\mu$, the coordinate is just rescaled, $x^\mu \to x^\mu(1+\epsilon)$. The last line is called the special conformal transformation.

One way to derive (15.1) is to find the conformal Killing vectors of Minkowski space. These are defined to be vectors V^μ obeying

$$\mathcal{L}_V \eta_{\mu\nu} = f(x)\eta_{\mu\nu} \ , \tag{15.2}$$

where f is any function. This is the infinitessimal version of the definition of a conformal symmetry, which maps $ds^2 \to e^{\Omega(x)} ds^2$.

Operators in a CFT can be organized under representations of the conformal group. We define *primary operators* to obey*

$$
\begin{aligned}
[D, O(0)] &= -i\Delta O(0) && (15.3)\\
[K_\mu, O(0)] &= 0\ . && (15.4)
\end{aligned}
$$

The dilatation eigenvalue Δ is called the scaling dimension of O. The second condition is like a highest weight condition. We can build the full representation by acting on $O(x)$ with the conformal generators, so for example $\partial_m u O(x)$ is a *descendant operator*.

The finite version of of the D commutator says that under a rescaling $x \to \lambda x$, we have $O(x) \to \lambda^\Delta O(\lambda x)$. More generally, primaries obey

$$
O'(x') = \left|\det \frac{\partial x'^\mu}{\partial x^\nu}\right|^{-\Delta/d} O(x)\ . \tag{15.5}
$$

For a correlation function of n primaries this implies

$$
\langle O_1(\lambda x_1)\cdots O_n(\lambda x_n)\rangle = \lambda^{-\Delta_1-\Delta_2-\cdots-\Delta_n}\langle O_1(x_1)\cdots O_n(x_n)\rangle\ . \tag{15.6}
$$

The special conformal transformations also impose requirements on correlators. It turns out that all the conformal generators together completely fix the 2 and 3-point functions of a CFT, up to overall factors. The two-point function of equal-weight fields is

$$
\langle O_1(x_1) O_2(x_2)\rangle = \frac{c_{12}}{|x_1 - x_2|^{2\Delta}} \qquad (\Delta_1 = \Delta_2 \equiv \Delta) \tag{15.7}
$$

and it must vanish if $\Delta_1 \neq \Delta_2$. The number c_{12} can be rescaled by rescaling our normalization of the operators. Often we pick an orthonormal basis of primary operators, so that $c_{ij} = \delta_{ij}$.

*This is for scalars. Operators with spin would also have the usual rule for action by the Lorentz group, $[L_{\mu\nu}, O(0)] = \Sigma_{\mu\nu} O(0)$.

Similarly, the only 3-point function allowed by conformal invariance is

$$\langle O_1(x_1)O_2(x_2)O_3(x_3)\rangle = \frac{c_{123}}{|x_{12}|^{\Delta_1+\Delta_2-\Delta_3}|x_{23}|^{\Delta_2+\Delta_3-\Delta_1}|x_{31}|^{\Delta_3+\Delta_1-\Delta_2}} \ , \tag{15.8}$$

where

$$x_{ij} \equiv x_i - x_j \ . \tag{15.9}$$

The number c_{ijk}, called an OPE coefficient (for operator product expansion), is a real physical prediction of the theory, since we've already fixed normalizations via the 2-point function.

In fact, the set of scaling dimensions Δ_i and the OPE coefficients c_{ijk} are *all* the data of a CFT. This is because higher correlators can be computed, at least in principle, by sewing together 3-point functions and summing over intermediate states.

The 4-point function is not completely fixed by conformal symmetry, but it is highly constrained. With equal external weights $\Delta_{1,2,3,4} = \Delta$, the most general form of the 4-point correlator is

$$\langle O(x_1)O(x_2)O(x_3)O(x_4)\rangle = |x_{12}|^{-2\Delta}|x_{34}|^{-2\Delta}F(u,v) \tag{15.10}$$

where F is an arbitrary function of the *conformal cross ratios*,

$$u = \frac{x_{12}^2 x_{34}^2}{x_{13}^2 x_{24}^2}, \qquad v = \frac{x_{14}^2 x_{23}^2}{x_{13}^2 x_{24}^2} \ . \tag{15.11}$$

15.2 CFT Correlators from AdS Field Theory

We assume a strongly coupled, large-N CFT with a semiclassical holographic dual. This is a limit where the gravity theory is weakly coupled, $G_N \ll \ell$, and higher curvature corrections can be neglected, $\ell_{string} \ll \ell$ (here ℓ is the AdS radius). According to the GKPW dictionary (14.2), we can compute the generating function of CFT correlators on the gravity side by

$$\begin{aligned} Z_{cft}[\phi_0] \ &\equiv \ \langle e^{-\int \phi_0 O}\rangle_{CFT} && (15.12)\\ &\approx \ \exp\left(-S_{grav} + O(G_N^0) + O(\frac{\ell_{string}}{\ell_{AdS}})\right) && (15.13)\end{aligned}$$

where S_{grav} is the on-shell action for gravity subject to the boundary condition

$$\phi \to z^{-\Delta+d}\phi_0(x) \tag{15.14}$$

as we approach the AdS boundary $z \to 0$.

So to compute CFT correlators, we need to understand how to compute the classical action in AdS as a functional of the boundary conditions.

This material is explained very clearly in many places, so I will not repeat it here. I recommend reading Witten's original paper on the subject, where 'Witten diagrams' were introduced [hep-th/9802150]. In class, I followed, almost exactly Kiritsis's String Theory in a Nutshell sections 13.8.1 and 13.8.2. Read those sections before continuing!

15.3 Quantum corrections

So far we only used the classical theory on the gravity side. (Though on the CFT side, this is a strongly coupled QFT calculation which is not at all classical!) What happens when we include loop corrections in the gravity? The gravitational loop expansion is organized into powers of G_N. The classical term is $\sim 1/G_N$. If we compute Witten diagrams with loops, then we find an expansion in G_N.

On the CFT side, this is an expansion in $1/N_{dof}$, since recall the dictionary $\ell^{d-1}/G_N \sim N_{dof}$.

This implies something very special about CFTs with a semiclassical holographic dual: These CFTs, although strongly coupled, have a meaningful expansion in $1/N_{dof}$. Defining $N_{dof} = N^2$ (since this notation holds in $SU(N)$ gauge theory), this can be restated as the fact that connected correlation functions are suppressed. That is, if we normalize our operators by setting

$$\langle OO \rangle \sim 1 \tag{15.15}$$

(with the appropriate factors of x suppressed), then the 3-point function is suppressed,

$$\langle OOO \rangle \sim \frac{1}{N} \tag{15.16}$$

and higher-point functions are dominated by their connected piece:

$$\langle OOOO\rangle \sim \langle OO\rangle\langle OO\rangle + O(1/N^2) \ . \tag{15.17}$$

The explicit theories we know of with this sort of behavior are large-N gauge theories. These have been studied for a long time, starting with a beautiful paper by 't Hooft in the 70s where he showed that the Feynman diagrams of $SU(N)$ gauge theory in the large-N limit naturally reorganize themselves into something that looks roughly like a string theory. I will not cover this, but I highly recommend you read about it in section 13.1 of Kiritsis, or the big AdS/CFT review [hep-th/9905111].

Another consequence of the weak coupling constant G_N on the gravity side is that gravity has an approximate Fock space. That is, if we have a weakly coupled scalar field on the gravity side, then we can construct 1-particle states, 2-particles states, etc, by acting with creation operators. On the CFT side, this means for example that if we have a primary O_1 of dimension Δ_1, and another primary O_2 of dimension Δ_2, then there is a third primary O_{1+2} of dimension

$$\Delta_{O_{1+2}} \approx \Delta_1 + \Delta_2 + O(1/N) \ . \tag{15.18}$$

This is very special; it does not happen in general CFT, where states are just a some strongly coupled mess and there is no way to 'add' some stuff to other stuff without getting large corrections to the conserved charges from the strong interactions.

Following the gauge theory language, the operators dual to single bulk fields are called 'single-trace operators', and the operators like O_{1+2} are called 'multi-trace operators' and usually just denoted by the product $O_1 O_2$ (or more complicated things like $O_1 \Box^n \partial_{\mu_1\cdots\mu_\ell} O_2$).

In words, (15.18) says that in a CFTs with a semiclassical holographic dual, low dimension operators have 'small anomalous dimensions.' I've restricted this statement to low-dimension operators because these are the operators dual to bulk fields; high dimension operators are dual to non-perturbative stuff like black hole microstates.

16 Black hole thermodynamics in AdS_5

Now we return to black holes, and some of the techniques introduced in the beginning of the course.

The basic starting point is that *thermal states in CFT are dual to black holes in quantum gravity.* In fact, this is a special case of the dictionary (14.2), where we impose boundary conditions appropriate for thermal field theory. That is,

$$Z_{cft}[\phi_0; M] = Z_{grav}[\phi_0; \text{boundary} = M] \tag{16.1}$$

where we take the manifold on which the CFT lives to be

$$M = \Sigma_{d-1} \times S^1_\beta \ . \tag{16.2}$$

Here Σ_{d-1} is space. We will mostly set $\Sigma_{d-1} = S^3_\ell$, a 3-sphere of size ℓ. And S^1_β is a circle of size β. As we saw earlier in the course, the Euclidean path integral on $\Sigma_{d-1} \times S^1_\beta$ defines the finite-temperature state on Σ_{d-1}.

The meaning of the notation in (16.1) is that we calculate the gravity partition with boundary condition ϕ_0 on bulk fields, and boundary condition M on the bulk manifold. Explicitly, fields obey the usual fall-off $\phi \sim r^{-d+\Delta}\phi_0(x)$ as $r \to \infty$, and the metric itself obeys the boundary condition

$$ds^2 \to \frac{r^2}{\ell^2}dt_E^2 + \frac{\ell^2}{r^2}dr^2 + r^2 d\Omega_3^2 \ , \qquad t_E \sim t_E + \beta \ . \tag{16.3}$$

Our goal is to compute the free energy at temperature β. For this we can turn off all fields besides the metric, so $\phi_0 = 0$, and we just have the relation

$$Z_{cft}[\beta] = Z_{grav}[\beta] \ . \tag{16.4}$$

We will compute the rhs in gravity, and interpret it in CFT. The result will exhibit rich behavior, including phase transitions as a function of temperature. This will turn out to be related to confinement/deconfinement in the CFT.

16.1 Gravitational Free Energy

To compute $Z_{grav}[\beta]$, in principle, we should compute the quantum gravity path integral subject to the boundary condition (16.3). That's impossible, but in the semiclassical limit we can evaluate it approximately by expanding around classical solutions of the equations of motion. We need to find all of the classical solutions that obey this boundary condition, and evaluate their on-shell actions using

$$I_E = -\frac{1}{16\pi G_N}\int d^5x\sqrt{g}\left(R+\frac{12}{\ell^2}\right) \ . \tag{16.5}$$

If there are several solutions, then the semiclassical approximation to the path integral is

$$Z_{grav}(\beta) \approx e^{-I_E^{(1)}} + e^{-I_E^{(2)}} + \cdots \tag{16.6}$$

Each saddlpoint also comes with an infinite series of perturbative (loop) corrections but we won't worry about those, we will just evaluate the classical contributions.

There are three classical solutions (in pure gravity) obey the thermal boundary condition (16.3): small black holes, large black holes, and thermal AdS.

16.1.1 Schwarzschild-AdS

The Euclidean black hole satisfying the boundary condition (16.3) is called Schwarzschild-AdS, with metric

$$ds^2 = f dt_E^2 + \frac{dr^2}{f} + r^2 d\Omega_3^2 \ , \quad f = 1 + \frac{r^2}{\ell^2} - \frac{\mu}{r^2} \ , \tag{16.7}$$

with the thermal identification

$$t_E \sim t_E + \beta \ . \tag{16.8}$$

μ is a constant that will be related to the mass. The explicit metric on the unit 3-sphere is

$$d\Omega_3^2 = d\psi^2 + \sin^2\psi(d\theta^2 + \sin^2\theta d\phi^2) \ , \tag{16.9}$$

where ψ and θ run from 0 to π and $\phi \in [0, 2\pi)$. The horizon is the outermost solution of the equation $f(r_+) = 0$, which gives

$$r_+^2 = \frac{\ell^2}{2}\left(-1 + \sqrt{1 + \frac{4\mu}{\ell^2}}\right) . \tag{16.10}$$

As usual, the condition that this solution is non-singular at $r = r_+$ relates β to r_+. From the path integral point of view, this is because only smooth classical solutions are good saddlepoints; solutions with a conical defect do not satisfy the equations of motion at the defect. The conical defect trick gives

$$\beta = \frac{2\pi\ell^2 r_+}{2r_+^2 + \ell^2} , \tag{16.11}$$

ie

$$r_+ = \frac{\pi\ell^2}{2\beta}\left[1 \pm \sqrt{1 - \frac{2\beta^2}{\pi^2\ell^2}}\right] . \tag{16.12}$$

Note two things: first, there is a maximum β, ie minimum temperature,

$$\beta_{max} = \frac{\ell\pi}{\sqrt{2}} . \tag{16.13}$$

Second, for any given temperature β, there are *two* different black holes, corresponding to the sign choice in (16.12). Call these the 'small' (minus sign) and ''large' (plus sign) black holes. The turnover is at

$$r_* = \ell/\sqrt{2} \tag{16.14}$$

so each β allows a small black hole with $r_+ < r_*$ and a large black hole with $r_+ > r_*$.

We are working with thermal boundary conditions that fix the temperature β. So in field theory language, we are working in the canonical ensemble. Therefore we should sum over the allowed solutions; the thermodynamics will be determined by whichever has the lower free energy. We will see below that the large black hole always has lower free energy than the small black hole (but there is also a third solution, so the large black hole is not always dominate).

The behavior (16.12) is quite different from flat spacetime. In flat spacetime, larger black holes always have smaller temperature; this means ordinary Schwarzschild has negative specific heat and so is thermodynamically unstable. (It cannot be held in equilibrium with a bath, because it will absorb radiation from the bath and get *colder* the more radiation it absorbs!)

On the other hand in AdS, according to (16.12), large black holes have positive specific heat. If you make them bigger (higher energy), they get hotter. Small black holes have negative specific heat, and very small black holes $r_+ \ll \ell$ don't care about the cosmological constant at all so are just like the flat spacetime Schwarzschild solution.

On-shell action

The free energy $F = -\frac{1}{\beta}\log Z$ is computed using the on-shell Einstein action. We must be careful about the boundary terms. The full action we will use is

$$I_E = -\frac{1}{16\pi G_N}\int d^5x\sqrt{g}(R + \frac{12}{\ell^2}) + \frac{1}{8\pi G_N}\int_{r=1/\epsilon} d^4x\sqrt{\gamma}K + \int_{r=1/\epsilon} d^4x\sqrt{\gamma}L_{ct}[\gamma] \ . \tag{16.15}$$

The first term is the usual Einstein term. The second term is the Gibbons-Hawking boundary term. The last term is a counterterm; it can be any function of the intrinsic boundary geometry $\gamma_{\mu\nu}$ and will be picked to cancel divergences. Note that we are cutting off the spacetime at $r = 1/\epsilon$, we will take $\epsilon \to 0$ at the end.

Bulk term

The Einstein equation in empty spacetime implies $R = -20/\ell^2$. Thus the bulk term in (16.15) is

$$\begin{aligned} I_{bulk} &= \frac{1}{2\pi G_N \ell^2}\int_{r<1/\epsilon} d^5x\sqrt{g} && (16.16)\\ &= \frac{1}{2\pi G_N \ell^2}\int_0^\beta dt_E \int_{r_+}^{1/\epsilon} dr r^3 \int d\Omega_3 && (16.17)\\ &= \frac{\pi\beta}{4G_N\ell^2}\left[\frac{1}{\epsilon^4} - r_+^4\right] \ . && (16.18)\end{aligned}$$

Boundary terms

To compute the Gibbons-Hawking-York boundary term we just plug into the definition of extrinsic curvature, and eventually find

$$I_{GHY} = -\frac{\pi\beta}{G_N \ell^2}\left[\frac{1}{\epsilon^4} + \frac{3\ell^2}{4\epsilon^2} - \frac{\mu\ell^2}{2}\right] \ . \tag{16.19}$$

So far our answer $I_{bulk} + I_{GHY}$ is divergent as $\epsilon \to 0$. To fix this we need to add a boundary counterterm which is a functional of γ. It turns out the only choice that makes everything finite is

$$I_{ct} = \frac{3}{8\pi G_N \ell}\int_{r=1/\epsilon} d^4x\sqrt{\gamma}\left(1 + \frac{\ell^2}{12}R[\gamma]\right) \tag{16.20}$$

where $R[\gamma]$ is the Ricci scalar for the metric γ.

Total

Plugging into the counterterm action, evaluating it, and adding everything up we find (as $\epsilon \to 0$)

$$I_E = I_{bulk} + I_{GHY} + I_{ct} = \frac{\pi^2\beta}{8G_N\ell^2}\left[r_+^2\ell^2 - r_+^4 + \frac{3\ell^4}{4}\right] \ . \tag{16.21}$$

This is finite, by design. This should be viewed as a function of temperature $I_E = I_E(\beta)$, so we should plug in for $r_+(\beta)$ using (16.12). If we pick the plus sign we get the action of the large black hole, and if we pick the minus sign we get the action of the small black hole. You can easily check that

$$I_E(r_+^{small}(\beta)) \geq I_E(r_+^{large}(\beta)) \ . \tag{16.22}$$

Thus the dominant solution, with lower free energy, is the larger black hole, at any β.

Energy and entropy

Now that we have the partition function we can use all the usual thermodynamic relations to derive things like energy and entropy. The energy is

$$E = -\partial_\beta \log Z = \frac{3\pi^2}{8G_N}\left(\mu + \frac{\ell^2}{4}\right) \ . \tag{16.23}$$

The first term is the relation between mass and μ. The second term is a little surprising, since it is independent of the mass! It can be interpreted as a Casimir energy induced by putting the theory on S^3. Had we chosen boundary conditions $R^3 \times S^1_\beta$, this term would be zero.

The thermodynamic entropy is

$$S = (1 - \beta\partial_\beta) \log Z = \frac{\pi^2 r_+^3}{2G_N} \tag{16.24}$$

which you can check agrees with the area law $S = \text{area}/4G_N$.

16.1.2 Thermal AdS

We're not done: there is a third solution with thermal boundary conditions (16.3). It is called *Euclidean thermal AdS* and the metric is simply

$$ds^2 = \left(1 + \frac{r^2}{\ell^2}\right) dt_E^2 + \frac{dr^2}{1 + \frac{r^2}{\ell^2}} + r^2 d\Omega_3^2 \ , \tag{16.25}$$

with the identification

$$t_E \sim t_E + \beta \ . \tag{16.26}$$

This is just the metric of empty Euclidean AdS, except that we have identified the Euclidean time circle. Note that β is not related to any parameter in the metric itself, since there is no horizon r_+ in this metric. Therefore β is not fixed by any regularity condition, it is just a free parameter in this solution.

Lorentzian interpretation

We are discussing Euclidean manifolds, so let's pause to comment on the Lorentzian interpretation of all this. As discussed earlier in the course, path integrals of quantum fields on a Euclidean manifold with $t_E \sim t_E + \beta$ prepare the fields in a thermal state. For both of the Euclidean manifolds discussed here — the Euclidean black hole and thermal AdS — the path integral on the Euclidean manifold prepares a thermal state on the Lorentzian manifold. That is, the Euclidean path integral on Euclidean Schwarzschild-AdS prepares the Hartle-Hawking thermal state for fields on the the Lorentzian black hole. The Euclidean path integral on Euclidean thermal AdS prepares fields in a thermal state on ordinary Lorentzian AdS.

So in other words: In Lorentzian signature, thermal AdS is exactly the same classical solution as empty AdS, but the state of the perturbative fields is different — they are thermally populated, but their energy is small $O(\hbar)$ and does not backreact on the geometry itself.

Contractible vs non-contractible time circle

In the Euclidean black hole, t_E is the angle in polar coordinates. Together, the coordinates (r, t_E) with $r > r_+$ and $t_E \in (0, \beta)$ make a disk. The origin of the disk is a smooth point which corresponds to the Euclidean horizon.

In thermal AdS, t_E is a circle, but the circle does not contract anywhere. There is no origin. So in this geometry the coordinates (r, t_E) make a cylinder rather than a disk.

Action

The calculation of the on-shell action is similar to what we did for the black hole. Skipping the details, the answer in the end is just the Casimir term:

$$I_E^{(th)} = \frac{\pi^2\beta}{8G_N\ell^2}\left(\frac{3\ell^4}{4}\right) . \tag{16.27}$$

This is the free energy, which we can use to calculate the energy and entropy.

16.1.3 Hawking-Page phase transition

We found three Euclidean geometries obeying the thermal boundary condition (16.3). They are the small black hole, large black hole, and thermal AdS. The free energy of the large black hole is always smaller than that of the small black, so in understanding the phases, we can forget about the small black hole – it never dominates the canonical ensemble.

This leaves the large black hole and thermal AdS with actions

$$\begin{aligned} I_E^{(bh)}(\beta) &= \frac{\pi^2\beta}{8G_N\ell^2}\left[r_+^2\ell^2 - r_+^4 + \frac{3\ell^4}{4}\right] \qquad (16.28)\\ I_E^{(th)}(\beta) &= \frac{\pi^2\beta}{8G_N\ell^2}\left(\frac{3\ell^4}{4}\right) , \end{aligned}$$

where in $I_E^{(bh)}$ we choose the larger root for $r_+(\beta)$.

The semiclassical approximation to the gravitational path integral is the sum

$$Z_{grav}(\beta) \approx \exp(-I_E^{(bh)}) + \exp(-I_E^{(th)}) + \dots \ . \tag{16.29}$$

Each of the exponents is very large, since they are order $1/G_N$. Therefore the sum is exponentially dominated by whichever term is bigger:

$$\log Z_{grav}(\beta) \approx \max\left(-I_E^{(bh)}, \quad -I_E^{(th)}\right) \ . \tag{16.30}$$

There is a sharp (1st order) phase transition where the two solutions exchange dominance, ie at $I_E^{(th)} = I_E^{(bh)}$. Comparing the two actions, the critical temperature, and corresponding black hole radius, for this phase transition is

$$\beta_{crit} = \frac{2\pi\ell}{3} \ , \qquad r_+^{crit} = \ell \ . \tag{16.31}$$

The low-temperature phase is thermal AdS; the high temperature phase is the black hole. This phase transition is called the Hawking-Page transition and was discovered well before AdS/CFT. The story is qualitatively the same in any number of dimensions, AdS_{d+1} (with a few differences in AdS_3).

Entropy

The entropy of thermal AdS is zero. We can see this either by noting there is no horizon, or computing $(1 - \beta\partial_\beta)\log Z = 0$. Actually, it is not exactly zero, since we have only computed the semiclassical term. There are quantum corrections, and the true entropy is $O(G_N^0)$ from the one-loop contribution (*i.e.,* determinant of gravitons matter fields in AdS).

Thus full thermal entropy $S(\beta)$, accounting for the phase transition, is $O(G_N^0)$ at low temperatures and then suddenly jumps to a very large number $O(1/G_N)$ at β_{crit}. In the microcanonical ensemble, where we view this as a function of energy $S(E)$, the entropy is related to the density of states

$$S(E) = \log \rho(E) \ . \tag{16.32}$$

The Hawking-Page transition indicates that theories with a semiclassical gravity description must have a small number of states at low energy, but an enormous number of states at high energy, with a sharp transition.

16.1.4 Large volume limit

We have been computing the free energy at temperature β for the theory on the space S^3_ℓ, a 3-sphere of size ℓ. In fact since we are in conformal field theory, only the ratio ℓ/β is meaningful, as this is the only dimensionful parameter. In other words the only parameter is ℓT. Going to high temperatures is therefore the same as going to large ℓ.

If we are interested in the theory on R^3 we can take $\ell \to \infty$. This is the same as taking the temperature $T \to \infty$. In this limit, from (16.28), the free energy becomes (with $I_E = \beta F$)

$$F \approx -\left(\frac{\ell^2}{G_N}\right)\ell^3\pi^6 T^4 \ , \tag{16.33}$$

i.e.,

$$F \sim -\left(\frac{\ell}{\ell_P}\right)^3 V T^4 \tag{16.34}$$

where V is the volume of the system. Up to the prefactor, we could have guessed this answer from dimensional analysis. In a conformal field theory on R^{d-1} the only dimensionful scale is the temperature, and F must be proportional to volume, so conformal invariance implies $F \sim VT^d$.

Note that the theory on the plane has only one phase: the black hole phase. There is no Hawking-Page transition on R^3. It is essentially always in the high temperature phase.

16.2 Confinement in CFT

Any CFT with a semiclassical holographic dual must share the same thermodynamics, summarized by (16.30). What does this mean about the CFT? The microcanonical entropy tells us about the spectrum: we must have an enormous number of degrees of freedom to reproduce the high-energy density of states. However we must have a small number of states at low energies. This sounds like confinement! In a confining $SU(N)$ gauge theory, in the confining phase, the physical states are color singlet hadrons, and the free energy is $F = O(1)$. In a deconfined phase, the states are gluons, and the free energy is $F \sim O(N^2)$. This agrees with a our results above (after subtracting off the contribution from the Casimir energy, which is a temperature-independent contribution to F and does not affect the entropy). The black hole phase is like the deconfined phase, and the confined phase is like thermal AdS.

This analogy comes with some caveats, so let's compare and contrast QCD with a holographic theory like $\mathcal{N} = 4$ Super-Yang-Mills. In QCD, there is a confinement/deconfinement phase transition in infinite volume, ie for the theory on R^3. It is confining at low temperatures and deconfined at high temperatures. According to our gravity results, $\mathcal{N} = 4$ Super-Yang-Mills (at strong coupling) does *not* have a confining phase on R^3. CFT's on R^3 cannot have phase transitions, because the temperature can always be rescaled (unless there is some other parameter turned on, like a chemical potential). So in fact $\mathcal{N} = 4$ SYM is not a confining gauge theory in the same sense as QCD.

In gravity, the phase transition is on S^3. Normally it is not possible to have a phase transition in finite volume – with a finite number of degrees of freedom, the free energy is an analytic function of β, and we get sharp phase transitions only in the thermodynamic limit. However this is possible in gravity because of the large-N limit. The same statements are true of $\mathcal{N} = 4$ SYM on a sphere: it has something like a confinement/deconfinement transition on the sphere, in the $N \to \infty$ limit. It is not the same sort of confinement as QCD, which comes from dynamics in a very complicated way. In gauge theory on the sphere, we get 'kinematic confinement' just from the Gauss law constraint, which does not allow charges states on a compact space. Therefore the physical states on a sphere cannot have any net color. This is what 'confines' the theory so that there are $O(1)$ physical states at low energies.

Temporal Wilson loop

(This will be very brief; see Kirtsis for more discussion.)

Define the temporal Wilson loop $W = \text{Tr} \exp \oint A$ where the integral is over a worldline going around the thermal time circle. This is an order parameter for the deconfinement transition:

$$\langle W \rangle \neq 0 \quad \Rightarrow \quad \text{deconfinement} \quad \Rightarrow \quad \text{black hole phase} \tag{16.35}$$

$$\langle W \rangle = 0 \quad \Rightarrow \quad \text{confinement} \quad \Rightarrow \quad \text{thermal AdS phase} \tag{16.36}$$

($\langle W \rangle \neq 0$ actually breaks a symmetry: the center of the gauge group, $Z_N \in SU(N)$.) A rough explanation is that you can think of the temporal Wilson loop as a free quark. If a free quark has finite energy, then you get $\langle W \rangle \neq 0$, but if the free quark has infinite energy then $\langle W \rangle = 0$.

To compute a Wilson loop in AdS/CFT from the gravity side, the rule is to find a string worldsheet ending on the Wilson line and extending into the bulk. This classical string diagram computes the leading contribution to the Wilson loop at large N.

In the Euclidean black hole, since (t_E, r) make a disk, it is easy to find a string worldsheet ending on this Wilson line. In thermal AdS, however, since the thermal circle is not contractible — ie (t_E, r) make a cylinder — you cannot find such a string worldsheet, and the Wilson line vanishes.

Thus the deconfined phase is the phase with a contractible thermal circle in the dual geometry, and the confined phase has a non-contractible thermal circle.

16.3 Free energy at weak and strong coupling

So can we calculate the free energy of $\mathcal{N} = 4$ SYM, and compare to (16.30)? Unfortunately, no. The gravity calculation is dual to $\mathcal{N} = 4$ SYM at very strong 't Hooft coupling, $\lambda \equiv g_{YM}^2 N \to \infty$. The free energy is not protected by supersymmetry, and it is unknown how to do this calculate in gauge theory at strong coupling.

But we can do the CFT calculation at weak coupling. The free energy of a weakly coupled QFT is just a 1-loop calculation, ie a determinant for each of the fields. This calculation has been done. It agrees qualitatively, but not quantitatively, with the gravity calculation. For example, after translating all the parameters of CFT to gravity parameters, the free energy of free $\mathcal{N}=4$ SYM on R^3 is

$$F_{free} = \frac{4}{3} F_{gravity} \ , \tag{16.37}$$

where $F_{gravity}$ is given in (16.33). This famous factor of 4/3 is not a contradiction. It just means that the free energy at strong coupling is different from the free energy at weak coupling.

In principle, or perhaps on a lattice, the free energy is some function of the coupling,

$$F = -f(\lambda)\frac{\pi^2}{6}N^2 V T^4 \ . \tag{16.38}$$

We know the behavior of $f(\lambda)$ as $\lambda \to 0$ and as $\lambda \to \infty$. We only described the leading terms above, but it is also possible to calculate corrections. On the CFT side, corrections come from higher loops. On the gravity side, corrections come from including higher curvature (stringy) contributions to the classical action. (In principle we could also ask about $1/N$ corrections, which would require quantum calculations on the gravity side, but we'll restrict to the leading-N behavior.) These corrections have been calculated and lead to

$$\text{gravity:} \quad f(\lambda) = \frac{3}{4} + \frac{45}{32}\frac{\zeta(3)}{\lambda^{3/2}} + \cdots \quad \text{as} \quad \lambda \to \infty \tag{16.39}$$

and

$$\text{CFT:} \quad f(\lambda) = 1 - \frac{3}{2\pi^2}\lambda + \cdots \quad \text{as} \quad \lambda \to 0 \tag{16.40}$$

Evidently the corrections are heading the right direction, but the full function $f(\lambda)$ is unknown.

Exercise: Hawking-Page in Three Dimensions

Recall the metric of the Euclidean BTZ black hole in AdS_3,

$$ds^2 = \ell^2 \left[(r^2 - 8M)dt_E^2 + \frac{dr^2}{r^2 - 8M} + r^2 d\phi^2 \right] \ . \tag{16.41}$$

In a previous exercise, you computed the on-shell Euclidean action of this black hole. The answer, including all boundary terms and counterterms, is

$$S_E^{(bh)}(\beta) = -\frac{\pi^2 c}{3\beta} \tag{16.42}$$

where

$$c = \frac{3\ell}{2G_N} \ . \tag{16.43}$$

1. Like in AdS_5, there is also a thermal AdS solution with the same boundary condition.[*] We will use a trick to compute its action. The trick is to note that (16.41) is a solid torus with boundary $S^1_{2\pi\ell} \times S^1_\beta$. (The subscript is the circumference of the S^1). The thermal circle S^1_β is 'filled in' to make the solid torus.

   Thermal AdS_3 is a solid torus where instead the *other* circle $S^1_{2\pi\ell}$ is 'filled in' to make solid torus. Argue that this implies

   $$S_E^{(th)}(\beta) = S_E^{(bh)}(\frac{4\pi^2\ell^2}{\beta}) = -\frac{c\beta}{12} \ . \tag{16.44}$$

   Comment: This is a special case of a *modular transformation*. It is a 'large'[†] conformal transformation acting on a torus, which roughly speaking relates a fat torus to a skinny torus.

2. Sketch a plot of the free energies $F^{(bh)}$ and $F^{(th)}$. Find the critical temperature β_{crit} of the Hawking-Page phase transition, and write $\log Z(\beta)$ as a piecewise function.

3. Find the thermodynamic entropy $S(\beta)$ for all $\beta > 0$.

4. Find the energy $E(\beta)$ for all $\beta > 0$.

5. Use part (4) in part (3) to find the entropy in the microcanonical ensemble, $S(E)$. (Be careful about what ranges of E your formulas apply to; in particular you cannot find $S(E)$ for all E by this method.)

[*] Unlike AdS_5, there is only one black hole with temperature β.

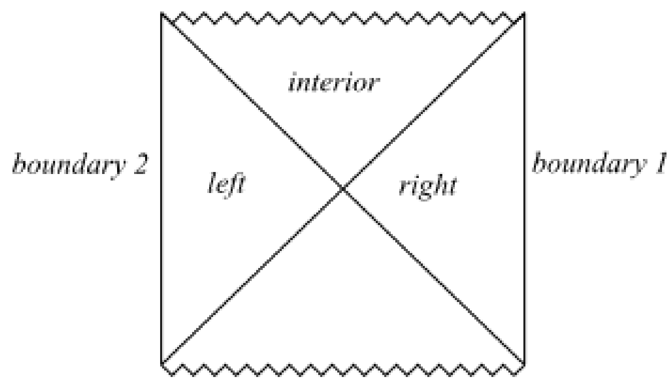
[†] *i.e.,* not continuously connected to the identity

6. Interpret your results in terms of the density of states in a 2d CFT dual to 3d gravity.

17 Eternal Black Holes and Entanglement

References: This section is based mostly on Maldacena hep-th/0106112; see also the relevant section of Harlow's review lectures, 1409.1231.

An *eternal* black hole is the black hole with the full, two-sided Penrose diagram. It has a past singularity, a future singularity, and two asymptotic regions:



(17.1)

This is to be distinguished from a black hole that forms from gravitational collapse, which has no past singularity and no second asymptotic region on the 'left' of the Penrose diagram. Although we often use the maximally extended Penrose diagram to discuss all sorts of black holes, it is only in the eternal black hole that we should really take the left side of the Penrose diagram seriously.

An eternal black hole in AdS — the maximally extended AdS-Schwarzschild spacetime — has two boundaries. This means that it is dual to two copies of the CFT. In fact, the connection between thermal field theory and this 'doubling' of degrees of freedom was well known long ago, and is called the thermofield double formalism. First we will describe this formalism in QFT, then we'll make the connection to AdS black holes.

17.1 Thermofield double formalism

Consider any QFT, with Hamiltonian H and complete set of eigenstate $|n\rangle$,

$$H|n\rangle = E_n|n\rangle \ . \tag{17.2}$$

The thermofield double formalism is a trick to treat the thermal, mixed state $\rho = e^{-\beta H}$ as a pure state in a bigger system. First we double the degrees of freedom, *i.e.,* we consider a new QFT which is two copies of the original QFT. If the theory is defined by a Lagrangian, then for every field ϕ in the original QFT, there are two fields $\phi_1(x_1)$ and $\phi_2(x_2)$ in the doubled QFT. These two fields live in different spacetimes x_1 and x_2, and are not coupled in the Lagrangian at all. The states of the doubled QFT are

$$|m\rangle_1 |n\rangle_2 \ . \tag{17.3}$$

Now in this doubled system we consider the *thermofield double state*:

$$|TFD\rangle = \frac{1}{\sqrt{Z(\beta)}} \sum_n e^{-\beta E_n/2} |n\rangle_1 |n\rangle_2 \ . \tag{17.4}$$

This is a particular pure state in the doubled system. The density matrix of the doubled QFT in this state is

$$\rho_{total} = |TFD\rangle\langle TFD| \ . \tag{17.5}$$

The reduced density matrix of system 1 is

$$\begin{aligned} \rho_1 &= \mathrm{tr}_2 \, \rho_{total} \\ &= \sum_m {}_2\langle m| \left(\sum_{n,n'} e^{-\beta E_n/2} |n\rangle_1 |n\rangle_2 \, {}_2\langle n'| \, {}_2\langle n'| e^{-\beta E_{n'}/2} \right) |m\rangle_2 \\ &= \sum_n e^{-\beta E_n} |n\rangle_1 \, {}_1\langle n| \\ &= e^{-\beta H_1} \end{aligned} \tag{17.6}$$

Therefore, *if we restrict our attention to system 1, this pure state in the doubled system is indistinguishable from a thermal state.* For example, if O_1 is made of local operators acting on system 1, $O_1 = \phi_1(x_1)\chi_1(y_1)\cdots$, then

$$\langle TFD|O_1|TFD\rangle = \frac{1}{Z(\beta)} \, \mathrm{Tr}_{\,\mathcal{H}_1} \, e^{-\beta H_1} O_1 \ . \tag{17.7}$$

This procedure is called *purifying* the thermal state. In fact, any mixed state can be purified by adding enough auxiliary states and tracing them out.

Although systems 1 and 2 are not coupled in the Lagrangian of the doubled system, they are correlated because we are in this particular entangled state. For example, if O_1 is built from operators acting on system 1 and O_2 is built from operators acting on system 2, then

$$\langle TFD|O_1O_2|TFD\rangle \tag{17.8}$$

can be non-zero.

The Hamiltonian

The choice of Hamiltonian acting on the doubled system is up to us. Two convenient choices are

$$H_{tot} = H_1 - H_2 \qquad \text{and} \qquad \tilde{H}_{tot} = H_1 + H_2 \ . \tag{17.9}$$

For our purposes, we will just use H_{tot}, but $\tilde{H}_{tot}$ is also useful in other contexts. Under H_{tot}, the TFD state is time-independent, since the phases cancel:

$$|TFD(t)\rangle \equiv e^{-iH_{tot}}|TFD\rangle = \sum_n e^{-\beta E_n/2} e^{-i(H_1-H_2)}|n\rangle_1\, |n\rangle_2 = |TFD\rangle \ . \tag{17.10}$$

17.2 Holographic dual of the eternal black hole

The statement

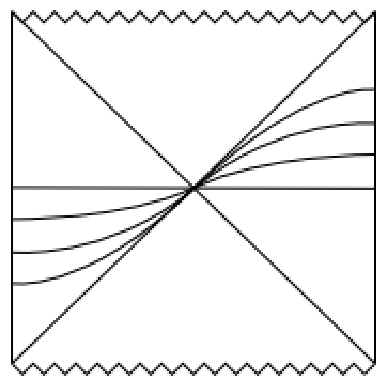
Maldacena's proposal is that the eternal black hole depicted in (17.1) is dual to two copies of the CFT, in the thermofield double state $|TFD\rangle$. Each asymptotic boundary of AdS is a copy of the original dual CFT. So, for example, to compute correlation functions like

$$\langle TFD|\phi_1(x_1)\chi_2(x_2)|TFD\rangle \tag{17.11}$$

we would use Witten diagrams with χ inserted on the left boundary, and ϕ inserted on the right boundary. Note that the local bulk fields are not doubled: there is just one bulk field Φ dual to the boundary operators ϕ_1 and ϕ_2, but this makes sense because we have to specify double the boundary conditions for Φ. The boundary condition for Φ on the left acts like a source for ϕ_2, and the boundary condition for Φ on the right acts like a source for ϕ_1.

The Hamiltonian

The Hamiltonian H_{tot} in (17.9) has a natural bulk interpretation. It is dual to the bulk Hamiltonian that generates time evolution along the isometry ∂_t, where t is the usual Schwarzschild coordinate. Recall (or look back at a textbook on the Kruskal coordinate change) that the Schwarzschild t coordinate runs 'backwards' on the left side of the Penrose diagram. That is, all of the spatial slices drawn in this figure are equivalent under the ∂_t isometry:



(17.12)

This corresponds to the minus sign in $H_{tot} = H_1 - H_2$.

Derivation

To justify the claim that the eternal black hole is dual to the TFD state, we will apply the AdS/CFT dictionary (14.2), in the form

$$Z_{gravity}[\partial M = \Sigma] = Z_{cft}[\Sigma] \ . \tag{17.13}$$

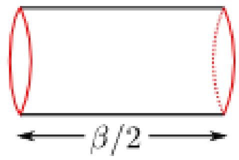
(Here M is the bulk manifold, and the meaning of the lhs is the gravity path integral with boundary condition $\partial M = \Sigma$.)

First, the CFT: The Euclidean path integral that prepares the TFD state is a path integral on an interval of length $\beta/2$, times a circle:

$$\Sigma = \text{Interval}_{\beta/2} \times S^{d-1} \ . \tag{17.14}$$

Pictorially,

$$|TFD\rangle = \tag{17.15}$$



This path integral has two open cuts (red), at the ends of the interval. We interpret the left cut as defining a state in system 2, and the right cut as defining a state in system 1. That is, this picture should be interpreted as a rule for computing the transition amplitude with field data φ_1 and φ_2 specified at the ends of the interval. To confirm that this path integral really prepares the TFD state, all we need to do is check that it computes the correct transition amplitudes. The path integral with these boundary conditions is*

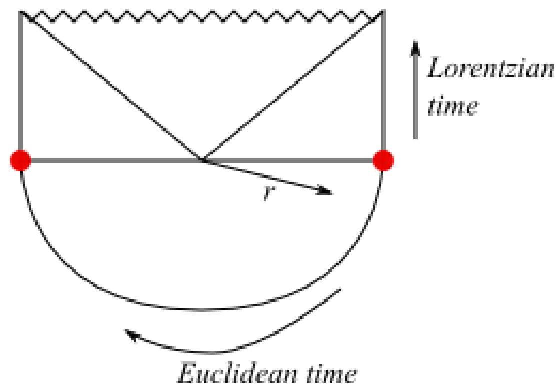
$$\begin{aligned}
{}_1\langle\varphi_1|\,{}_2\langle\varphi_2|TFD\rangle &= \langle\varphi_1|e^{-\beta H/2}|\varphi_2^*\rangle && (17.16)\\
&= \sum_n \varphi_1|n\rangle\langle n|\tilde{\varphi}_2\rangle e^{-\beta E_n/2} && (17.17)\\
&= \sum_n e^{-\beta E_n/2}\langle\varphi_1|n\rangle_1\langle\varphi_2|n\rangle_2 && (17.18)
\end{aligned}$$

These are precisely the matrix elements of the state $|TFD\rangle$ defined in (17.4). So, as claimed, this is the Euclidean path integral that prepares $|TFD\rangle$.

Now that we've produced this state from a Euclidean path integral on the manifold Σ, we can apply (17.13). We must find a Euclidean gravity solution with conformal boundary condition $\partial M = \text{Interval}_{\beta/2} \times S^{d-1}$. In fact, *half* of the Euclidean black hole has precisely this boundary condition. That is, we consider the Euclidean Scharzschild-AdS solution and restrict to $t_E \in [0, \beta/2]$ instead of the full range $t_E \in [0, \beta]$. The (t_E, r) portion of this Eulidean spacetime makes a half-disk; the boundary of the half-disk is $\text{Interval}_{\beta/2} \times S^{d-1}$. The half-disk is cut down the middle; this cut is interpreted as the time=0 surface of the Lorentzian spacetime. Pictorially, the bulk spacetime has a Euclidean piece that prepares the state, then a Lorentzian piece describing the time

*The tildes indicate the conjugate state.

evolution in Minkowski signature:



(17.19)

The red blobs in this picture denote the S^{d-1}'s at the end of the interval on the boundary, the red circles in (17.15).

17.3 ER=EPR

Let's describe this result in words. The left side of the Penrose diagram is dual to CFT_2, and the right side is dual to CFT_1. The Einstein-Rosen bridge connecting to the two sides, the black hole interior itself, are somehow 'created' by entangling CFT_1 with CFT_2. In fact this is a precise statement: In CFT language, correlators between the two CFTs like $\langle TFD|O_1 O_2|TFD\rangle$ are non-zero only because we are in the entangled state $|TFD\rangle$. After all, the two CFTs are not coupled. In the bulk, these correlations are nonzero because we can draw Witten diagrams going through the interior. For very massive fields or high energies, the left-right correlation functions can be approximated by geodesics that pass through the black hole interior. Without a wormhole connecting the two sides, there would be no such correlations.

This idea and its generalizations have recently been given the slogan 'ER=EPR' (by Maldacena and Susskind): Einstein-Rosen bridges are equivalent to entanglement (as discussed by Einstein-Podolsky-Rosen). This slogan is only entirely precise and well defined in the semiclassical limit, describing the eternal black hole and similar spacetimes, but the idea is that some more general construction should make sense in the very quantum, non-geometrical limit.

17.4 Comments in information loss in AdS/CFT

Hawking's information loss paradox relied on a black hole that forms from collapse, then evaporates. In AdS, this only happens for small black holes. These black holes are not in thermal equilibrium, and are difficult to address precisely using AdS/CFT. Of course, the CFT is always unitary, so if we believe AdS/CFT (or use AdS/CFT to define a theory of quantum gravity) then obviously this evaporation process, however it is described in CFT, must be unitary. This strongly suggests that unitarity should be preserved, and locality or some other tenet of effective field theory must be violated. However it is not very satisfying, since it does not answer the question of what went wrong with Hawking's calculation. Presumably the answer is that local effective field theory is not quite right in non-perturbative quantum gravity, but we do not really understand how to characterize this breakdown. This is a very important open question in current research.

17.5 Maldacena's information paradox

Maldacena introduced a different version of the information paradox that applies to large, eternal black holes. This version is easier to address in AdS/CFT. The idea is to first perturb the thermal state by inserting an operator O_2 in CFT_2,

$$|TFD\rangle \to |\widetilde{TFD}\rangle = (1 + \epsilon O_2)|TFD\rangle \ . \tag{17.20}$$

This changes the reduced density matrix of system 1,

$$\rho_1 \to \tilde{\rho}_1 = e^{-\beta H_1} + \text{tiny corrections} \ . \tag{17.21}$$

Now, we compute expectation values in CFT_1,

$$\langle\widetilde{TFD}|O_1|\widetilde{TFD}\rangle \tag{17.22}$$

in the perturbed state. To first order in the perturbation, this is the two-sided correlation function

$$\langle O_1\rangle \sim G_{12} \equiv \langle TFD|O_1 O_2|TFD\rangle \ . \tag{17.23}$$

Now we can produce a contradiction by waiting a very long time, so this correlation function decays. On the gravity side, if we hold O_2 at a particular time and send O_1 to very late times, then the geodesic distance between these two points grows linearly with time, forever. Therefore the correlation function must decay as

$$G_{12}^{gravity} \sim e^{-\mathrm{const}\times t/\beta} \tag{17.24}$$

for $t \gg \beta$. This decays exponentially to zero. At very late times, it therefore becomes exactly thermal, with arbitrarily small corrections.

This contradicts unitarity of the CFT. In the CFT, any perturbation of the thermal state should stay forever a perturbation of the thermal state: it will of course become scrambled and appear to thermalize, but it should never forget the initial perturbation completely, so it should never become arbitrarily close to the thermal state. In fact the corrections to the thermal state should be suppressed by the entropy, but finite:

$$G_{12}^{CFT} \sim e^{-\mathrm{const}\times S} \tag{17.25}$$

for $t \gg \beta$. In summary, at very late times, gravity 'forget' the initial perturbation, but a unitary CFT does not:

$$G_{12}^{gravity} \ll G_{12}^{unitary} \ . \tag{17.26}$$

However is this paradox resolved? The answer is that we have neglected non-perturbative contributions of the gravity side of order $e^{-1/G_N} \sim e^{-S}$. For example, there is another saddlepoint (the thermal AdS saddle) and fluctuations around this saddle will also contribution to the two-sided correlation function at this order.

Although this tells us where the gravity derivation went wrong, it does not tell us exactly how to recover the lost information in quantum gravity, *i.e.,* without referring to the dual CFT. Presumably this would require treating the full non-perturbative string theory, which is currently not possible.

17.6 Entropy in the thermofield double

Our next topic will be entanglement. It will be a while before we get back to gravity, so as a brief preview, let's consider the interpretation of entropy in the thermofield double formulation of the black hole. The state $|TFD\rangle$ is a pure state; pure states have no ordinary entropy, *i.e.*,

$$\rho_{total} \equiv |TFD\rangle\langle TFD| \tag{17.27}$$

has entropy

$$S_{tot} = -\operatorname{tr} \rho_{total} \log \rho_{total} = 0 \ . \tag{17.28}$$

However, if we trace out half the system, we know this gives a thermal state, so it should have some entropy. The reduced density matrix of system 1 is

$$\rho_1 = \operatorname{tr}_2 \rho_{total} = e^{-\beta H_1} \tag{17.29}$$

Therefore the state of system 1 has entropy,

$$S_1 = -\operatorname{tr} \rho_1 \log \rho_1 = S_{thermal}(\beta) \neq 0 \ . \tag{17.30}$$

We started in a pure state, with no entropy. Where did the entropy of system 1 come from? The answer is entanglement. Thermal entropy is just one special case of *entanglement entropy*.

18 Introduction to Entanglement Entropy

The next few lectures are on entanglement entropy in quantum mechanics, in quantum field theory, and finally in quantum gravity. Here's a brief preview: Entanglement entropy is a measure of how quantum information is stored in a quantum state. With some care, it can be defined in quantum field theory, and although it is difficult to calculate, it can be used to gain insight into fundamental questions like the nature of the renormalization group. In holographic systems, entanglement entropy is encoded in geometric features of the bulk geometry.

We will start at the beginning with discrete quantum systems and work our way up to quantum gravity.

References: Harlow's lectures on quantum information in quantum gravity, available on the arxiv, may be useful. See also Nielsen and Chuang's introductory book on quantum information for derivations of various statements about matrices, traces, positivity, etc.

18.1 Definition and Basics

A *bipartite system* is a system with Hilbert space equal to the direct product of two factors,

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B \ . \tag{18.1}$$

Starting with a general (pure or mixed) state of the full system ρ, the reduced density matrix of a subsystem is defined by the partial trace,

$$\rho_A = \mathrm{tr}_B\, \rho \tag{18.2}$$

and the *entanglement entropy* is the von Neumann entropy of the reduced density matrix,

$$S_A \equiv -\,\mathrm{tr}\, \rho_A \log \rho_A \ . \tag{18.3}$$

Example: 2 qubit system

If each subsystem A or B is a single qubit, then the Hilbert space of the full system is spanned by

$$|00\rangle, \quad |01\rangle, \quad |10\rangle, \quad |11\rangle \ , \tag{18.4}$$

where the first bit refers to A and the second bit to B, *i.e.,* we use the shorthand

$$|ij\rangle \equiv |i\rangle_A |j\rangle_B \equiv |i\rangle_A \otimes |j\rangle_B \ . \tag{18.5}$$

Suppose the system is in the pure state

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|00\rangle + |11\rangle \right) \ , \tag{18.6}$$

so $\rho = |\psi\rangle\langle\psi|$. As a 4x4 matrix, ρ has diagonal and off-diagonal elements. Diagonal density matrices are just classical probability distributions, but the off-diagonal elements indicate entanglement and are intrinsically quantum.

The reduced density matrix of system A is

$$\begin{aligned} \rho_A &= \mathrm{tr}_B\, \rho \\ &= \frac{1}{2}\, {}_B\langle 0| \left(|00\rangle + |11\rangle \right) \left(\langle 11| + \langle 00| \right) |0\rangle_B \\ &\quad + \frac{1}{2}\, {}_B\langle 1| \left(|00\rangle + |11\rangle \right) \left(\langle 11| + \langle 00| \right) |1\rangle_B \\ &= \frac{1}{2} \big(|0\rangle_A\, {}_A\langle 0| + |1\rangle_A\, {}_A\langle 1| \big) \\ &\propto \mathbf{1_{2x2}} \ . \end{aligned} \tag{18.7}$$

The last line says ρ_A is proportional to the identity matrix of a 2-state system. In this case we say ρ_A is *maximally mixed*, and the initial state $|\psi\rangle$ is *maximally entangled*.

The entanglement entropy of subsystem A is easy to calculate for a diagonal matrix,

$$\begin{aligned} S_A &= -\,\mathrm{tr}\, \rho_A \log \rho_A \\ &= -2 \times \frac{1}{4} \log \frac{1}{4} \\ &= \log 2 \ . \end{aligned} \tag{18.8}$$

Interpretation of entanglement entropy

In fact the 2-qubit example illustrates a useful way to put entanglement entropy into words:

Entanglement entropy counts the number of entangled bits between A and B.

If we had k qubits in system A and k qubits in system B, then in a maximally entangled state $S_A = k \log 2$. So S_A counts the number of bits, or equivalently, e^{S_A} counts the number of entangled states (since k qubits have 2^k states).

Rephrased slightly:

Given a state ρ_A with entanglement entropy S_A, the quantity e^{S_A} is the minimal number of auxiliary states that we would need to entangle with A in order to obtain ρ_A from a pure state of the enlarged system.

Schmidt decomposition

A very useful tool is the following theorem, called the Schmidt decomposition: Suppose we have a system AB in a pure state $|\psi\rangle$. Then there exist orthonormal states $|i\rangle_A$ of A and $|i\rangle_B$ of B such that

$$|\psi\rangle = \sum_i \lambda_i |i\rangle_A |\tilde{i}\rangle_B \ , \tag{18.9}$$

with λ_i real numbers in the range $[0, 1]$ satisfying

$$\sum_i \lambda_i^2 = 1 \ . \tag{18.10}$$

The number of terms in the sum is (at most) the dimension of the smaller Hilbert space $\mathcal{H}_A$ or $\mathcal{H}_B$.

Proof: See Wikipedia, or Nielsen and Chuang chapter 2.

If A is small and B is big, this is intuitive. It says we can pick a basis for $|i\rangle_A$, and each of these states will be correlated with a particular state of system B. The thermofield double is an obvious example.

Complement subsystems

An immediate consequence of the Schmidt decomposition is that a pure state of system

AB has

$$S_A = S_B \qquad \text{(pure states)} \ . \tag{18.11}$$

To see this, write the reduced density matrices in the Schmidt basis,

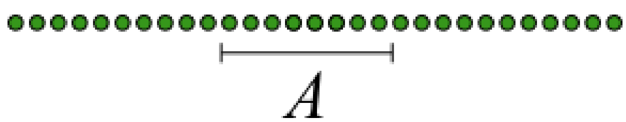
$$\rho_A = \sum_i \lambda_i^2 |i\rangle_{A\,A}\langle i| \ , \qquad \rho_B = \sum_i \lambda_i^2 |i\rangle_{B\,B}\langle i| \ . \tag{18.12}$$

Both density matrices have eigenvalues λ_i^2 so they have the same entropy. (18.11) does not hold for mixed states of AB.

18.2 Geometric entanglement entropy

Entanglement entropy can be defined whenever the Hilbert space splits into two factors. A very important example is when we define A as a subregion of space.

Example: N spins on a lattice in 1+1 dimensions
Let's arrange N spins in a line. Define A to be a spatial region containing k spins, and $B = A^C$ is everything else:



The most general state of this system is

$$|\psi\rangle = \sum_{\{s_i\}} c_{s_1\cdots s_N} |s_1\rangle|s_2\rangle\cdots|s_N\rangle \tag{18.13}$$

where $s_i = 0$ or 1 (meaning 'up' or 'down'), and the c's are complex numbers.

Scaling with system size
Let's restrict to $1 \ll |A| \ll |B|$, so that we can think of subsystem A as large and B as infinite. In a random state, *i.e.,* one in which the coefficients $c_{ij\cdots}$ are drawn from a uniform distribution, we expect any subsystem A to be almost maximally entangled with B. In the language of the Schmidt decomposition, this means that λ_i is nonzero

and $\sim 1/\sqrt{2^k}$ for a complete basis of states $|i\rangle_A$. In fact this is a theorem, see Harlow's lectures for the exact statement.

Accordingly, the entanglement entropy scales as the number of spins in region A. In 1+1d this is linear in the size of A, and more generally,

$$S_A \sim \text{Volume(A)} \qquad \text{(random state)}. \tag{18.14}$$

In other words, most states in the Hilbert space of the full system have entanglement scaling with volume.

However, often we are interested in the groundstate. Ground states of a local Hamiltonian are very non-generic, and the corresponding entanglement entropies obey special scaling laws. Usually, if the system is gapped (*i.e.,* correlations die off exponentially), the ground state must obey the *area law*:

$$S_A \sim \text{Area(A)} \qquad \text{(ground state of local, gapped Hamiltonian)} \ . \tag{18.15}$$

(This is a theorem in 1+1d, and usually true in higher dimensions.)

Thus groundstates occupy a tiny, special corner of the Hilbert space. This is a corner with especially low 'complexity.' Intuitively speaking, a large degree of entnaglement is what makes quantum information exponentially more powerful than classical information; so states with lower entanglement entropy are less complex. More specifically, this actually means that you can encode a groundstate wavefunction with far fewer parameters than the 2^N complex numbers appearing in (18.13).

DMRG

In 1+1d, the area law becomes simply

$$S_A \sim \text{const} \tag{18.16}$$

independent of the system size. This special feature is responsible for a hugely important technique in quantum condensed matter called the *density matrix renormalization group* (DMRG). This technique is used to efficiently compute groundstate wavefunctions of 1+1d systems using a computer. This would not be possible for general states, since (we think) classical computers require exponential time to simulate quantum systems. But (18.16) means that, in a precise sense, groundstates of gapped 1d systems are no more complex than classical systems.

Scaling at a critical point

The area law applies to gapped systems. Near a critical point, where dof become massless and long-distance correlations are power-law instead of exponentially suppressed, the area law can be violated. In a 1+1d critical system, and therefore also in 1+1d conformal field theory, (18.16) is replaced by

$$S_A \sim \log L_A \tag{18.17}$$

where L_A is the size of region A. This is bigger than the area law, but still much lower than the volume-scaling of a random state.

18.3 Entropy Inequalities

Relative entropy

Much of the recent progress in QFT based on entanglement comes from a few inequalities obeyed by entanglement entropy. Define the *relative entropy*

$$S(\rho||\sigma) \equiv \operatorname{tr} \rho \log \rho - \operatorname{tr} \rho \log \sigma \ . \tag{18.18}$$

(Note that this is not symmetric in ρ, σ.) This obeys

$$S(\rho||\sigma) \geq 0 \tag{18.19}$$

with equality if and only if $\rho = \sigma$. The proof of this statement is straightforward, see Wikipedia. It just involves some matrix manipulations. The key ingredient is the fact that density matrices in quantum mechanics are very special: they have a positive spectral decomposition,

$$\rho = \sum_i p_i v_i v_i^* \tag{18.20}$$

where p_i is non-negative and v_i is a basis vector. This is necessary for quantum mechanics to have a sensible probabilistic interpretation and is closely related to unitarity.

The relative entropy can be viewed as a measure of how 'distinguishable' ρ and σ are. In the classical case (diagonal ρ and σ), it is error we will make in predicting the uncertainty of a random process if we think the probability distribution is σ, but actually it is ρ. Given this interpretation, positivity is obvious — clearly we will never do *better* using the wrong distribution.

Triangle inequality

Positivity or relative entropy implies the *triangle inequality*,

$$|S_A - S_B| \le S_{AB} \ . \tag{18.21}$$

Mutual information

Define the *mutual information*,

$$I(A,B) \equiv S_A + S_B - S_{AB} \ . \tag{18.22}$$

This can be written as a relative entropy, and is therefore non-negative:

$$I(A,B) = S(\rho_{AB}||\rho_A \otimes \rho_B) \ge 0 \ . \tag{18.23}$$

Roughly, $I(A,B)$ measures the amount of information that A has about B (or vice-versa, since it is symmetric).

In a pure state of AB, the only correlations between A and B come from entanglement, so in this case $I(A,B)$ measures entanglement between A and B. However, in a mixed state, $I(A,B)$ also gets classical contributions. For example in a 2-qubit system, it is easy to check that the classical mixed state

$$\rho_{AB} \propto |00\rangle\langle 00| + |11\rangle\langle 11| \tag{18.24}$$

has non-zero mutual information.

Strong subadditivity

So far we have discussed partitioning a system into two pieces A and B, but we can partition further and find new inequalities. The *strong subadditivity* inequality (*SSA* for short) applied to a tripartite system $\mathcal{H}_{ABC} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, is

$$S_{ABC} + S_B \leq S_{AB} + S_{BC} \ . \tag{18.25}$$

This is less mysterious if written in terms of the mutual information,

$$I(A,B) \leq I(A,BC) \ . \tag{18.26}$$

Although this inequality seems obvious — clearly A has more information about BC than about B alone — and is 'just' a feature of positive matrices, it is surprisingly difficult to prove. See Nielsen and Chuang for a totally unenlightening derivation.

Sometimes it is useful to express (18.25) in different notation, where A and B are two overlapping subsystems, which are not independent:

$$S_{A\cup B} + S_{A\cap B} \leq S_A + S_B \ . \tag{18.27}$$

Exercise: Positivity of classical relative entropy

Prove that the classical relative entropy is non-negative. That is, prove (18.19), assuming ρ and σ are diagonal.

Exercise: Mutual information practice

Consider a 2-qubit system. First, calculate the mutual information of the two bits in the classical mixed state

$$\rho = \tfrac{1}{2}\left(|00\rangle\langle 00| + |11\rangle\langle\, 11|\right) \ . \tag{18.28}$$

This is a clearly a state with the maximal amount of classical correlation — if we measure one bit, we know the value of the second bit.

Now, what is the maximal amount of mutual information for a *quantum* (pure or mixed) state of 2 qubits? Write an example of a state with this maximal amount of mutual information. (Quantum states with more mutual information than is possible in any classical state are sometimes called *supercorrelated*.)

Exercise: Purification and the Triangle Inequality

Use strong subadditivity to prove the following identities for a tripartite system:

$$S_A \leq S_{AB} + S_{BC} \tag{18.29}$$

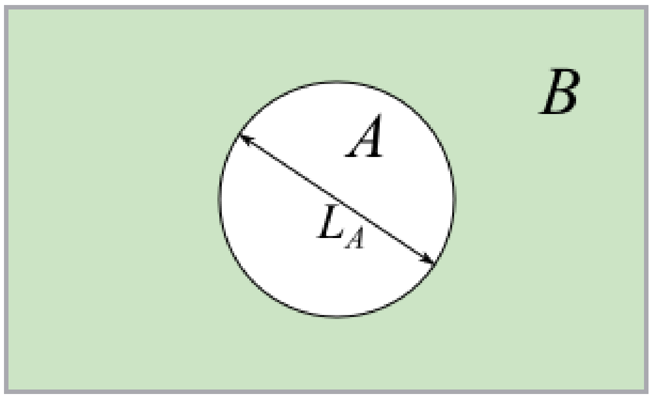
$$S_A \leq S_{AB} + S_{AC} \tag{18.30}$$

$$S_{AB} \geq |S_A - S_B| \tag{18.31}$$

Hint: Purify the tripartite system that appears in strong subadditivity by adding a 4th system, D, with $ABCD$ in a pure state. This is always possible.

19 Entanglement Entropy in Quantum Field Theory

So far we have discussed entanglement in ordinary quantum mechanics, where the Hilbert space of a finite region is finite dimensional. Now we will discuss geometric entanglement entropy in quantum field theory. Space (not spacetime) is divided into two regions, A and B, by a continuous curve:



(19.1)

This picture is at a fixed time. Region A is drawn as a circle, but for now it could be any shape. (It could also be disjoint, but we will assume it is connected unless specified otherwise.)

Quantum field theory is strictly speaking *not* bipartite,

$$\mathcal{H}_{AB} \neq \mathcal{H}_A \otimes \mathcal{H}_B \ . \tag{19.2}$$

There are two things to worry about: first, in gauge theories, you cannot really localize states. The gauge constraint is applied to the full system, so by looking at any subregion, you cannot decide whether it is a physical state obeying the constraint. This issue (which also appears in ordinary quantum mechanical gauge systems) has been addressed in some nice papers just in the last year or so, and we will ignore it entirely. It turns out to not affect the discussion that follows very much.

The second issue is UV divergences. In a continuum QFT there are UV modes at arbitrarily small scales across the dividing surface ∂A, and this makes it impossible to actually split the full Hilbert space. To deal with this, we must impose a UV cutoff by introducing the 'lattice scale' ϵ_{UV}. With a finite cutoff, the Hilbert space of a finite region is finite-dimensional, and most of the results of the previous section — in particular, positivity of relative entropy and strong subadditivity — apply to QFT. In the end we usually want to regulate the divergences somehow, but the leftover finite pieces do not immediately obey the same properties, so we need to be careful about tracking cutoff dependence throughout the problem.

19.1 Structure of the Entanglement Entropy

The divergent terms in S_A come from UV physics. In the UV, any finite energy state is the same as the vacuum state. Therefore to discuss the structure of the divergent terms we can restrict to $\rho = |0\rangle\langle 0|$, the vacuum state of the full system.

UV divergences

The divergent terms depend on the theory and on the shape of region A. In a local QFT, we expect the divergent piece to be a *local* integral over the entangling surface ∂A,

$$S_A^{(div)} \sim \int_{\partial A} d^{d-2}\sigma \sqrt{h}\, F[K_{ab}, h_{hab}] \ , \tag{19.3}$$

where F is some (theory-dependent) functional of the extrinsic curvature and induced metric on ∂A. This is for the same reason that when we do renormalization, we are only allowed to add local counterterms to the Lagrangian; non-local terms come from IR physics.

Let's organize (19.3) as an expansion in powers of K_{ab}. Since $K_{ab} \sim 1/L_A$, this is an expansion in powers of the size L_A. What sort of terms can appear? In a pure state, $S_A = S_B$, an in particular $S_A^{(div)} = S_B^{(div)}$. The extrinsic curvature is $K \sim \nabla n$ with n the unit normal; this flips sign if we consider region A vs its complement, region B. Therefore $S_A^{(div)} = S_B^{(div)}$ implies that only even powers of K_{ab} are allowed:

$$S_A^{(div)} \sim a_1 L_A^{d-2} + a_2 L_A^{d-4} + \cdots \ , \tag{19.4}$$

where a_i depend on the theory but not on L_A.

The leading term in (19.4) is a UV divergence proportional to Area(A). This makes sense: UV modes entangled across ∂A give a divergent contribution, and the number of these modes is proportional to the area.

General structure and universal terms

Now let us further assume the theory is scale invariant. In the vacuum state of a scale invariant theory, the only scales in the problem are ϵ_{UV} and L_A. Therefore, by dimensional analysis, $a_1 \sim \epsilon_{UV}^{2-d}$, $a_2 \sim \epsilon_{UV}^{4-d}$, etc. Thus, allowing also for a finite contribution, we find the general behavior of the entanglement entropy in a CFT. In odd dimensions d:

$$S_A^{CFT} \sim b_{d-2}\left(\frac{L_A}{\epsilon_{UV}}\right)^{d-2} + b_{d-4}\left(\frac{L_A}{\epsilon_{UV}}\right)^{d-4} + \cdots + b_1\frac{L_A}{\epsilon_{UV}} + (-1)^{\frac{d-1}{2}}\tilde{S} + O(\epsilon_{UV}) \ , \quad (19.5)$$

and in even dimensions:

$$\begin{aligned} S_A^{CFT} \ \sim \ & b_{d-2}\left(\frac{L_A}{\epsilon_{UV}}\right)^{d-2} + b_{d-4}\left(\frac{L_A}{\epsilon_{UV}}\right)^{d-4} + \cdots + b_2\left(\frac{L_A}{\epsilon_{UV}}\right)^{2} \\ & + (-1)^{\frac{d-2}{2}}\tilde{S}\log\frac{L_A}{\epsilon_{UV}} + \text{const} + O(\epsilon_{UV}) \ , \end{aligned} \quad (19.6)$$

The difference between even and odd comes from the fact that the "$1/\epsilon_{UV}^0$" term that would appear in even dimensions actually turns into a log divergence (just as it does in Feynman diagrams). The powers of (-1) are inserted by convention.

In the vacuum state, the b_i and $\tilde{S}$ depend on the theory, but not on L_A or ϵ_{UV}.

In a non-scale-invariant QFT, or in an excited state of a CFT, there are other scales.* So in general, $\tilde{S}$ depends on the theory, the shape, and the state ρ_{total}. Furthermore, $\tilde{S}$ is *universal* in the sense that it does not depend on ambiguities in the choice of regulator. For this reason it is sometimes called the *renormalized entanglement entropy.* †

*This is sometimes confusing in the CFT case but obviously true: even in a scale invariant theory, an excited state with lumps of stuff 1 meter apart is different from an excited state with lumps of stuff 2 meters apart!

†The fact that it is independent of regulator is clear in the even dimensional case, since it is the coefficient of a log. It is less clear for the constant term odd dimensions, since evidently shifting $\epsilon_{UV} \to \epsilon_{UV} + a$ would change the finite term. In practice it seems to be well defined in CFT for reasons I won't get into here, but I'm not sure about the non-conformal case.

Area vs volume terms

The leading UV divergence is always proportional to Area(A), in any state. In the vacuum we do not expect any extensive contribution to $\tilde{S}$, but in a random excited state, we expect

$$\tilde{S} \sim \text{Volume}(A) \ . \tag{19.7}$$

This is for the same reason that we argued for volume scaling in a random state of a lattice system. In a highly excited random state, the IR modes that contribute to $\tilde{S}$ should all be highly entangled with the outside, and the number of such modes scales with volume.

Example: 2d CFT in vacuum

As a simple example, consider a 2d CFT in the vacuum state of the full system. Space is a line, and region A is an interval of length L_A. In this case the entanglement entropy can be computed exactly (we will do this calculation later in the course) with the result

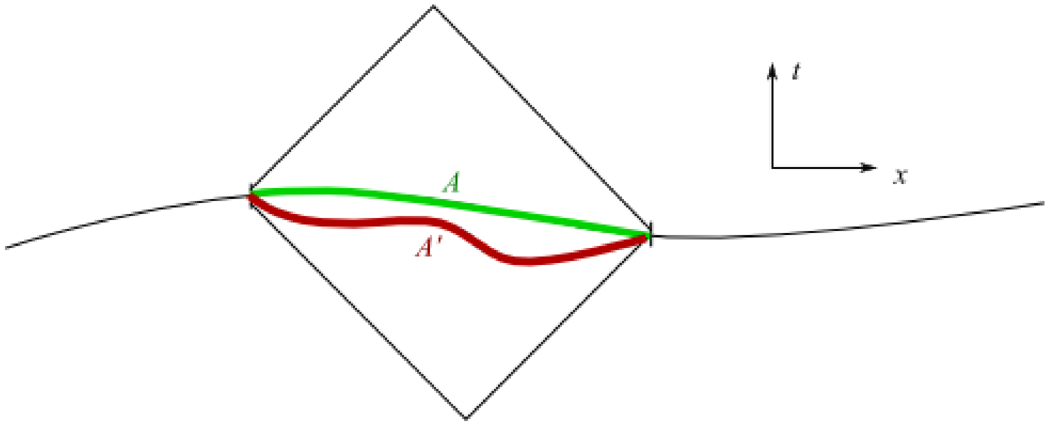
$$S_A = \frac{c}{3} \log \frac{L_A}{\epsilon_{UV}} \ . \tag{19.8}$$

Here c is the central charge of the CFT (which, remember, roughly speaking counts the degrees of freedom). This agrees with the general formula in even spacetime dimensions (19.6), with $\tilde{S} = \frac{c}{3}$.

If we instead consider a highly excited state, then we can't do the calculation in general, but in cases where it can be done the result in a typical state scales as $\tilde{S} \sim cL_A$.

19.2 Lorentz invariance

In a Lorentz-invariant QFT, the density matrix of a spatial region A must contain all of the same information as the density matrix of a spatial region A' that shares the same causal diamond. That is, for this setup:



(19.9)

we must have

$$S_A = S_{A'} \ . \tag{19.10}$$

In words, this is because if we know everything about A, we can time-evolve to learn everything about A'. In formulas, it is because the reduced density matrices are related by (perhaps very complicated and nonlocal!) unitary operation,

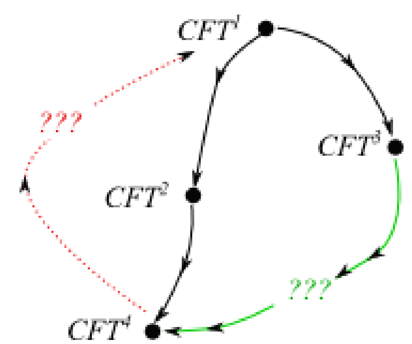
$$\rho_{A'} = U^\dagger \rho_A U \ . \tag{19.11}$$

20 Entanglement Entropy and the Renormalization Group

Entanglement entropy is very difficult to actually calculate in QFT. There are only a few cases where it can be done. So what is it good for? One answer is the relation to quantum gravity, which we'll get to later. Another answer is that entanglement entropy has led to deep insights into the structure of QFT. It is a tool that is almost orthogonal to the usual tools of QFT, and can be used to prove general facts about QFT that, so far, cannot be proved using any other method. The most important example is on the irreversibility of the renormalization group in $d = 3$. We'll now take a brief detour to describe this result and the relevance of entanglement, as pioneered by Casini and Huerta. We restrict to Lorentz-invariant QFTs.

20.1 The space of QFTs

The renormalization group connects conformal field theories:*



(20.1)

Starting with CFT^1 in the UV, we deform by a relevant operator and flow down to CFT^2 or, depending on the deformation, perhaps CFT^3 in the IR. These CFTs might be free, or trivial, as in QCD, which is an RG flow between a free theory in the UV and a gapped (empty) theory in the IR. The IR fixed points may also have relevant perturbations, so we can continue the process and flow to new theories. Two natural questions are:

*Strictly speaking, it connects scale-invariant theories. It is widely suspected that scale invariant QFTs are necessarily conformal, but this is proven only in 2d and in 4d under certain assumptions.

1. Which CFTs can flow to which other CFTs? For example, can the green flow in the figure exist, connecting CFT^3 to CFT^4? Or should it flow from CFT^4 to CFT^3 instead?

2. Can there by closed cycles, connecting the IR back up to the UV like the red dotted flow in the figure?

The RG involves integrating out degrees of freedom, so it would be very strange to find closed cycles! We expect that each time we do into the IR, we reduced the number of degrees of freedom. To make this intuition precise has been a longstanding problem in quantum field theory.

20.2 How to measure degrees of freedom

To make this precise we need to define 'number of degrees of freedom.'

Free energy is no good

One 'obvious' guess fails. Let's try to measure degrees of freedom by computing the thermodynamic free energy, $\log Z$. This can be computed by the Euclidean path integral on $R^{d-1} \times S^1_\beta$. At a fixed point, dimensional analysis fixes

$$F(\beta) = -c_{therm} V_{d-1} T^d \tag{20.2}$$

where c_{therm} is a dimensionless number that we might guess counts degrees of freedom. However, c_{therm} does *not* necessarily decrease along RG flows. An example is the flow from the interacting critical point of N bosons in $d = 3$, to the Goldstone phase with $N - 1$ free bosons.

So we need a more sophisticated measure of degrees of freedom. The correct measure depends on dimension, as do known results about the irreversibility of the RG.

d=2: c-theorem

This case is the easiest and has been understood since the 80s, when Zamolodchikov proved that the correct quantity to consider is the central charge c. Zamolodchikov's c-theorem states

$$c_{UV} \geq c_{IR} \tag{20.3}$$

in a unitary, Lorentz-invariant RG flow.

c plays many roles in a 2d CFT: It appears in the Virasoro algebra, in the trace anomaly, in the stress-tensor correlation functions, in the Casimir energy on a circle, in the thermodynamic free energy, and in the groundstate entanglement entropy. In higher dimensions, these different quantities can have different constants associated to them, so it is not obvious how to generalize (20.3) to higher dimensions. The picture that has emerged in the last few years (conjectured in even dimensions long ago by Cardy) is that the correct quantity to consider is the partition function on S^d. Exactly how this works depends on the dimension.

d=3: F-theorem

The correct measure of degrees of freedom in a 3d CFT is

$$F = -\log |Z_{S^3}| \ . \tag{20.4}$$

It can be shown that this is equal to the finite term in the entanglement entropy of a spherical region. That is, let A be a ball of radius L_A. In the vacuum state the quantity appearing in (19.5) obeys

$$\tilde{S} = F \ . \tag{20.5}$$

This quantity obeys the 'F-theorem',

$$F_{UV} \geq F_{IR} \ . \tag{20.6}$$

This was proved by Casini and Huerta using entanglement methods, described below.

d=4: a-theorem

In even dimensions, the partition function on S^d has a log divergence due from the conformal anomaly. The coefficient of this log divergence is called a:

$$\log Z_{S^4} \sim a \log \frac{R}{\epsilon_{UV}} \ . \tag{20.7}$$

The same number appears in the entanglement entropy of a spherical region. In the notation of (19.6),

$$\tilde{S} \propto a \ . \tag{20.8}$$

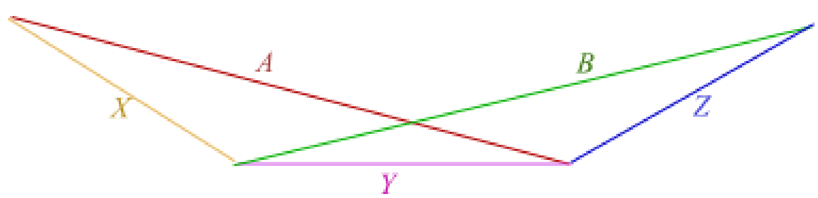
This obeys the 'a-theorem',

$$a_{UV} \geq a_{IR} \ . \tag{20.9}$$

20.3 Entanglement proof of the c-theorem

Zamolodchikov derived the c-theorem in $d = 2$ using standard QFT methods, without reference to entanglement entropy. Later, it was derived using entanglement entropy by Casini and Huerta. Their derivation is very elegant, and exemplifies how entanglement inequalities can be applied in QFT. Unlike Zamolodchikov's proof, it also generalizes to $d = 3$.

We consider any Lorentz-invariant QFT in 2d. Consider two boosted, overlapping intervals A and B, arranged as follows:



(20.10)

We have also labeled the regions X, Y, Z. All of these are spacelike regions. Comparing causal diamonds, Lorentz invariance, as discussed in section 19.2, implies

$$S_A = S_{X\cup Y} \ , \quad S_B = S_{Y\cup Z} \tag{20.11}$$

and

$$S_{A\cup B} = S_{X\cup Y\cup Z}, \qquad S_{A\cap B} = S_Y \ . \tag{20.12}$$

Now, strong subadditivity implies

$$S_A + S_B \geq S_{A\cup B} + S_{A\cap B} \tag{20.13}$$

i.e., (with $\cup$'s implied)

$$S_{XY} + S_{YZ} \geq S_Y + S_{XYZ} \ . \tag{20.14}$$

Parameterize the region lengths by r and R with

$$\ell(A) = \sqrt{rR}, \qquad \ell(Y) = R \ . \tag{20.15}$$

In the vacuum state, the entanglement entropy can depend only on the proper length of the region. Thus SSA becomes

$$2S(\sqrt{rR}) \geq S(R) + S(r) \ . \tag{20.16}$$

Expanding with $R = r + \epsilon$, this means

$$rS''(r) + S'(r) \leq 0 \tag{20.17}$$

or equivalently

$$C'(r) \leq 0, \qquad C(r) = rS'(r) \ . \tag{20.18}$$

(20.18) is the main technical result: the function $C(r)$ is monotonic as a function of interval size. Now for the interpretation. First, suppose our QFT is scale invariant. In this case, from (19.8), the entanglement entropy is

$$S_{cft}(r) = \frac{c}{3}\log\frac{r}{\epsilon_U V} \ . \tag{20.19}$$

Thus the Casini-Huerta C-function $C(r)$ is proportional to the central charge at a critical point,

$$C_{cft}(r) \equiv rS'(r) = \frac{c}{3} \ . \tag{20.20}$$

Now, if the QFT is not scale invariant, then it describes an RG flow between some UV CFT and some IR CFT. That is, the QFT at very short distances is equivalent to CFT_{UV}, and the QFT at very long distances is CFT_{IR}. We are interpreting the *physical* distance r as the RG scale. But we know that at the fixed points, $C(r)$ is given by the central charge,

$$C(r \to 0) = \frac{c_{UV}}{3} \ , \qquad C(r \to \infty) = \frac{c_{IR}}{3} \ . \tag{20.21}$$

Integrating the equation $C'(r) \leq 0$ from short to long distances,

$$\int_0^\infty drC'(r) \leq 0 \ . \tag{20.22}$$

This proves the c-theorem,

$$c_{UV} \geq c_{IR} \ . \tag{20.23}$$

Note that nowhere in this proof have we used the concept of a quantum field!!! We used only locality, Lorentz invariant, quantum mechanics, and unitarity (in the guise of the SSA inequality).

20.4 Entanglement proof of the F theorem

Casini and Huerta's proof of the F theorem $d = 3$ is quite similar. In this case, there is no other known way to prove that RG flows are irreversible – standard field theory methods in even dimension rely on the conformal anomaly, which does not exist in odd dimensions.

We will just briefly sketch the argument, since it is similar to $d = 2$. In a 3d CFT in vacuum,

$$S_A^{CFT} \sim \frac{r}{\epsilon_{UV}} - \tilde{S} \tag{20.24}$$

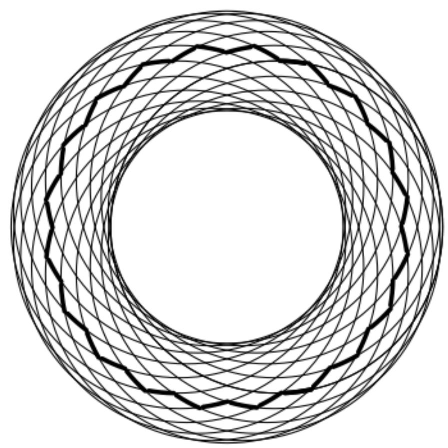
where $\tilde{S}$ is a constant, independent of r and ϵ_{UV}. Therefore a natural guess for the monotonic function is

$$F(r) = rS'(r) - S(r) \ , \tag{20.25}$$

which agrees with $\tilde{S}$ at a critical point,

$$F_{CFT} = \tilde{S} \ . \tag{20.26}$$

To use SSA, we use a more clever version of the boosted-interval setup. Two boosted balls, won't work, because the union of the causal domain of two boosted balls is not the causal domain of any ball. Instead we must arrange an infinite number of boosted regions. Projected onto a single time slice, they look like this:*



(20.27)

An argument similar to 2d implies that $F'(r) \leq 0$, which establishes the F-theorem:

$$F_{UV} \geq F_{IR} \ . \tag{20.28}$$

*Figure taken from Casini and Huerta 1202.5650.

21 Holographic Entanglement Entropy

21.1 The formula

We now turn to entanglement entropy in CFTs with a semiclassical holographic dual. That is, we assume the CFT has a large number of degrees of freedom $N_{dof} \gg 1$ (so that $\ell_{AdS} \gg \ell_{Planck}$) and a sparse low-lying spectrum (to suppress higher curvature corrections, *i.e.*, $\ell_{AdS} \gg \ell_{string}$). We also assume that the CFT is in a state ρ with a geometric dual. This last assumption is needed since even in a holographic CFT, not every state corresponds to a particular geometry (consider, for example, a linear superposition of two black hole microstates with very different energies).

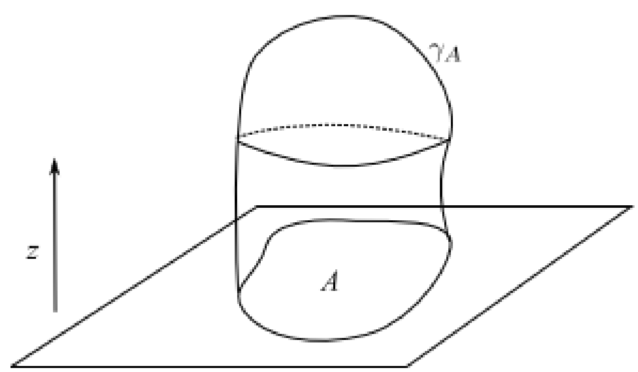
The entanglement entropy in this case is given by the *holographic entanglement entropy formula*:

$$S_A = \frac{\text{area}(\gamma_A)}{4G_N} \ , \tag{21.1}$$

where γ_A is a codimension-2, spacelike extremal surface in the dual geometry, anchored to the AdS boundary such that

$$\partial\gamma_A = \partial A \ . \tag{21.2}$$

An extremal surface is a surface of extremal area. This looks roughly as follows, with z the radial direction in AdS:



(21.3)

γ_A lives in a particular spacelike slice, so that is what is drawn here with the orthogonal (time) direction suppressed.

Two additional comments: First, in (21.1), we are only allowed to include extremal surfaces γ_A which are homologous (continuously deformable) to region A. Second, if there are multiple extremal surfaces satisfying the homology condition, then the rule is to apply (21.1) to the one which has minimal area.

History and Nomenclature

In the static case, this is generally referred to as the Ryu-Takayanagi formula, after the authors who conjectured it in 2006. In time-dependent geometries it's called the HRT formula, after Hubeny, Rangamani, and Takayanagi. Important refinements, discussed below, were also made by Headrick, and many others. The static formula, and the time-dependent formula in certain special states, were derived from the AdS/CFT dictionary $Z_{cft}(M) = Z_{grav}(\text{bdry} = M)$ by Lewkowycz and Maldacena in 2013. Because RTHHRTLM is a mouthful, I will refer to the general formula (21.1) as the HEE (holographic entanglement entropy) formula.

Static case

In a static geometry there is a natural t coordinate, and symmetry implies that γ_A will always lie within a fixed-t slice. An extremal surface in a fixed-t slice is the same as a 'minimal area surface' inside this slice, so in this case the HEE formula reduces to finding a minimal-area surface in a $d-1$-dimensional space geometry.

Extremal surfaces are minimal-area with respect to deformations inside a fixed-t slice, but maximal-area with respect to deformations in the t direction (since we can always reduce the area of a surface by making it 'more null'). The same is true of spacelike geodesics, which extremize the length of a curve in spacetime, rather than minimizing or maximizing it.

21.2 Example: Vacuum state in 1+1d CFT

Consider a 2d CFT in vacuum. Let region A by an interval of length L_A,

$$x \in [-\frac{L_A}{2}, \frac{L_A}{2}] \ . \tag{21.4}$$

The dual geometry is empty AdS_3, with metric

$$ds^2 = \frac{\ell^2}{z^2}(-dt^2 + dx^2 + dz^2) \ . \tag{21.5}$$

z is the radial direction, with the boundary at $z = 0$.

The state is static, so we can set $t = 0$. A codimension-2 extremal 'surface' in AdS_3 is one-dimensional, *i.e.,* a geodesic. So the HEE formula instructs us to find a spacelike geodesic, in the space geometry

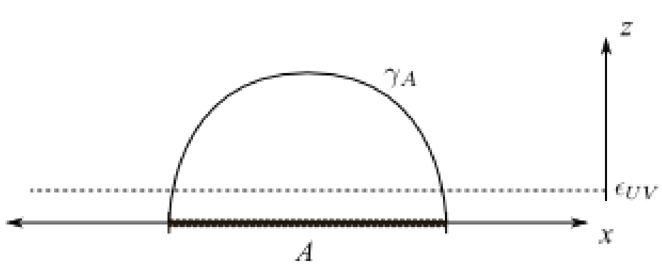
$$ds^2 = \frac{\ell^2}{z^2}(dx^2 + dz^2) \ , \tag{21.6}$$

connecting the points

$$P_1 = (z_1, x_1) = (0, -\frac{L_A}{2}) \quad \text{and} \quad P_2 = (z_2, x_2) = (0, \frac{L_A}{2}) \ . \tag{21.7}$$

However, a geodesic that reaches the boundary like this will have infinite length, since $\int \frac{dz}{z} = \infty$. This is the gravity dual of the statement that entanglement entropy in QFT is UV divergent. To regulate the divergence, we follow the same procedure we used to regulate the on-shell action, or holographic correlation functions: cut off the spacetime at $z = \epsilon_{UV}$.

Thus we want to compute the length of this curve:



(21.8)

Parameterizing the curve $(x(\lambda), z(\lambda))$ by z, the regulated geodesic length is

$$\begin{aligned} Length &= \int ds \\ &= 2L_A \int_{\epsilon}^{z_{max}} \frac{dz}{z} \sqrt{x'(z)^2 + 1} \end{aligned} \tag{21.9}$$

The factor of 2 is because we the geodesic goes out, and comes back, and we will only integrate $z \in [\epsilon, z_{max}]$ once. Treating (21.9) as a 1d "action", it is easy to show that the geodesic is a semicircle,

$$x = \frac{L_A}{2}\cos\lambda, \quad z = \frac{L_A}{2}\sin\lambda, \quad \lambda \in (\frac{\epsilon}{L_A}, \pi - \frac{\epsilon}{L_A}) \ . \tag{21.10}$$

Plugging this back into (21.9) and doing the integral gives

$$Length = 2L_A \log\left(\frac{L_A}{\epsilon_{UV}}\right) \ . \tag{21.11}$$

Therefore, applying the HEE formula (21.1),

$$S_A = \frac{L_A}{2G_N}\log\left(\frac{L_A}{\epsilon_{UV}}\right) \ . \tag{21.12}$$

The map between gravity parameters and CFT parameters in AdS_3/CFT_2 is

$$c = \frac{3\ell}{2G_N} \ , \tag{21.13}$$

where c is the central charge, so

$$S_A = \frac{c}{3}\log\left(\frac{L_A}{\epsilon_{UV}}\right) \ . \tag{21.14}$$

This agrees perfectly with our general discussion of the structure of entanglement entropy in QFT in even spacetime dimensions, (19.6), including the UV divergence. The prefactor also agrees exactly with the known result in 2d CFT, (19.8).

Exercise: 2d HEE

Fill in all the missing steps — *i.e.,* solve for the geodesic and do the length integral — in the derivation of (21.14). Don't forget to use conserved quantities to efficiently solve the geodesic equation.

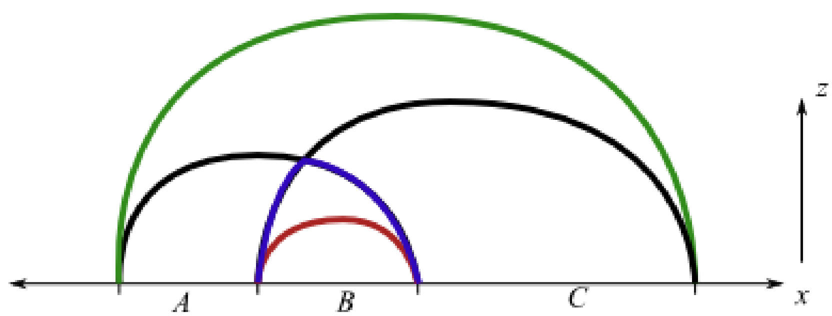
Exercise: Strips in d dimensions
Compute the holographic entanglement entropy of an infinite strip of width L_A in a d-dimensional CFT in the vacuum state. That is, with CFT coordinates $(t, x, \vec{y})$, region A is the region $x \in [-L_A/2, L_A/2]$, $t = 0$, $\vec{y}$ =anything.

21.3 Holographic proof of strong subadditivity

The proof of the strong subadditivity inequality in quantum mechanics is rather technical and tricky. The holographic proof, in static states, is easy! The statement of SSA for a tripartite system is

$$S_{ABC} + S_B \leq S_{AB} + S_{BC} \ . \tag{21.15}$$

Let's draw the various minimal-area surfaces:



(21.16)

We've picked a color scheme to reorganize the inequality a bit, so now it says

$$\text{red} + \text{green} \leq \text{blue} + \text{black} \ . \tag{21.17}$$

The fact that the curves are minimal area immediately implies

$$\text{red} \leq \text{blue}, \qquad \text{green} \leq \text{black} \tag{21.18}$$

so SSA follows.

This argument has also been extended to the time-dependent case. It is much trickier, since the extremal surfaces need not all lie in the same spatial slice.

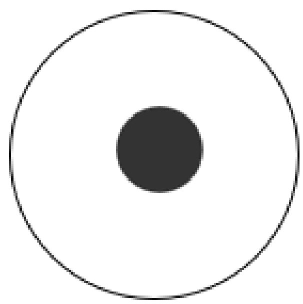
21.4 Some comments about HEE

HEE and Bekenstein-Hawking

The HEE formula is a generalization of the Bekenstein-Hawking area law for black hole entropy. To see this, let's apply the formula to a static black hole spacetime, and choose region A to be all of space. In this case the boundary condition (21.2) on the extremal surface is

$$\partial\gamma_A = \emptyset \ . \tag{21.19}$$

It is tempting to say γ is the empty set, but this would not satisfy the homology condition. A spatial slice of the black hole spacetime looks like



(21.20)

This is not simply connected, and the 'empty set' curve is not deformable to region A. So, in fact, we must choose γ_A to be the horizon itself (which is extremal). Thus

$$S_A = \frac{\text{area(horizon)}}{4G_N} \tag{21.21}$$

in agreement with Bekenstein-Hawking.

But why is this 'entanglement entropy'? Actually, it might not be. More accurately, the HEE formula computes the von Neumann entropy of the reduced density matrix, $S_A = -\operatorname{tr}\rho_A \log\rho_A$. This von Neumann entropy may or may not come from entanglement — we can't tell the difference without knowing the full system. In the black hole spacetime, the ordinary thermal entropy is the von Neumann entropy of the thermal state $\rho = e^{-\beta H}$, so the HEE formula applied to the full space gives the thermal entropy. You can also think of this as actual entanglement entropy coming from the entanglement of the CFT with the thermal double.

Some words

Entanglement entropy is a measure of how quantum information is spatially organized in a quantum state. In a general QFT, it is extremely complicated, and we do not expect any tractable simple formula. The fact that it simplifies, and becomes geometric, in holographic CFTs is a deep fact about strongly coupled systems. It means that the organization of quantum information approaches a sort of simplified, universal limit at strong coupling. How this happens and exactly how it is related to emergent geometry is an unsolved, and presumably very important, problem in current research.

22 Holographic entanglement at finite temperature

In this section we will discuss the HEE computation in a black hole spacetime. For explicitness, we will talk about the BTZ black hole in AdS_3, but all of the results hold qualitatively in higher dimensions, too.

The BTZ metric is static, so we need only the fixed-time metric ($\ell = G = 1$)

$$ds^2 = \frac{dr^2}{r^2 - 8M} + r^2 d\phi^2 \ . \tag{22.1}$$

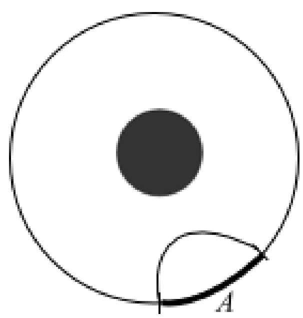
This is dual to a finite-temperature state $\rho = e^{-\beta H}$ with temperature

$$T = \sqrt{8M}/2\pi \ . \tag{22.2}$$

Choose region A to be the boundary segment

$$A : \qquad \phi \in (0, R) \ . \tag{22.3}$$

So the figure is



(22.4)

There are many geodesics connecting the endpoints of region A. In fact there are an infinite number, labeled by the integer number of times that the geodesic winds the black hole. If $R \ll 2\pi$, then there is one that is obviously the shortest, which does not wind the black hole. This is the one drawn in the figure. The length of this geodesic is infinite, but if we impose a cutoff at $r = 1/\epsilon_{UV}$, the resulting entropy is

$$\begin{aligned} S_A^{(0)} &= \frac{\text{length}(\gamma_A^{(0)})}{4G_N} \\ &= \frac{c}{3} \log \left[\frac{1}{\pi T \epsilon_{UV}} \sinh(\pi R T) \right] \ , \end{aligned} \tag{22.5}$$

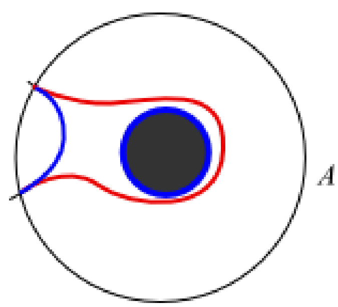
with $c = 3\ell/2G_N$.

For $0 < R \ll 2\pi$, this is the final answer. For $R > 2\pi$, the same formula (22.5) computes the length of a geodesic that winds (possibly multiple times) around the horizon. The winding geodesics do not satisfy the homology condition, *i.e.,* they cannot be continuously deformed to A. But we must also consider disjoint geodesics. For example, the horizon itself is a geodesic $\mathcal{H}$. This can be added to to the wrapped geodesic $\gamma^{(1)}$. The union

$$\gamma^{(1)} \cup \mathcal{H} \tag{22.6}$$

is homologous to region A, if we view the orientation of $\mathcal{H}$ as opposite that of $\gamma_A^{(1)}$.

If region A is large, we must choose the dominant (minimal area) surface. We are choosing between the red (wrapping) and blue (disjoint) surfaces in this figure:



(22.7)

The length of the red curve is what we computed above,

$$S_A^{red} = \frac{c}{3} \log \left[\frac{1}{\pi T \epsilon_{UV}} \sinh(\pi R T) \right] . \tag{22.8}$$

The length of the blue curve gets a contribution from the short part, and a contribution from the horizon:

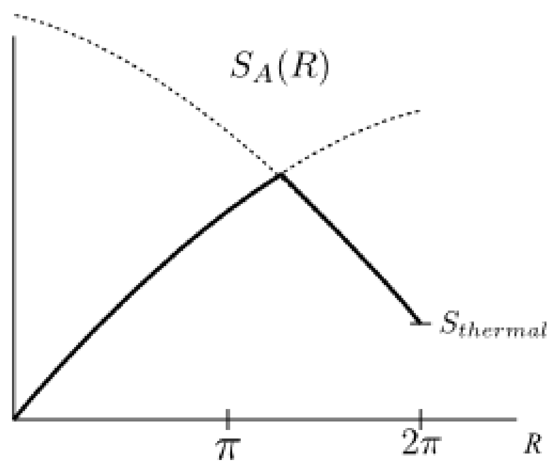
$$S_A^{blue} = \frac{c}{3} \log \left[\frac{1}{\pi T \epsilon_{UV}} \sinh(\pi(2\pi - R)T) \right] + \frac{c}{3} 2\pi^2 T \tag{22.9}$$

The first term is the answer above, applied to A^C. The second term is the thermal entropy, which we know is area(horizon)/4.

The final answer is

$$S_A(R) = \min \left[S_A^{red}, \quad S_A^{blue} \right] \tag{22.10}$$

The two contributions exchange dominance at some point $R_* > 2\pi$. At this point there is a sharp transition in the behavior of the entanglement entropy. The plot as a function of system size is something like the solid line in this figure:



(22.11)

Pure vs mixed

Clearly with a black hole, $S_A \neq S_{A^C}$ due to the homology condition. In fact we found (for $R > R_*$)

$$S_A = S_{A^C} + S_{thermal} \ . \tag{22.12}$$

Since $S_{thermal}$ is the von Neumann entropy of the full space, this can be written

$$S_A = S_B + S_{AB} \tag{22.13}$$

Thus the entanglement entropy of the black hole saturates the Araki-Lieb triangle inequality,

$$S_{AB} \geq |S_A - S_B| \ . \tag{22.14}$$

This is a special feature of thermal states in holographic systems.

22.1 Planar limit

The infinite-volume limit of the CFT is a limit of the black hole where the horizon becomes planar. We can take this limit for BTZ by assuming

$$TR \gg 1, \qquad R \ll 2\pi \ . \tag{22.15}$$

In this limit, our answer (22.5) reduces to

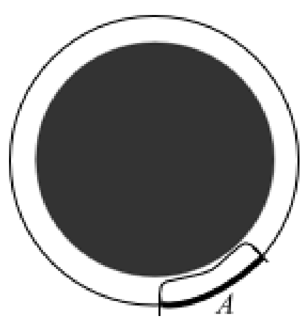
$$S_A \approx \frac{c}{3}\pi TR + \text{subleading} \tag{22.16}$$

This is equal to the thermal entropy density,

$$S_A \approx S_{thermal}(\beta) \times \frac{R}{2\pi} \ . \tag{22.17}$$

This makes sense: in the thermodynamic limit, the state is very mixed up, and the subsystem itself just looks thermal at temperature β.

Geometrically, the reason behind (22.17) is that, in this limit, the extremal surface 'hugs' the black hole horizon for most of its length:



(22.18)

The contribution from the horizon is proportional to the horizon area, *i.e.,* to the thermal entropy.

We have discussed BTZ, but the same feature generalizes to planar horizons in higher dimensions: for $L_A \gg \beta$, the extremal surface hugs the black hole horizon, giving a volume-law contribution to the finite-temperature entanglement entropy.

23 The Stress Tensor in 2d CFT

In the last few lectures, we will go into more depth on the AdS_3/CFT_2 correspondence. First we need to cover some more ground in the basics of 2d CFT.

References: For 2d CFT I recommend: Chapter 2 of Polchinski's text; Chapter 4 of Kiritsis's text; the big book of Di Francesco et al; and the string theory lectures notes by David Tong, available online. For the optimal introduction to the subject, I recommend working through the chapter of Tong's notes first, then working through chapters 4-6 of Di Francesco et al.

23.1 Infinitessimal coordinate changes

Recall that in two dimensions, with complex coordinates in Euclidean $\mathbf{R}^2$,

$$ds^2 = dzd\bar{z} \ , \tag{23.1}$$

conformal transformations are holomorphic coordinate changes:

$$w = w(z), \quad \bar{w} = \bar{w}(\bar{z}) \ . \tag{23.2}$$

The coordinate change

$$z' = z + \epsilon(z) \tag{23.3}$$

corresponds to the vector field

$$\zeta^\mu \partial_\mu = -\epsilon(z)\partial_z \tag{23.4}$$

They act on fields as

$$\phi'(z', \bar{z}') = \phi(z', \bar{z}') + \zeta^\mu \partial_\mu \phi \tag{23.5}$$

The infinitessimal generators can be taken as

$$\zeta_n = -z^{n+1}\partial_z, \quad \bar{\zeta}_n = -\bar{z}^{n+1}\partial_{\bar{z}} \ . \tag{23.6}$$

The conformal generators make an algebra

$$[\zeta_m, \zeta_n] = (m-n)\zeta_{m+n} \ , \tag{23.7}$$

and similarly for the barred generators. This is called the 'Witt algebra' (or centerless Virasoro algebra).

The algebra is infinite-dimensional. There is one subalgebra, which consists of the 3 generators $\zeta_{0,1,-1}$. These make the *global subalgebra* SL(2):

$$\zeta_{-1}, \zeta_0, \zeta_1 \quad \text{and} \quad \bar{\zeta}_{-1}, \bar{\zeta}_0, \bar{\zeta}_1 \quad \Rightarrow \quad SL(2,R) \times SL(2,R) \sim SO(2,2) \ . \tag{23.8}$$

It is called the global subalgebra because these are the only $\zeta_n' s$ that are non-singular on the Riemann sphere. To see this, first look near $z \sim 0$. Clearly the vector field $\zeta_n = -z^{n+1}\partial_z$ is regular only for $n \geq -1$. Now, do the coordinate change $w = 1/z$, and look near $w \sim 0$. The vector field becomes

$$\zeta_n = -z^{n+1}\partial_z = -w^{-n-1}(-w^2)\partial_w \tag{23.9}$$

which is regular only for $n \leq 1$. So the generators in (23.8) are the only ones regular at both poles of the Riemann sphere.

23.2 The Stress Tensor

Translation invariance implies the action S is invariant under $x^\mu \to x^\mu + \epsilon^\mu$, for constant ϵ^μ. The classical stress tensor is defined by applying the Noether to this symmetry. Promoting ϵ^μ to an arbitrary function of x^μ and varying the action must give something proportional to $\partial\epsilon$,

$$\delta S = -2 \int d^2 z \sqrt{g} T^{\mu\nu} \partial_\mu \epsilon_n \ . \tag{23.10}$$

This defines the stress tensor*

$$T_{\mu\nu} = -\frac{4\pi}{\sqrt{g}} \frac{\delta S}{\delta g^{\mu\nu}} \ . \tag{23.11}$$

*Note the extra -2π compared to our when we computed gravitational stress tensors a while back. This is just a convention but will be important to remember when we compare the two.

It's conserved

The Noether procedure guarantees that

$$\nabla_\mu T^{\mu\nu} = 0 \ . \tag{23.12}$$

In a flat background (23.1), in complex coordinates the two components of this equation are

$$\bar{\partial} T_{zz} = 0 \ , \qquad \partial T_{\bar{z}\bar{z}} = 0 \ . \tag{23.13}$$

where we have introduced the shorthand

$$\partial \equiv \partial_z, \qquad \bar{\partial} = \partial_{\bar{z}} \ . \tag{23.14}$$

From (23.13), T_{zz} is a holomorphic function of z, and $T_{\bar{z}\bar{z}}$ is anti-holomorphic. These will be denoted

$$T(z) \equiv T_{zz}, \qquad \bar{T}(\bar{z}) = T_{\bar{z}\bar{z}} \ . \tag{23.15}$$

And traceless

So far we have used only translation invariance. In a theory that is classically scale invariant, we can also conclude that the trace of the stress tensor vanishes. The symmetry in this case is the infinitessimal rescaling

$$x^\mu \to x^\mu + \lambda x^\mu \ . \tag{23.16}$$

Since under a rescaling

$$\partial_\mu \epsilon_\nu = \lambda g_{\mu\nu} \ , \tag{23.17}$$

the variation of the action is

$$\delta S \propto \int d^2 z \sqrt{g} \lambda T^\mu_\mu . \tag{23.18}$$

Scale invariance implies that this integral vanishes; conformal invariance, or local scale invariance, means we can make $\lambda \to \lambda(z, \bar{z})$ so $T^\mu_\mu = 0$. In complex coordinates,

$$T_{z\bar{z}} = 0 \ . \tag{23.19}$$

This is the classical stress tensor. Even if it is traceless, the quantum stress tensor might have a non-zero trace, for two different reasons. First, the UV regulator introduces a scale, and may introduce a trace. In fact, in a renormalizable theory,

$$T^\mu_\mu(x) = \sum_i \beta_{g_i} O_i(x) \ , \tag{23.20}$$

where O_i are the relevant operators of the theory, g_i are the corresponding couplings, and β_g are their beta functions. So, for example, in massless QCD, although the classical stress tensor is traceless, the quantum stress tensor has a contribution from the non-zero QCD beta function. In a CFT, all the beta functions are exactly zero, so the equation

$$T^\mu_\mu = 0 \tag{23.21}$$

is true as an operator statement. The phrase "as an operator statement" means the equation is true *inside* correlation functions, up to delta-functions where T^μ_μ hits other operator insertions (we'll see some of these delta functions below).

The second origin of a non-zero trace is a quantum anomaly. This happens even in CFT, if we place the theory on a curved background. This is called the Weyl anomaly and is important but we probably won't have time to cover it.

Noether currents for conformal symmetries

$T(z)$ is the Noether current for translations. That is, the current J^μ with

$$J^{\bar{z}} = T(z) \tag{23.22}$$

is the current associated to $z \to z+$const. What are the Noether currents associated to the general conformal transformation $z \to z + \epsilon(z)$? These are simply

$$J^{\bar{z}} = \epsilon(z) T(z) \ . \tag{23.23}$$

This is sort of obvious; to reproduce it from the Noether procedure, you can promote $\epsilon(z) \to \epsilon(z, \bar{z})$ and apply the usual Noether procedure.

As expected, the current is conserved,

$$\partial_\mu J^\mu = \bar{\partial}(\epsilon T) = 0 \ . \tag{23.24}$$

23.3 Ward identities

Ward identities are the quantum version of the Noether procedure.

Suppose we have a general symmetry $\phi' = \phi + \epsilon\delta\phi$. The fact that this is a symmetry means the action and the path integral measure are invariant,

$$S[\phi'] = S[\phi], \qquad D\phi' = D\phi \ . \tag{23.25}$$

Thus, promoting $\epsilon \to \epsilon(x^\mu)$,

$$\int D\phi e^{-S[\phi]} = \int D\phi' e^{-S[\phi']} \tag{23.26}$$

$$= \int D\phi e^{-S[\phi]-\int J^\mu \partial_\mu \epsilon} \tag{23.27}$$

$$= \int D\phi (1 - \int J^\mu \partial_\mu \epsilon) e^{-S[\phi]} \tag{23.28}$$

The first line is just renaming a dummy variable; the second line defines the current, J^μ, which may have contributions from both the classical action and the measure; and the third line expands to linear order. It follows that

$$\langle \int J^\mu \partial_\mu \epsilon \rangle = 0 \tag{23.29}$$

for all ϵ, and so

$$\langle \partial_\mu J^\mu \rangle = 0 \ . \tag{23.30}$$

This is the quantum version of the Noether procedure. The same exact steps, starting instead with $\int D\phi O_1(x_1)\cdots O_n(x_n) e^{-S[\phi]}$ can be used to show that

$$\langle \partial_\mu J^\mu(y) O_1(x_1)\cdots O_n(x_n)\rangle = 0 \quad \text{if} \quad x_i \neq y \ . \tag{23.31}$$

The restriction to $x_i \neq y$ is necessary for the derivation to work. Just set the support of

ϵ to a small circle around y that does not include any of the other operator insertions, and repeat the steps above.

The equation (23.31) means

$$\partial_\mu J^\mu = 0 \tag{23.32}$$

This is an operator equation, ie it holds inside correlators, up to delta functions.

If there are insertions that collide with $\partial_\mu J^\mu$, we have to be more careful. Suppose $x_1 = y$. Then when we do the transformation inside the path integral, the transformation also affects this operator insertion,

$$O_1 \to O_1 + \epsilon\delta O_1 \ . \tag{23.33}$$

So now,

$$\int D\phi O_1(x_1)\cdots O_n(x_n)e^{-S[\phi]} = \int D\phi(1 - \int J^\mu \partial_\mu \epsilon)(O_1 + \epsilon\delta O_1)O_2\cdots O_n e^{-S[\phi]} \tag{23.34}$$

and the conservation law is modified to (restoring some dropped constants)

$$-\frac{1}{2\pi}\int_{D(y)} \partial_\mu \langle J^\mu(y)O_1(x_1)\cdots O_n(x_n)\rangle = \langle \delta O_1(x_1)O_2(x_2)\cdots O_n(x_n)\rangle, \tag{23.35}$$

where $D(y)$ is a disk enclosing y, where we've chosen ϵ to be non-zero. Allowing for any of the operators to collide with the current, (23.35) becomes

$$\partial_\mu \langle J^\mu(y)O_1(x_1)\cdots O_n(x_n)\rangle = \sum_i \delta(y - x_i)\langle O_1\cdots \delta O_i \cdots O_n\rangle \ . \tag{23.36}$$

(23.36) is called the Ward identity.

As a residue

In two dimensions, using complex coordinates we can write (23.36) in a nice way. By Stokes, we have

$$\int_{D(y)} \partial_\mu J^\mu \sim \oint_y (J_z dz - J_{\bar z} d\bar z) \tag{23.37}$$

and

$$\frac{i}{2\pi}\oint_x dz J(z) O(x) = -\text{res}_{z\sim x} J(z) O(x) \ . \tag{23.38}$$

Therefore, the Ward identity (23.36) can be written as the operator equation

$$\delta O(x) = -\text{res}_{z\sim x} \left[J(z) O(x)\right] \ . \tag{23.39}$$

Conformal Ward Identities

The Noether current for the conformal symmetry $z \to z + \epsilon(z)$, from (23.23), is

$$J(z) = \epsilon(z) T(z) \ . \tag{23.40}$$

Therefore the Ward identity for conformal transformations, allowing for both holomorphic and anti-holomorphic transformations, is

$$\delta_{\epsilon,\bar{\epsilon}} O(x) = -\text{res}_{z\sim x} \left[\epsilon(z) T(z) O(x)\right] \ . \tag{23.41}$$

23.4 Operator product expansion

The product of two local operators can be Taylor-expanded as local operators at a single point:

$$O_1(x) O_2(y) = \sum_j f_{12j}(x-y) O_j(u) \ . \tag{23.42}$$

The sum is over all local operators in the theory. This is called the operator product expansion (OPE). The OPE coefficients $f_{12j}(x-y)$ depend on the theory, but they are restricted by conformal invariance, so that in fact the f_{ijk}'s of primary operators determine the f_{ijk}'s of descendants as well.

The Ward identity, in the form (23.41), means that the singular terms in the $O(x)T(y)$ OPE are related to the conformal transformations of O.

Primaries

Recall that primary operators transform as

$$O'(z',\bar{z}') = \left(\frac{dz'}{dz}\right)^{-h}\left(\frac{d\bar{z}'}{d\bar{z}}\right)^{-\bar{h}} O(z,\bar{z}) \tag{23.43}$$

where $(h,\bar{h})$ are called the conformal weights of O. The corresponding infinitessimal transformation is

$$\begin{aligned} \delta_{\epsilon,\bar{\epsilon}} O(z,\bar{z}) &\equiv O'(z,\bar{z}) - O(z,\bar{z}) && (23.44)\\ &= -(hO\partial\epsilon + \epsilon\partial O) - (\bar{h}O\bar{\partial}\bar{\epsilon} + \bar{\epsilon}\bar{\partial}O) && (23.45) \end{aligned}$$

Comparing to (23.41), this means

$$T(z)O(w,\bar{w}) \sim \frac{hO(w,\bar{w})}{(z-w)^2} + \frac{\partial O(w,\bar{w})}{z-w} \ . \tag{23.46}$$

The symbol '$\sim$' means that we only written the singular terms in the OPE; there is also an infinite series of contributions regular at $z = w$. To check this gives the correct residue, expand

$$\epsilon(z)T(z)O(w) = T(z)\epsilon(w)O(w) + (z-w)T(z)\epsilon'(w)O(w) + \cdots \tag{23.47}$$

and look at the simple pole in (23.46).

Upshot

The lesson is that *singular terms in the $T\phi$ OPE contain exactly the same information as the transformations of ϕ under conformal symmetry.* The conformal algebra is one and the same as the data in the singular part of the OPE.

23.5 The Central Charge

Now we will examine the transformation of the stress tensor under conformal symmetry, or equivalently, the TT OPE. Much of the discussion in this section is easiest to understand first using an example like a free scalar field. This example is worked in every reference so I will not repeat here but encourage the reader to jump to the free scalar section of Polchinski, Kiritsis, or Tong's notes before continuing.

The stress tensor is not a primary, so we cannot plug $O = T$ into (23.46). But, it does behave similar to a primary under rescalings, with $(h, \bar{h}) = (2, 0)$:

$$T'(\lambda z) = \lambda^{-2} T(z) \ . \tag{23.48}$$

This is basically dimensional analysis; the energy should have mass dimension 1, and $E \sim \int T$. Similarly, $\overline{T}(\bar{z})$ has scaling weights $(h, \bar{h}) = (0, 2)$. Requiring both sides of the OPE to have the same scaling weight, it must have the form

$$T(z)T(w) = \frac{c/2}{(z-w)^4} + \frac{X_1(w)}{(z-w)^3} + \frac{X_2(w)}{(z-w)^2} + \frac{X_3(w)}{z-w} + \cdots \ , \tag{23.49}$$

where c is a number (the 2 is inserted by convention), X_1 is some field of dimension $(1, 0)$, X_2 is dimension $(2, 0)$, and X_1 is dimension $(3, 0)$. There are no terms more singular than $(z - w)^4$, since fields must have positive scaling weight in a unitary theory.* The exchange symmetry

$$T(z)T(w) = T(w)T(z) \tag{23.50}$$

set $X_1 = 0$. The OPE is always invariant under permutations (in Euclidean signature). One way to see this is that equal-time commutators in Lorentzian signature must vanish by causality, and these equal-time commutators correspond to permutations of the Euclidean correlator. This property is sometimes called locality or causality.

To fix the other two terms, we use the Ward identity for the scale symmetry (23.48). The Noether current for scale symmetry is, from (23.23)

$$J_{scale}(z) = zT(z) \ . \tag{23.51}$$

Then using the Ward identity (23.41) gives our final answer for the singular terms in the OPE:

$$T(z)T(w) \sim \frac{c/2}{(z-w)^4} + \frac{2T(w)}{(z-w)^2} + \frac{\partial T(w)}{z-w} \ . \tag{23.52}$$

*I will not explain why, but this can be found in the references, in the section on the state-operator correspondence and conformal reps. This also explains why c must be a number, not a field.

The constant c is called the central charge. For a free boson, it turns out to be $c = 1$, and for a free fermion $c = 1/2$. (see references).

Transformation law for the stress tensor

Knowing the TT OPE, we can use the Ward identity $\delta T(w) = -\text{res}_{z\sim w}[\epsilon(z)T(z)T(w)]$ to find how T transforms under a conformal symmetry:

$$\delta T = -\epsilon T' - 2\epsilon' T - \frac{c}{12}\epsilon''' \ . \tag{23.53}$$

The first two terms are the usual transformations of a primary. The last term is an 'anomolous' term coming from the central charge.

The finite transformation law, obtained by integrating (23.53) is

$$T'(z') = \left(\frac{dz'}{dz}\right)^{-2}\left[T(z) - \frac{c}{12}\{z', z\}\right] \ , \tag{23.54}$$

where

$$\{f(z), z\} \equiv \frac{f'''}{f'} - \frac{3}{2}\frac{(f'')^2}{(f')^2} \tag{23.55}$$

is called the *Schwarzian derivative*. The easiest way to derive (23.54) is to check that it agrees infinitessimally with (23.53), and that it composes correctly under multiple transformations.

Aside: Virasoro algebra

We won't cover the operator formalism in this course, so I will not discuss the Virasoro algebra in detail. Roughly, you can decompose the stress tensor into modes, $L_n \sim \oint dz z^{n+1}T(z)$. Then the TT OPE (23.52), combined with the Ward identity, become the Virasoro algebra

$$[L_m, L_n] = (m-n)[L_m - L_n] + \frac{c}{12}(m^3 - m)\delta_{m,-n} \ . \tag{23.56}$$

This is where the name 'central charge' comes from. It is another way of writing (23.52): the TT OPE, the Schwarzian derivative, and the Virasoro algebra all contain the same information.

23.6 Casimir Energy on the Circle

c is related to the Casimir energy of the theory on a circle. The mapping from the plane to the cylinder of radius L is

$$z = e^{2\pi w/L} \ . \tag{23.57}$$

The cylinder coordinate is identified $w \sim w + iL$, since this takes us around a circle on the z plane. Using the finite transformation law (23.54) gives

$$T_{cyl}(w) = \left(\frac{2\pi}{L}\right)^2 z^2 \left(T_{plane}(z) - \frac{c}{24z^2}\right) \ , \tag{23.58}$$

where now we're using 'plane' and 'cyl' to distinguish the stress tensor before and after the transformation.

Let's calculate the expectation value. On the plane, scale invariance sets all 1-point functions to zero:

$$\langle T_{plane}(z)\rangle = 0 \ . \tag{23.59}$$

This is because the only scale-invariant function of z with weight 2 is $1/z^2$, but this would not be translation invariant. Now using (23.58),

$$\langle T_{cyl}(w)\rangle = -\frac{c}{24}\left(\frac{2\pi}{L}\right)^2 \ . \tag{23.60}$$

This is a Casimir energy, in the usual sense: we started with a theory on a line (*i.e.*, space is a line), then imposed periodic boundary conditions with period L, and found energy $\sim 1/L$. The size of the Casimir energy is fixed by the central charge. This is our first indication that c might be a good way to the measure the degrees of freedom of a CFT.

To compute the value of the energy explicitly, choose real coordinates $w = \tau + i\phi$, where $\phi \sim \phi + L$. The energy is defined in the usual way by integrating the stress tensor over a fixed-time slice:

$$E_{cyl} = \frac{1}{2\pi} \in_0^L dx \langle T_{\tau\tau}^{cyl}(\tau = 0, x) \rangle \tag{23.61}$$

$$= \frac{1}{2\pi} \int_0^L dx \langle T_{ww}^{cyl} + T_{\bar{w}\bar{w}}^{cyl} \rangle \tag{23.62}$$

In the second line, we just did the change of coordinates $w = \tau + i\phi, \bar{w} = \tau - i\phi$, and used tracelessness of the stress tensor, $T_{w\bar{w}} = 0$. Evaluating this in vacuum gives the Casimir energy

$$E_{cyl}^{vac} = -\frac{c}{12}\frac{2\pi}{L} \ . \tag{23.63}$$

In a general state, (23.61) gives

$$E_{cyl} = \Delta - \frac{c}{12} \tag{23.64}$$

where

$$\Delta = \langle \frac{1}{2\pi i} \left(\oint dz z T(z) + \oint d\bar{z} \bar{z} \bar{T}(\bar{z}) \right) \rangle \ . \tag{23.65}$$

Recall that $J = zT(z)$ is the Noether current for scale transformations. Therefore, the rhs is the expectation value of the scale operator. In an eigenstate, Δ is the scaling dimension of that state. So (23.64) says that the energy on the cylinder is the scaling dimension on the plane, shifted by the central term. This makes sense, since time translations on the cylinder correspond to scale transformations on the plane.

Exercise: free boson

The action of a free boson is $S = \int d^2z \partial\phi\bar{\partial}\phi$. Use the Noether procedure to find the stress tensor. Then, compute the TT OPE and confirm that $c = 1$. (See Polchinski, Kiritsis, Di Francesco, or Tong's notes if you get stuck).

24 The stress tensor in 3d gravity

Now we will compare the stress tensor of 3d gravity to our results of the previous section. Consider the asymptotically-AdS_3 metric

$$ds^2 = \frac{\ell^2}{r^2}dr^2 + \frac{r^2}{\ell^2}dzd\bar{z} + h_{zz}dz^2 + h_{\bar{z}\bar{z}}d\bar{z}^2 + 2h_{z\bar{z}}dzd\bar{z} \ . \tag{24.1}$$

We have not written every possible term in the perturbation $h_{\mu\nu}$, but it turns out that other terms can be removed by a diff. Also, we will keep only the leading term in $h_{\mu\nu}$ at large r, *i.e.,* near the boundary, so we can assume that $h_{\mu\nu}$ is independent of r.

The Einstein equations imply that the perturbation is traceless and conserved:

$$h_{z\bar{z}} = 0 \tag{24.2}$$

$$\partial h_{\bar{z}\bar{z}} = \bar{\partial} h_{zz} = 0 \ . \tag{24.3}$$

24.1 Brown-York tensor

The stress tensor of this geometry was computed in an exercise from one of the early lectures. Let's briefly review how this works. The Brown-York stress tensor is*

$$T_{ij} \equiv -\frac{4\pi}{\sqrt{g}}\frac{\delta S^{on-shell}_{Einstein}}{\delta g^{ij}} \tag{24.4}$$

$$= -\frac{1}{4}\left(K_{ij} - Kg_{ij} - \frac{1}{\ell}g_{ij}\right) \ . \tag{24.5}$$

The first two terms came from varying the Einstein action plus the Gibbons-Hawking boundary term. The last term comes from the counterterm, with the coefficient set in order to make the answer finite as $r \to \infty$. Plugging the metric (24.1) into (24.5), using (24.2) and doing a lot of work, eventually

$$T_{zz} = -\frac{1}{4\ell}h_{zz} \ , \qquad T_{\bar{z}\bar{z}} = -\frac{1}{4\ell}h_{\bar{z}\bar{z}} \ . \tag{24.6}$$

*We've changed conventions by a factor of 2π compared to some earlier lectures. This is just a choice, made to agree with our convention for the CFT stress tensor.

Thus the Brown-York metric is traceless, conserved, and therefore holomorphic/anti-holomorphic just like in CFT.

24.2 Conformal transformations and the Brown-Henneaux central charge

Under diffeos, the metric (24.1) transforms as

$$\delta g_{\mu\nu} = \mathcal{L}_\zeta g_{\mu\nu} \ . \tag{24.7}$$

What vector fields ζ preserve the form of the metric (24.1)? The answer is

$$\begin{aligned} z &\rightarrow z + \epsilon(z) - \frac{\ell^4}{2r^2}\bar{\epsilon}''(\bar{z}) \\ \bar{z} &\rightarrow \bar{z} + \bar{\epsilon}(\bar{z}) - \frac{\ell^4}{2r^2}\epsilon''(z) \\ r &\rightarrow r - \frac{r}{2}\epsilon'(z) - \frac{r}{2}\bar{\epsilon}'(\bar{z}) \end{aligned} \tag{24.8}$$

for arbitrary functions $\epsilon(z)$ and $\bar{\epsilon}(\bar{z})$. Near the boundary, these act on $z, \bar{z}$ just like conformal transformations. The extra ∂_r piece acts as a rescaling.

Thus transformations of AdS_3 that preserve the asymptotics of the metric coincide with 2d conformal transformations!

Let's set $\bar{\epsilon} = 0$ and focus on the holomorphic conformal transformations. Under (24.8), the metric transforms as

$$ds^2 \rightarrow ds^2 + \left(-2h_{zz}\epsilon' - \epsilon h'_{zz} + \frac{\ell^2}{2}\epsilon'''\right) dz^2 \tag{24.9}$$

Thus the dz^2 piece of the metric, which we interpreted as the gravitational stress tensor up to a factor of $-1/4\ell$, transforms as

$$\delta_\epsilon T = -\epsilon\partial T - 2T\partial\epsilon - \frac{\ell}{8}\epsilon''' \ . \tag{24.10}$$

This is exactly the transformation law in 2d CFT derived in (23.53). Comparing the coefficient of the anomalous term, we see

$$c = \frac{3\ell}{2G_N} \tag{24.11}$$

where we've reinserted a factor of G_N (previously set to 1) by dimensional analysis.

This is called the *Brown-Henneaux central charge*, after Brown and Henneaux who computed it way back in 1987 – long before AdS/CFT, and even well before the discovery of the BTZ black hole or the Brown-York stress tensor. They used a different method, based on conserved charges, which directly produces the Virasoro algebra (23.56) as the asymptotic symmetry group. As far as I know, they did not recognize the relation to the 2d conformal group.

24.3 Casimir energy on the circle

Recall the metric of the 3d black hole (BTZ):

$$ds^2 = -(\frac{r^2}{\ell^2} - 8M)dt^2 + \frac{dr^2}{r^2/\ell^2 - 8M} + r^2 d\phi^2 \tag{24.12}$$

It is up to us whether to identity $\phi \sim \phi + 2\pi$ (since there is no conical defect even if we leave $\phi \in [-\infty, \infty]$). The BTZ black hole is the solution with $\phi \sim \phi + 2\pi$. The boundary of this spacetime is the Lorentzian (t, ϕ) cylinder, so this is dual to the CFT on a cylinder.

In an exercise in a previous lecture you computed the energy of this solution, and found $E = M$.

Global AdS can also be written in the form (24.12), by choosing $M = -\frac{\ell}{8}$. Therefore the gravitational energy of the groundstate on the cylinder is

$$E_{vac} = -\frac{\ell}{8} = -\frac{c}{12} \ , \tag{24.13}$$

where c take the Brown-Henneaux value (24.11).

This is equal to the Casimir energy of a 2d CFT (23.63). It was actually guaranteed to agree once we found the transformation law (24.10) agrees with CFT, because the finite version of this infinitessimal transformation must be the Schwarzian derivative, on the gravity side just as it was in CFT.

25 Thermodynamics of 2d CFT

In this lecture we will discuss the thermodynamics of 2d CFT, including the torus partition function and the famous Cardy formula for CFT entropy. Later we'll come back to holography, but this note is entirely in CFT.

The partition function of a theory at finite temperature is

$$Z = \text{ Tr } e^{-\beta H} = \sum_{\text{states}} e^{-\beta E} \ . \tag{25.1}$$

This is a sum over states in some Hilbert space; that Hilbert space depends on our choice of the space in which the theory lives, *i.e.,* our choice of boundary conditions on the fields. If we choose space to be a circle, then

$$Z = \sum_{\text{states}} e^{-\beta E_{cyl}} = \sum_{\text{states}} e^{-\beta(\Delta - \frac{c}{12})} \ , \tag{25.2}$$

where Δ is the scaling dimension of the state. The second equality comes from (23.64), and assumes $L_{cyl} = 2\pi$.

As usual, the trace (25.1) is equal to a path integral in periodic imaginary time, with period β. Since space is also periodic, this is a path integral on a torus, with a 'space' identification and a 'thermal' identification:

$$w \sim w + L \sim w + i\beta \ . \tag{25.3}$$

$Z(\beta)$ is equal to the Euclidean path integral on the torus (25.3).

Our aim is to compute this path integral at high temperature. First we discuss general properties of CFT on a torus in the next couple subsections.

25.1 A first look at the S transformation

Consider a 2d QFT (not necessarily conformal yet) on a Euclidean torus. The most general torus is specified by two lattice vectors $\vec{v}_1, \vec{v}_2$ on the (t_E, ϕ) plane, meaning that we identify all points related by

$$(t_E, \phi) \sim (t_E, \phi) + m\vec{v}_1 + n\vec{v}_2, \qquad m, n \in \mathbf{Z} \ . \tag{25.4}$$

If the theory is rotationally invariant (*i.e.,* if its Lorentzian counterpart is Lorentz invariant), then we may w.l.o.g. rotate $\vec{v}_1$ to lie on the t_E axis. If the theory is scale invariant, then we can also w.l.o.g. set its length to $\vec{v}_2 = (0, 2\pi)$. Thus in a conformal field theory, we are led to consider the theory on the torus

$$(t_E, \phi) \sim (t_E + \beta, \phi + \theta) \sim (t_E, \phi + 2\pi) \ , \tag{25.5}$$

where β and θ are arbitrary real numbers.

In the complex coordinate $z = \frac{1}{2\pi}(\phi + it_E)$, the torus is specified by a complex number

$$\tau = \frac{1}{2\pi}(\beta + i\theta) \ . \tag{25.6}$$

The identifications are

$$z \sim z + 1 \sim z + \tau \ . \tag{25.7}$$

The path integral on this torus will be denoted $Z(\tau, \bar{\tau})$. τ is called the modulus, or complex structure modulus of the torus.

Converting to trace

The path integral on this torus can be converted to operator language by declaring that ϕ is 'space' and t_E is 'time.' Then we take states on a constant-t_E surface and evolve them along the vector $\beta\partial_{t_E} + i\theta\partial_\phi$. Thus the path integral can be rewritten as

$$Z(\tau, \bar{\tau}) = \ \mathrm{Tr}\ _{\mathcal{H}(0,2\pi)} e^{-\beta H + i\theta J} \tag{25.8}$$

where J is the angular momentum (which, by convention, generates motion along $-\partial_\phi$, hence the sign flip). In the subscript, we have noted explicitly what Hilbert space to trace over: it is the Hilbert space of states defined on the spatial slice where $\phi \in [0, 2\pi]$ with t_E held fixed. That is, the fields obey the boundary condition $X(t_E, \phi) = X(t_E, \phi + 2\pi)$. This is not standard notation; usually people just write ' Tr ', with $\mathcal{H}(0, 2\pi)$ implicit.

Large conformal transformations

We already know how conformal transformations act on the plane, as holomorphic/anti-holomorphic coordinate changes. Now we want to examine conformal symmetry on the torus. All of the usual conformal transformations $z \to z + \epsilon(z)$ are still symmetries, but there are also new, 'large' conformal transformations, which cannot be continuously connected to the identity.

The S transformation

One way to see this is to re-slice the torus path integral in a different way. We'll start with an intuitive explanation in the simplest case $\theta = 0$, then come back to the general story below. The torus is Euclidean, so we are free to switch the roles of t_E and ϕ when we construct the trace by declaring t_E is 'space' and ϕ is 'time'. Then following the usual logic, we find

$$Z(\tau, \bar{\tau}) = \text{ Tr }_{\mathcal{H}(\beta,0)} e^{-2\pi J} . \tag{25.9}$$

Therefore we have written the same path integral in two different ways (25.8) and (25.9),

$$Z(\tau, \bar{\tau}) = \text{ Tr }_{\mathcal{H}(0,2\pi)} e^{-\beta H} = \text{ Tr }_{\mathcal{H}(\beta,0)} e^{-2\pi J} \ . \tag{25.10}$$

This is true in any respectable quantum field theory. In a general (non-conformal) QFT, the Hilbert spaces in these two expressions are not the same. The first is the Hilbert space on the circle $\phi \sim \phi + 2\pi$, and the second is the Hilbert space on the circle $t_E \sim t_E + \beta$. Now, in a scale-invariant theory, these are related by a rotation by 90° followed by a rescaling by $2\pi/\beta$. Thus, in a CFT,

$$\text{Tr }_{\mathcal{H}(\beta,0)} e^{-2\pi J} = \text{ Tr }_{\mathcal{H}(0,2\pi)} e^{-\frac{4\pi^2}{\beta} H} \ . \tag{25.11}$$

Here the rotation by 90° sends $J \to H$, and the rescaling inserts the factor of $\frac{2\pi}{\beta}$.

It follows that

$$Z(\beta) = \text{ Tr } e^{-\beta H} = Z\left(\frac{4\pi^2}{\beta}\right) \ . \tag{25.12}$$

This is called the S transformation. It came from a large conformal transformation that swapped space and Euclidean time, and it relates the thermodynamics at high temperatures to the thermodynamics at low temperatures.

25.2 $SL(2,Z)$ transformations

The S transformation is one of an infinite group of large conformal transformations on the torus. These come from all the different ways of slicing the torus. We can think of these as ways of choosing the fundamental domain for the torus on the z-plane. The usual fundamental domain is the tilted rectangle with

$$z \sim z + 1 \sim z + \tau \ . \tag{25.13}$$

But we can instead choose the even-more-tilted rectangle

$$w \sim w + 1 \sim w + \tau + 1 \ . \tag{25.14}$$

This is precisely the same torus, since the lattice $m + n\tau$ with $m, n \in \mathbf{Z}$ is unchanged. From the point of view of the z coordinate, the w coordinate system 'winds' around the torus:

$$FIGURE: Twotori, windingcoords \tag{25.15}$$

This coordinate transformation is called 'T'. It acts on the modulus as

$$T \ : \quad \tau \to \tau + 1 \ . \tag{25.16}$$

The winding means that this coordinate transformation cannot be continuously deformed to the identity, *i.e.,* there is no infinitessimal version.

Every theory, conformal or not, is invariant under T:

$$Z(\tau + 1, \bar\tau + 1) = Z(\tau, \bar\tau) \ . \tag{25.17}$$

This is because we haven't actually done anything, we've just rewritten the same torus path integral in a different coordinate system. We can also check explicitly from the trace formula:

$$Z(\tau + 1, \bar\tau + 1) = \ \mathrm{Tr} \ e^{-\beta H + i(\theta + 2\pi)J} \ . \tag{25.18}$$

For a theory with only bosons, this is equal to $Z(\tau, \bar\tau)$ since angular momentum is integer quantized. In a theory with fermions, we have to be more careful about imposing boundary conditions on the fermions, but in the end it still works.

There are in fact an infinite number of ways to slice the torus. The general choice of lattice vectors v_1', v_2' that generate the same lattice is

$$\begin{pmatrix} v_1' \\ v_2' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} , \tag{25.19}$$

where

$$a, b, c, d \in \mathbf{Z} \qquad \text{and} \qquad ad - bc = 1 \ . \tag{25.20}$$

The matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is therefore an element of the group $SL(2, Z)$— *i.e.*, , 2x2 matrices with integer elements and unit determinant. This re-slicing maps

$$\tau \to \frac{a\tau + b}{c\tau + d} \ . \tag{25.21}$$

$SL(2, Z)$ is generated by the S and T transformations. The S transformation is the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, which acts as

$$S : \quad \tau \to -1/\tau \ . \tag{25.22}$$

Above, we discussed the S transformation with zero angular potential, $\tau = i\beta/(2\pi)$. In this case $S : \beta \to \frac{4\pi^2}{\beta}$ as claimed above.

Unlike T, the other $SL(2, Z)$ transformations are *not* symmetries of a general QFT. This is because the new space circle, which defines the Hilbert space, is not the same as the old space circle. It is only in a scale-invariant theory that we can rescale these two circles and see that they have the same Hilbert space. Thus in conformal field theory, $SL(2, Z)$ is a symmetry:

$$Z(\tau, \bar{\tau}) = Z\left(\frac{a\tau + b}{c\tau + d}, \frac{a\bar{\tau} + b}{c\bar{\tau} + d}\right) \ . \tag{25.23}$$

25.3 Thermodynamics at high temperature

(We will restrict to the case $\theta = 0$, but angular potential can be included at the expense of slightly more complicated formulas.) First let's calculate the partition function at very low temperature:

$$Z(\beta) = \sum_{\text{states}} e^{-\beta E} \approx e^{-\beta E_{vac}} \quad (\beta \to \infty) \ . \tag{25.24}$$

This is simply the statement that at very low temperature, the vacuum state dominates the canonical ensemble. Above we found the Casimir energy $E_{vac} = -\frac{c}{12}$, so

$$Z(\beta) \approx e^{\frac{\beta c}{12}} \quad (\beta \to \infty) \ . \tag{25.25}$$

Now, what about high temperatures? Modular invariance requires

$$Z(\beta) = Z(4\pi^2/\beta) \ , \tag{25.26}$$

so we can repeat the derivation at low temperature using the S-transformed partition function. Replacing $\beta \to 4\pi^2/\beta$ in (25.25) gives

$$Z(\beta) \approx e^{\frac{\pi^2 c}{3\beta}} \quad (\beta \to 0) \ . \tag{25.27}$$

The corresponding free energy, defined by $Z = e^{-\beta F}$, is

$$F = -\frac{\pi^2}{3} c T^2 \left(\frac{V}{2\pi}\right) \ , \tag{25.28}$$

where we've written the formula for a general 'volume', which we've been assuming is 2π. This is a remarkable formula. The scaling with the temperature in the free energy is fixed by dimensional analysis in a CFT, but the coefficient is not. The high-temperature free energy dominated by very heavy states, and is usually impossible to calculate in a strongly interacting QFT. But in this case, modular invariance relates it to the Casimir energy of the vacuum state.

There are two immediate consequences. First, this confirms once again that c should be interpreted as a measure of degrees of freedom. Second, it fixes the asymptotic density of states. To see this, write the result as

$$\sum_{states} e^{-\beta E} \approx \int dE \rho(E) e^{-\beta E} \approx e^{\pi^2 c/(3\beta)} \qquad (\beta \to 0) \ . \tag{25.29}$$

The first sum is over all states of the theory; the integral is over energies, with $\rho(E)$ the density of states.

The rhs is very singular as $\beta \to 0$. Each individual term on the lhs is regular at $\beta = 0$, so the singularity can only come from the infinite sum. The strength of the divergence must be related to the growth of $\rho(E)$ as $E \to \infty$.

To make this more quantitative, we can use standard thermodynamic formulae. The thermodynamic entropy and energy are

$$S = (1 - \beta\partial_\beta) \log Z = \frac{2\pi^2 c}{3\beta} \tag{25.30}$$

$$E(\beta) = -\partial_\beta \log Z = \frac{\pi^2 c}{3\beta^2} \ . \tag{25.31}$$

Putting these together, we have

$$S(E) = 2\pi\sqrt{\frac{c}{3}E} \ . \tag{25.32}$$

As usual in stat mech, this formula should give the density of states via

$$\rho(E) = e^{S(E)} \ . \tag{25.33}$$

As a check, let's plug this into the partition function and see that it reproduces the expected singularity as $\beta \to 0$:

$$Z(\beta) \approx \int dE \rho(E) e^{-\beta E} \tag{25.34}$$

$$\approx \int dE \exp\left(S(E) - \beta E\right) \tag{25.35}$$

$$\approx \exp\left(S(E_*) - \beta E_*\right) \tag{25.36}$$

where E_* is the saddlepoint value, defined by

$$S'(E_*) - \beta = 0 \ . \tag{25.37}$$

This saddlepoint is just the usual thermodynamic value of the energy. Using (25.32), finding the saddlepoint, and plugging in, we find

$$Z(\beta) \approx \exp\left(\frac{\pi^2 c}{3\beta}\right) \tag{25.38}$$

as claimed. (This is not a surprise; the whole point of thermodynamics and Legendre transforms is to solve this saddlepoint equation for you.)

The entropy formula (25.32) is often called the Cardy formula. It applies to any CFT as $E \to \infty$.

Exercise: Cardy formula with angular potential
Derive the asymptotic density of states $\rho(E, J)$ by applying modular invariance to the partition function including angular potential, $Z(\tau, \bar{\tau}) = \text{ Tr } e^{-\beta H + i\theta J}$.

---

26 Black hole microstate counting

26.1 From the Cardy formula

The BTZ metric,

$$ds^2 = -\left(\frac{r^2}{\ell^2} - 8M\right) dt^2 + \frac{dr^2}{r^2/\ell^2 - 8M} + r^2 d\phi^2 \tag{26.1}$$

with $\phi \sim \phi + 2\pi$, has energy $E = M$ and entropy

$$\begin{aligned} S &= \frac{\text{area}}{4} \\ &= \frac{1}{4} 2\pi \rho_+ \\ &= \frac{\pi}{2}\sqrt{8M\ell} \\ &= 2\pi\sqrt{\frac{c}{3}E} \ . \end{aligned} \tag{26.2}$$

In the last line we used the Brown-Henneaux central charge, $c = 3\ell/2$ (with $G_N = 1$).

This exactly matches the Cardy formula, (25.32).* This match was found by Strominger in 1997. What is important about this result is that the Cardy formula *counts microstates in statistical mechanics.* They are microstates in the dual CFT, but by holographic duality, they must be microstates of quantum gravity in AdS_3 as well. We've just counted them without actually enumerating them, but a great deal of progress has been made enumerating them as well.

26.2 Strominger-Vafa

Historically, the first black hole microstate counting was a string theory calculation by Strominger and Vafa. It gave the same answer as (26.2). This calculation was important because it laid to rest any final doubts about whether black hole thermodynamics is *really* thermodynamics (*i.e.,* coming from stat mech) or just a mysterious analogy. In quantum gravity, black hole entropy counts microstates:

$$S_{BH}(E) \approx \log \rho_{QG}(E) \ , \tag{26.3}$$

where ρ_{QG} is the density of states in quantum gravity. This is an extraordinary window from low-energy physics into the theory of quantum gravity well above the Planck scale.

*Actually the Cardy formula was for $E \to \infty$, whereas the black hole formula is for $E > 0$. It is possible to get the match at all energies but requires further input from string theory, or some further assumptions about the spectrum of the CFT.