Lecture Notes

Quantum Gravity: Black Holes and the AdS/CFT Correspondence

Noppadol Mekareeya

INFN, sezione di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy

E-mail: n.mekareeya@gmail.com

Abstract: These are the notes of a 42-hour graduate course on quantum aspects of black holes and the AdS/CFT correspondence, for students with a first course in general relativity and in quantum field theory and no previous exposure to string theory or supersymmetry.

- Part I develops black hole thermodynamics: the laws of black hole mechanics, Bekenstein's entropy and the generalized second law, quantum fields in curved spacetime, Hawking radiation and gray-body factors, the Unruh effect and the thermofield double, and the Euclidean approach to gravitational thermodynamics.
- Part II is an introduction to the AdS/CFT correspondence: conformal field theory, anti-de Sitter space, the large-N limit, the holographic dictionary and correlation functions, holographic anomalies, a self-contained crash course on strings and D-branes, the Maldacena limit of D3-branes, and tests of the duality.
- Part III returns to black holes with holography in hand: the Hawking–Page transition, the two notions of entropy and the Page curve, black hole entropy as entanglement entropy, and the Ryu–Takayanagi formula.

All derivations are carried out in detail, so that the notes are suitable for self-study. The lectures will not cover everything here: the material not treated in class is left to the reader.

The latest version of these notes can be found at https://nmekareeya.github.io/QG.

Last updated: September 12, 2026

About these notes

These notes accompany a 42-hour course (21 lectures of two hours each) on quantum aspects of black holes and the AdS/CFT correspondence. The prerequisites are a first course in general relativity and in quantum field theory; no previous exposure to string theory or supersymmetry is assumed.

The course has a deliberate "sandwich" structure. **Part I** (Chapters 1–7) develops black hole thermodynamics: the classical laws of black hole mechanics, Bekenstein's entropy, Hawking radiation, the Unruh effect, and the Euclidean approach to gravitational thermodynamics. **Part II** (Chapters 8–17) is an introduction to the AdS/CFT correspondence, starting from conformal field theory, anti-de Sitter geometry and the large-N limit, formulating the holographic dictionary, and culminating in the D3-brane derivation of the duality between $\mathcal{N} = 4$ super Yang–Mills theory and Type IIB strings on $\mathrm{AdS}_5 \times S^5$. **Part III** (Chapters 18–21) returns to black holes equipped with holography: the Hawking–Page transition and its dual interpretation, von Neumann entropy and the Page curve, black hole entropy as entanglement entropy, and the Ryu–Takayanagi formula.

The notes are meant to be self-contained and suitable for self-study: derivations are carried out in detail. They contain more than can be covered in 42 hours. The lectures will follow the notes but will not treat every section; the material not covered in class is left to the reader.[1]

Sources and further reading. Part I and Part III are largely based on Witten's *Introduction to Black Hole Thermodynamics* [1] (Chapters 1–7 on its sections 1–6, Chapters 18–21 on sections 6.5–9); Part II on Zaffaroni's *Introduction to the AdS/CFT correspondence* [2], sections 1–3. Original papers are cited where each result is derived. For background beyond the notes: Wald [3] for general relativity; Birrell–Davies [4] and Wald [5] for quantum fields in curved spacetime; Di Francesco–Mathieu–Sénéchal [6] and Rychkov [7] for conformal field theory; Tong's lecture notes [8], cited by section in Chapter 15, and Polchinski [9, 10] for string theory; the review of Aharony et al. [11] for AdS/CFT; Skenderis's lecture notes [12] for holographic renormalization; Nielsen–Chuang [13] for quantum information; Rangamani–Takayanagi [14] for holographic entanglement entropy.

Conventions. We use the mostly-plus signature $(-,+,+,+)$ and units $\hbar = c = k_B = 1$, keeping Newton's constant G explicit. Note that G is dimensionful, $[G] = (\text{length})^{D-2}$ in D spacetime dimensions; expansions "in G" always refer to a dimensionless combination such as GE^2, G/A, or G_5/L^3, as discussed in Remark 2.1 of Chapter 2. Greek indices $\mu, \nu, \ldots$ are spacetime indices. The Riemann tensor is defined by $[\nabla_\mu, \nabla_\nu]V^\rho = R^\rho{}_{\sigma\mu\nu}V^\sigma$ and the Ricci tensor by $R_{\mu\nu} = R^\rho{}_{\mu\rho\nu}$, so that the sphere has positive curvature.

[1] A star $\star$ marks material planned for self-study, in a section title or at the start of a proof; more may be added as the course progresses.

Contents

Part I

Black Hole Thermodynamics

Chapter 1

Black Hole Geometry and the Laws of Black Hole Mechanics

This chapter assembles the classical general relativity that the course builds on: the Schwarzschild solution and its global structure, Killing horizons and surface gravity, the area theorem, the essentials of charged and rotating black holes, and the four laws of black hole mechanics. It ends with the analogy between those laws and thermodynamics, and with the puzzle that a classical black hole has temperature zero, which the rest of Part I resolves. Section 1.1 is a recap of material the reader knows; it fixes notation and the figures used later.

1.1 Schwarzschild, Eddington–Finkelstein, Kruskal

- *The solution.* The unique spherically symmetric solution of $R_{\mu\nu} = 0$ is

$$\mathrm{d}s^2 \;=\; -f(r)\,\mathrm{d}t^2 \;+\; \frac{\mathrm{d}r^2}{f(r)} \;+\; r^2\,\mathrm{d}\Omega^2, \qquad f(r) \;=\; 1 - \frac{2GM}{r}, \qquad r_S \;=\; 2GM, \tag{1.1}$$

  with r the areal radius, M the mass read off from the Newtonian potential at large r, and r_S the Schwarzschild radius where the coordinates degenerate.

- *Birkhoff's theorem.* Any spherically symmetric vacuum solution is locally (1.1). Hence the exterior of any spherically symmetric star, even a collapsing one, is exactly Schwarzschild, and collapse spacetimes are built by gluing a matter interior to a portion of (1.1).

- *Nothing is singular at r_S.* The invariant $R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} = 48G^2M^2/r^6$ is finite at $r = r_S$ and decreases with M, and diverges only at $r = 0$: the horizon is a coordinate singularity, the origin a curvature singularity.

- *Tortoise and null coordinates.* Radial null rays obey $\mathrm{d}t/\mathrm{d}r = \pm 1/f$, which is integrated by

$$r_* \;=\; \int \frac{\mathrm{d}r}{f(r)} \;=\; r + 2GM\,\ln\!\Big(\frac{r}{2GM} - 1\Big), \qquad v = t + r_*, \qquad u = t - r_*, \tag{1.2}$$

  so that the exterior is $-\infty < r_* < \infty$, the horizon sits at $r_* = -\infty$, and radial rays are $u = \text{const}$ (outgoing) or $v = \text{const}$ (ingoing).

- *Ingoing Eddington–Finkelstein coordinates.* In (v, r),

$$\mathrm{d}s^2 \;=\; -f(r)\,\mathrm{d}v^2 \;+\; 2\,\mathrm{d}v\,\mathrm{d}r \;+\; r^2\mathrm{d}\Omega^2, \tag{1.3}$$

  which is smooth for all $0 < r < \infty$. Its radial null directions are $\mathrm{d}v = 0$ (ingoing) and $\mathrm{d}r/\mathrm{d}v = f/2$ (outgoing): outside, outgoing rays move to larger r; at r_S they hover; inside,

$f < 0$ and both branches move to smaller r. Inside the horizon r is a timelike coordinate, $r = 0$ is a moment rather than a place, and $r = r_S$ is a one-way membrane, the event horizon.

- *Kruskal–Szekeres coordinates.* The exterior in double-null form is $\mathrm{d}s^2 = -f\,\mathrm{d}u\,\mathrm{d}v + r^2\mathrm{d}\Omega^2$, degenerate at the horizon because $f \to 0$ while $u \to \infty$. Compress the null coordinates exponentially,

$$U = -\,\mathrm{e}^{-u/4GM}, \qquad V = \mathrm{e}^{v/4GM} \qquad (U < 0 < V \text{ in the exterior}). \tag{1.4}$$

Then

$$\begin{aligned} \mathrm{d}U\,\mathrm{d}V &= \frac{\mathrm{e}^{(v-u)/4GM}}{16\,G^2M^2}\,\mathrm{d}u\,\mathrm{d}v = \frac{\mathrm{e}^{r_*/2GM}}{16\,G^2M^2}\,\mathrm{d}u\,\mathrm{d}v, \\ \mathrm{e}^{r_*/2GM} &= \mathrm{e}^{r/2GM}\left(\frac{r}{2GM} - 1\right) = \mathrm{e}^{r/2GM}\,\frac{r}{2GM}\,f(r), \end{aligned} \tag{1.5}$$

by (1.2): the factor f, which vanishes linearly at the horizon, appears explicitly and cancels against the f in $-f\,\mathrm{d}u\,\mathrm{d}v$:

$$\boxed{\mathrm{d}s^2 = -\frac{32\,G^3M^3}{r}\,\mathrm{e}^{-r/2GM}\,\mathrm{d}U\,\mathrm{d}V + r^2\,\mathrm{d}\Omega^2} \qquad UV = \left(1 - \frac{r}{2GM}\right)\mathrm{e}^{r/2GM}. \tag{1.6}$$

The right-hand side of the second relation decreases monotonically from 1 at $r = 0$ to $-\infty$ as $r \to \infty$, so $r(UV)$ is smooth on $UV < 1$, and the metric is regular there. The scale $4GM$ in (1.4) is forced by the cancellation; Section 1.2 identifies it as $1/\kappa$.

- *The four regions* (Figure 1.1), with $T = (U+V)/2$, $X = (V-U)/2$: I ($U < 0 < V$), the original exterior, where the time translation acts as a boost; II ($U, V > 0$), the black hole interior, ending on the future singularity $UV = 1$; III ($V < 0 < U$), a second exterior isometric to I and causally disconnected from it; IV ($U, V < 0$), the white hole. The horizons are $U = 0$ (future, $\mathcal{H}^+$) and $V = 0$ (past, $\mathcal{H}^-$), meeting at the bifurcation sphere $U = V = 0$ of radius r_S.

- *Eternal versus collapse.* The two-sided geometry is not produced by any collapse, but it is the black-hole analogue of an equilibrium state and returns in Chapters 6, 7 and Part III. In a collapse (Figure 1.2) regions III and IV are replaced by the interior of the star.

Penrose diagrams. A conformal compactification, $\mathrm{d}s^2 = \Omega^{-2}\mathrm{d}\tilde{s}^2$ with $\mathrm{d}\tilde{s}^2$ regular on a compact region, preserves null geodesics and causal relations, so the diagram of $\mathrm{d}\tilde{s}^2$ displays the causal structure of the whole spacetime on a finite page, with light rays at 45°. For Minkowski space, compactifying the null coordinates $u = t - r$, $v = t + r$ by $u = \tan\tilde{u}$, $v = \tan\tilde{v}$ produces a triangle whose boundaries are past and future null infinity $\mathscr{I}^{\mp}$, where all light rays begin and end, spatial infinity i^0, and past and future timelike infinity $i^{\mp}$, where timelike curves begin and end. For Schwarzschild one applies the same compactification, $U = \tan\tilde{U}$, $V = \tan\tilde{V}$, to the Kruskal coordinates. The physically relevant diagram is that of collapse (Figure 1.2), with two features to keep in mind.

- *The horizon is teleological.* The event horizon $\mathcal{H}^+$ is the boundary between the events from which a signal can still reach future null infinity and those from which it cannot. Deciding on which side an event lies requires knowing the entire future of the spacetime, so the horizon cannot be located by local measurements today: it starts growing at the centre of the star before the collapse is complete, anticipating the infall.

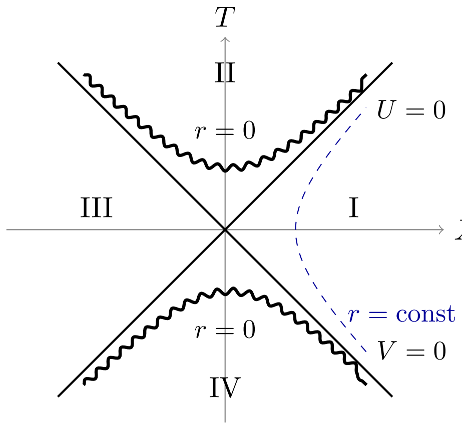


Figure 1.1: The Kruskal–Szekeres extension (1.6) in the (T, X) plane; each point is a two-sphere and light rays travel at 45°. The horizons $U = 0$ and $V = 0$ cross at the bifurcation sphere; the singularities $r = 0$ are the wavy hyperbolae; curves of constant $r > r_S$ are timelike hyperbolae in regions I and III. Its Penrose diagram is obtained by bringing infinity to finite distance with $U = \tan \tilde{U}$, $V = \tan \tilde{V}$.

- *The eternal diagram is an idealization.* Inside the star the geometry is regular, and the vacuum description holds only outside its surface. The two-sided geometry idealizes the collapse geometry the way an infinite thermal bath idealizes a large finite one.

1.2 Killing horizons and surface gravity

The Schwarzschild exterior is static: $\xi = \partial_t$ is a Killing vector with norm $\xi^\mu \xi_\mu = -f(r)$, which vanishes at the horizon. The horizon is a null hypersurface to which ξ is tangent and on which it is null.

Definition 1.1 (Killing horizon and surface gravity)**.** A *Killing horizon* is a null hypersurface $\mathcal{H}$ whose null generators are orbits of a Killing vector field ξ. Since $\xi^\mu \xi_\mu$ is constant (zero) along $\mathcal{H}$, its gradient is normal to $\mathcal{H}$, hence proportional to ξ_μ:

$$\nabla^\mu \left(\xi^\nu \xi_\nu\right)\Big|_{\mathcal{H}} \;=\; -\,2\kappa\, \xi^\mu, \qquad \text{equivalently} \qquad \xi^\nu \nabla_\nu \xi^\mu\Big|_{\mathcal{H}} \;=\; \kappa\, \xi^\mu, \tag{1.7}$$

which defines the *surface gravity* κ; the second form follows from Killing's equation and says that ξ is geodesic on $\mathcal{H}$ in a non-affine parametrization.

For Schwarzschild the event horizon is the Killing horizon of $\partial_t = \frac{1}{4GM}(V\partial_V - U\partial_U)$, the boost generator in Kruskal coordinates. We compute κ for any metric of the form (1.1) with f vanishing linearly at $r = r_h$.

1. A static observer at fixed (r, θ, ϕ) has four-velocity $u^\mu = \xi^\mu/\sqrt{f}$ and proper acceleration $a^\mu = u^\nu \nabla_\nu u^\mu$; only $\Gamma^r_{tt} = \frac{1}{2} f f'$ contributes, so

$$a^r \;=\; \frac{f'(r)}{2}, \qquad a \;\equiv\; \sqrt{g_{rr}\,(a^r)^2} \;=\; \frac{f'(r)}{2\sqrt{f(r)}}, \tag{1.8}$$

   which diverges at the horizon: no rocket can hover there.

2. The force applied from infinity, say through a string, is redshifted by $\sqrt{f}$, the norm of ξ. The redshifted acceleration stays finite and is the surface gravity,

$$\kappa \;=\; \lim_{r \to r_h} \sqrt{f(r)}\; a(r) \;=\; \frac{f'(r_h)}{2}, \tag{1.9}$$

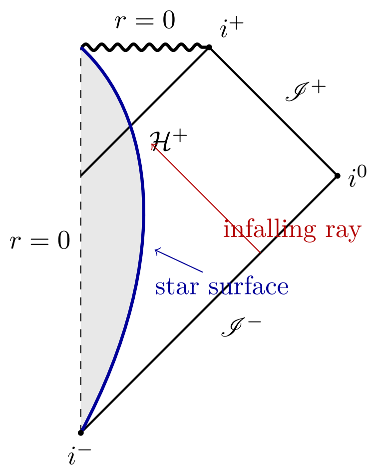


Figure 1.2: Penrose diagram of gravitational collapse. The shaded region is the interior of the star; outside it the geometry is exactly Schwarzschild by Birkhoff's theorem. There is a single exterior, a single future horizon and a single spacelike singularity: no white hole and no second universe. The horizon is generated by the outgoing rays that just fail to escape.

in agreement with (1.7).

For a Schwarzschild black hole of mass M, the surface gravity is

$$\kappa = \frac{f'(r_S)}{2} = \frac{GM}{r_S^2} = \frac{1}{4GM}. \tag{1.10}$$

It is the acceleration, measured from infinity, needed to hover at the horizon; equivalently, it is the rate at which Killing time exponentiates into affine parameter along the horizon, $U \propto -\mathrm{e}^{-\kappa u}$ in (1.4).

Two remarks.

- *Big black holes are mild.* κ decreases with M. The exponential relation between the clock of an infalling observer and that of a distant one, with rate κ, is the origin of the thermal nature of black holes (Chapters 4 and 6).

- *Near-horizon geometry is Rindler space.* With $f \simeq 2\kappa(r - r_h)$ and the proper distance $\rho = \sqrt{2(r - r_h)/\kappa}$, so that $f = \kappa^2\rho^2$,

$$ds^2 \simeq -\kappa^2\rho^2\,dt^2 + d\rho^2 + r_h^2\,d\Omega^2, \tag{1.11}$$

flat space in the coordinates of uniformly accelerated observers. Every non-extremal Killing horizon looks like this up close, which is why the Unruh effect of Chapter 6 captures the physics of Hawking radiation.

1.3 The area theorem

Theorem 1.1 (Hawking's area theorem [15])**.** *Let $(\mathcal{M}, g)$ satisfy the Einstein equations with matter obeying the null energy condition $T_{\mu\nu}k^\mu k^\nu \geq 0$ for all null k^μ, and suppose cosmic censorship holds (strong asymptotic predictability). Then the area of the event horizon cross-section is non-decreasing toward the future,*

$$\frac{dA}{dt} \geq 0. \tag{1.12}$$

⋆ *Self-study.* The proof has three ingredients: the kinematics of null congruences, the Raychaudhuri equation, and a global argument about horizon generators.

1. *Expansion of the generators.* The horizon is ruled by null geodesics with affine tangent k^μ. Decompose $B_{\mu\nu} = \nabla_\nu k_\mu$, restricted to the two-plane transverse to k, into trace, symmetric traceless and antisymmetric parts,

$$B_{\mu\nu}\big|_\perp = \tfrac{1}{2}\theta\, h_{\mu\nu} + \sigma_{\mu\nu} + \omega_{\mu\nu}, \qquad \frac{1}{\mathrm{d}A}\frac{\mathrm{d}(\mathrm{d}A)}{\mathrm{d}\lambda} = \theta, \tag{1.13}$$

the expansion θ, the shear σ and the twist ω. The twist vanishes for the generators of any null hypersurface, which are hypersurface-orthogonal.

2. *Raychaudhuri equation.* Differentiating $B_{\mu\nu}$ along the flow and commuting covariant derivatives,

$$\begin{aligned} k^\lambda \nabla_\lambda B_{\mu\nu} &= k^\lambda \nabla_\lambda \nabla_\nu k_\mu = k^\lambda \nabla_\nu \nabla_\lambda k_\mu + k^\lambda \left[\nabla_\lambda, \nabla_\nu\right] k_\mu \\ &= \nabla_\nu\Big(k^\lambda \nabla_\lambda k_\mu\Big) - (\nabla_\nu k^\lambda)(\nabla_\lambda k_\mu) + R_{\mu\sigma\lambda\nu}\, k^\sigma k^\lambda = -\, B_{\mu\lambda} B^\lambda{}_\nu + R_{\mu\sigma\lambda\nu}\, k^\sigma k^\lambda, \end{aligned} \tag{1.14}$$

the geodesic equation killing the first term of the second line. The transverse trace with $\omega = 0$ is

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{\theta^2}{2} - \sigma_{\mu\nu}\sigma^{\mu\nu} - R_{\mu\nu}\, k^\mu k^\nu. \tag{1.15}$$

3. *Focusing.* By the Einstein equations, $R_{\mu\nu}k^\mu k^\nu = 8\pi G\, T_{\mu\nu}k^\mu k^\nu \geq 0$ for null k, so every term on the right of (1.15) is non-positive and

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} \leq -\frac{\theta^2}{2} \qquad \Longrightarrow \qquad \frac{1}{\theta(\lambda)} \geq \frac{1}{\theta_0} + \frac{\lambda}{2}: \tag{1.16}$$

if $\theta_0 < 0$ anywhere, $\theta \to -\infty$ within affine parameter $2/|\theta_0|$. Neighbouring generators cross at a caustic.

4. *The global argument.* Suppose $\theta < 0$ at a point of the horizon. Cosmic censorship lets the generators be followed to the caustic. But by a theorem of Penrose horizon generators have no future endpoints, whereas beyond a caustic they would enter the region from which signals reach infinity, i.e. leave the horizon. Hence $\theta \geq 0$ on the horizon; every area element is non-decreasing by (1.13), generators can join the horizon but never leave it, and the total area cannot decrease. □

Example 1.1 (Black hole merger)**.** Two Schwarzschild black holes of masses M_1, M_2 merge into one of mass M. The area theorem demands $M^2 \geq M_1^2 + M_2^2$, so for $M_1 = M_2$ at most $1 - 1/\sqrt{2} \approx 29\%$ of the mass can be radiated in gravitational waves; the LIGO events radiate a few percent.

Remark 1.1. The null energy condition holds for classical matter but is violated by quantum fields: near a black hole $\langle T_{\mu\nu}\rangle k^\mu k^\nu$ can be negative. This is how Hawking evaporation evades the area theorem and shrinks the horizon (Chapter 4); finding the quantum replacement of the theorem is the theme of Chapter 2.

1.4 Charged and rotating black holes in brief

The course works almost exclusively with Schwarzschild and its anti-de Sitter cousin, but the four laws refer to charge and angular momentum, so we record the essentials; derivations are in [3].

- *Reissner–Nordström.* The unique spherically symmetric solution of Einstein–Maxwell theory with charge Q is (1.1) with $f = 1 - 2GM/r + GQ^2/r^2$, whose zeros $r_\pm = GM \pm \sqrt{(GM)^2 - GQ^2}$ are an outer event horizon and an inner Cauchy horizon, both Killing horizons of ∂_t, with $\kappa = (r_+ - r_-)/2r_+^2$ and electrostatic potential $\Phi = Q/r_+$. Horizons require $GM^2 \geq Q^2$; equality, the *extremal* case, gives $r_+ = r_-$ and $\kappa = 0$.

- *Kerr.* The unique stationary rotating vacuum black hole has $a = J/M$ and horizons at $r_\pm = GM \pm \sqrt{(GM)^2 - a^2}$, extremal at $J = GM^2$ where again $\kappa = 0$. The horizon is the Killing horizon of the corotating vector

$$\begin{aligned} \chi &= \partial_t + \Omega_H\,\partial_\phi, & \Omega_H &= \frac{a}{r_+^2 + a^2}, \\ \kappa &= \frac{r_+ - r_-}{2\,(r_+^2 + a^2)}, & A &= 4\pi\,(r_+^2 + a^2) \;=\; 8\pi G M r_+. \end{aligned} \tag{1.17}$$

  Between the horizon and the surface where $g_{tt} = 0$ lies the *ergoregion*, in which ∂_t is spacelike and every observer is dragged around the hole.

- *The Penrose process.* The conserved energy $E = -p \cdot \partial_t$ of a particle is positive where ∂_t is timelike, but inside the ergoregion, where ∂_t is spacelike, a future-directed timelike p can have $E < 0$; such a particle cannot reach infinity, but it can fall into the hole. So send in a particle of energy $E_0 > 0$ and let it split inside the ergoregion into a fragment with $E_1 < 0$ that crosses the horizon and one with $E_2 = E_0 - E_1 > E_0$ that escapes: more energy comes out than went in, paid for by the hole, which loses $\delta M = E_1 < 0$ and, the fragment counter-rotating, $\delta J = L_1 < 0$. The extraction is bounded: at the horizon p and the generator χ of (1.17) are both future-directed causal vectors, whose inner product is non-positive, so $-p \cdot \chi = E - \Omega_H L \geq 0$ for the absorbed fragment and the hole changes by [3]

$$\delta M \;\geq\; \Omega_H\,\delta J. \tag{1.18}$$

- *Irreducible mass* [16]. Defining $M_{\rm irr}^2 \equiv A/16\pi G^2$, the Kerr mass decomposes as

$$M^2 \;=\; M_{\rm irr}^2 \;+\; \frac{J^2}{4\,G^2 M_{\rm irr}^2}, \tag{1.19}$$

  and differentiating, with $\Omega_H = J/(4G^2 M M_{\rm irr}^2)$, gives $\delta M - \Omega_H \delta J = \frac{M_{\rm irr}}{M}\big(1 - \frac{J^2}{4G^2 M_{\rm irr}^4}\big)\delta M_{\rm irr}$ with positive prefactor below extremality: the bound (1.18) is $\delta M_{\rm irr} \geq 0$, i.e. the area cannot decrease, with equality for reversible processes. The rotational term is the extractable part, up to $1 - 1/\sqrt{2} \approx 29\%$ of the mass of an extremal Kerr hole. A state function that only grows, and reversible versus irreversible processes: the structure of thermodynamics, before Bekenstein.

1.5 The four laws of black hole mechanics

Bardeen, Carter and Hawking [17] organized the classical properties of stationary black holes into four laws, numbered and phrased to mirror thermodynamics.

- *Zeroth law.* The surface gravity κ is constant over the event horizon of a stationary black hole. For a static spherical hole this is symmetry; the content is that it holds for a rapidly rotating Kerr hole, whose horizon is far from round. The proof of [17] uses the Einstein equations with the dominant energy condition; for a Killing horizon with a bifurcation surface, as in Kruskal, constancy follows purely geometrically. A quantity uniform over a system in equilibrium, like a temperature.

- *First law.* For Schwarzschild, $A = 16\pi G^2 M^2$ gives $\mathrm{d}M = \mathrm{d}A/32\pi G^2 M = (\kappa/8\pi G)\,\mathrm{d}A$ with $\kappa = 1/4GM$. The general statement, proved in [17] by comparing nearby stationary solutions of Einstein–Maxwell theory, is

$$\mathrm{d}M \;=\; \frac{\kappa}{8\pi G}\,\mathrm{d}A \;+\; \Omega_H\,\mathrm{d}J \;+\; \Phi\,\mathrm{d}Q, \tag{1.20}$$

  with work terms analogous to $-p\,\mathrm{d}V$ and $\mu\,\mathrm{d}N$. The Penrose bound (1.18) is $\kappa\,\mathrm{d}A/8\pi G \geq 0$, and differentiating (1.19) at fixed J and at fixed A reproduces $\kappa/8\pi G$ and Ω_H for Kerr.

- *Second law.* The area theorem, Theorem 1.1: $\mathrm{d}A \geq 0$ in any classical process. Unlike the zeroth and first laws, which compare stationary states, it is a dynamical statement valid in mergers and accretion.

- *Third law.* No physical process can reduce κ to zero in finite advanced time if the stress tensor is bounded and obeys the weak energy condition: one cannot spin a Kerr hole up to $a = GM$ or charge a Reissner–Nordström hole to $Q^2 = GM^2$ by finitely many operations. This mirrors the unattainability form of Nernst's theorem, not the Planck form $S \to 0$ as $T \to 0$: extremal holes have $\kappa = 0$ but large area. The extremal case is exactly where string theory counts microstates (Remark 2.2).

- *Smarr formula.* Under $r \to \lambda r$ at fixed G, $M \to \lambda M$, $A, J \to \lambda^2 A, \lambda^2 J$, $Q \to \lambda Q$, so Euler's theorem for the homogeneous function $M(A, J, Q)$ with the derivatives of (1.20) gives

$$M \;=\; \frac{\kappa A}{4\pi G} \;+\; 2\,\Omega_H J \;+\; \Phi\,Q, \tag{1.21}$$

  the analogue of the Euler relation $E = TS - pV + \mu N$; for Schwarzschild $\kappa A/4\pi G = M$ by inspection.

1.6 The analogy with thermodynamics, and a puzzle

Law	**Thermodynamics**	**Black hole mechanics**
Zeroth	T uniform at equilibrium	κ uniform on the horizon
First	$\mathrm{d}E = T\,\mathrm{d}S - p\,\mathrm{d}V + \mu\,\mathrm{d}N$	$\mathrm{d}M = \dfrac{\kappa}{8\pi G}\,\mathrm{d}A + \Omega_H\,\mathrm{d}J + \Phi\,\mathrm{d}Q$
Second	$\mathrm{d}S \geq 0$	$\mathrm{d}A \geq 0$
Third	$T = 0$ unattainable	$\kappa = 0$ unattainable
	$E,\ T,\ S$	$M,\ \alpha\kappa,\ \dfrac{A}{8\pi G\alpha}$

- *The dictionary is forced up to one constant.* Energy is mass, literally; temperature is $\alpha\kappa$ and entropy $A/8\pi G\alpha$ for some undetermined α. Every structural feature matches, including the Euler/Smarr relation.

- *The puzzle.* A classical black hole absorbs and never emits: in contact with a bath at any $T > 0$ it only absorbs, so its physical temperature is

$$T_{\mathrm{BH}}^{\mathrm{classical}} \;=\; 0 \tag{1.22}$$

  regardless of κ. Zero temperature with finite "entropy" $\propto A$ is incompatible with $\mathrm{d}M = T\,\mathrm{d}S$, and Bardeen, Carter and Hawking initially concluded that the analogy could not be taken literally.

- *The resolution needs two quantum inputs.* Bekenstein's argument (Chapter 2) that the second law forces an entropy $\propto A/G\hbar$, invisible classically since $\hbar \to 0$ sends $S \to \infty$ and $T \to 0$ at fixed M; and Hawking's discovery (Chapters 3–4) that quantum fields make the black hole radiate at

$$T_H = \frac{\kappa}{2\pi}, \tag{1.23}$$

  fixing $\alpha = 1/2\pi$ and $S = A/4G$. The laws of black hole mechanics are the laws of thermodynamics applied to an unfamiliar system.

Chapter 2

Bekenstein's Proposal and the Generalized Second Law

Chapter 1 ended with a paradox: black hole mechanics looks exactly like thermodynamics, yet a classical black hole has temperature zero. Following Bekenstein [18], we trace the fate of the second law when entropy falls behind a horizon and are led to $S = A/4G\hbar$, the generalized entropy, and the Generalized Second Law. Bekenstein's quantitative test, absorption of black-body radiation, succeeds at short wavelengths and fails at long ones, and the failure is a prediction: black holes must radiate. We close with the implied $\mathrm{e}^{A/4G}$ microstates and the no-hair theorem read as thermalization.

2.1 The tea-cup problem

The Second Law states that in every process that occurs in practice the total thermodynamic entropy does not decrease, $\mathrm{d}S/\mathrm{d}t \geq 0$, with S the macroscopic entropy defined through $\mathrm{d}S = \delta Q/T$ and computed by Boltzmann as the logarithm of the number of microstates. (Its microscopic cousin, the von Neumann entropy, behaves differently and appears in Chapter 19; here "entropy" means thermodynamic entropy.)

- *Wheeler's question* (Figure 2.1). Drop a cup of hot tea into a black hole. Its entropy disappears behind the horizon, and classically nothing behind the horizon influences the exterior again. If the black hole has zero entropy, the second law for the exterior world is violated, not by a fluctuation but reliably, by a mundane process.
- *Bekenstein's answer.* Save the second law by assigning the black hole an entropy of its own, which grows by at least as much as the matter entropy lost. The candidate must then be a quantity that classical general relativity never lets decrease.

2.2 Why the area?

- *Not the mass.* The Penrose process of Chapter 1 extracts energy from a rotating hole, so M can decrease; J and Q change in either direction too.
- *The area.* By the area theorem (Theorem 1.1), granted the null energy condition and cosmic censorship, the total horizon area only grows. Bekenstein proposed that the entropy is a function of A.
- *A linear function.* Entropy must be additive, and the second law must hold in mergers, where the area theorem guarantees only $A_{\text{final}} \geq A_1 + A_2$. A concave function such as

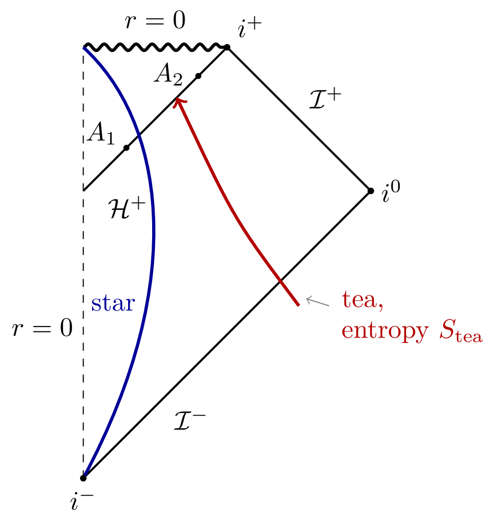


Figure 2.1: The tea-cup problem. A cup of tea carrying entropy S_{tea} falls into a black hole formed by collapse. For exterior observers its entropy is gone. Bekenstein's exchange rate: the horizon area grows from A_1 to A_2, and the growth of $A/4G$ more than compensates the loss of S_{tea}.

$\sqrt{A} \propto M_{\text{irr}}$ fails, since $\sqrt{A_1 + A_2} < \sqrt{A_1} + \sqrt{A_2}$. Linearity is the simplest choice for which the area theorem implies the second law in all such processes.

- *Dimensional analysis.* Entropy is dimensionless and area has dimensions of length squared; G and c alone give no length. With $\hbar$ one builds the Planck length, $\ell_P = (\hbar G/c^3)^{1/2} \approx 1.6 \times 10^{-33}$ cm, and hence a unique area.

Bekenstein–Hawking entropy. A black hole with horizon area A carries the entropy

$$S_{\text{BH}} \;=\; \frac{A}{4G\hbar} \;=\; \frac{A}{4\ell_P^2} \qquad (c = k_B = 1)\,: \tag{2.1}$$

one quarter of the horizon area in Planck units. The coefficient 1/4 is not fixed by Bekenstein's argument; it is determined by Hawking's calculation of black hole radiance (Chapter 4).

- *The $\hbar$ is essential.* As $\hbar \to 0$ at fixed A the entropy diverges, and a system with $S \to \infty$ at fixed energy has $T = (\partial S/\partial E)^{-1} \to 0$: this is why classical relativity sees a zero-temperature object with no visible entropy. From now on $\hbar = 1$ and $S_{\text{BH}} = A/4G$.

- *Schwarzschild numbers.* With $A = 16\pi G^2 M^2$,

$$S_{\text{BH}} \;=\; 4\pi G M^2 \;=\; 4\pi \left(\frac{M}{M_P}\right)^2, \qquad S_{\text{BH}}(M_\odot) \;\approx\; 10^{77}, \tag{2.2}$$

  with $M_P = G^{-1/2} \approx 2.2 \times 10^{-5}$ g. The Sun itself has $S_\odot \sim 10^{59}$ from its 10^{57} baryons: the black hole of the same mass is 10^{18} times more entropic, and the entropy of the observable universe is dominated by its supermassive black holes.

- *A strange thermodynamic system.* $S_{\text{BH}} \propto M^2$ rather than $\propto M$, and it scales with the area of the boundary rather than the volume. The first peculiarity underlies the negative specific heat of Chapter 5; the second is the germ of the holographic principle of Chapter 12.

Remark 2.1 (Newton's constant is dimensionful)**.** With $\hbar = c = 1$, a dimensionless Einstein–Hilbert action requires $[G] = (\text{length})^{D-2}$; in four dimensions $G = \ell_P^2 = 1/M_P^2$. An "expansion in G" is always an expansion in a dimensionless combination of G with a scale of the problem: GE^2 for graviton scattering at energy E, the power counting that makes gravity perturbatively non-renormalizable, and $\ell_P^2/L^2 \sim G/A$ for a black hole of curvature radius L. The thermodynamic description of this course is reliable when $S_{\rm BH} = A/4G \gg 1$. In Part II the expansion parameter of quantum gravity on AdS_5 is G_5/L^3, which the holographic dictionary equates with $1/N^2$ of the dual gauge theory (Chapters 12 and 16).

2.3 Generalized entropy and the Generalized Second Law

Definition 2.1 (Generalized entropy)**.** The *generalized entropy* of a black hole spacetime is the horizon contribution plus the ordinary entropy of matter and radiation outside the horizon:

$$S_{\rm gen} \;=\; \frac{A}{4G} \;+\; S_{\rm out}. \tag{2.3}$$

Generalized Second Law (GSL). In any process that can occur in practice, the generalized entropy is non-decreasing:

$$\frac{\mathrm{d}S_{\rm gen}}{\mathrm{d}t} \;=\; \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{A}{4G} + S_{\rm out}\right) \;\geq\; 0. \tag{2.4}$$

- *More than the sum of its parts.* The GSL permits the area to decrease if the exterior entropy grows enough (Hawking evaporation), and the exterior entropy to decrease if the area grows enough (the tea cup).
- *Which $S_{\rm out}$.* As stated, $S_{\rm out}$ is thermodynamic entropy, which presupposes local equilibrium outside. The modern formulation uses the von Neumann entropy of the exterior fields; its ultraviolet divergence at the horizon, and the renormalization of G that absorbs it, are the subject of Chapters 19 and 20.

2.4 The first test: absorbing black-body radiation

Shine black-body radiation of temperature T onto a Schwarzschild black hole of mass M, let it absorb an energy $E \ll M$, and check whether $S_{\rm gen}$ increased.

1. *Entropy of thermal radiation.* For a photon gas in a volume V at temperature T the free energy is $F = -\frac{\pi^2}{45}T^4 V$, so

$$S \;=\; -\frac{\partial F}{\partial T}\bigg|_V \;=\; \frac{4\pi^2}{45}\,T^3 V, \qquad E \;=\; F + TS \;=\; \frac{\pi^2}{15}\,T^4 V, \qquad \Longrightarrow \qquad S \;=\; \frac{4}{3}\,\frac{E}{T}, \tag{2.5}$$

   the 4/3 reflecting $p = E/3V$; the coefficient of F drops out of the ratio.

2. *Gain of horizon entropy.* With $S_{\rm BH} = 4\pi G M^2$ and $\delta M = E$, $\Delta S_{\rm BH} = 8\pi G M\, E$.
3. *Loss of exterior entropy.* $\Delta S_{\rm out} = -4E/3T$ by (2.5).
4. *Balance.*

$$\Delta S_{\rm gen} \;=\; \left(8\pi G M \;-\; \frac{4}{3T}\right) E \;\geq\; 0 \qquad \Longleftrightarrow \qquad T \;\geq\; T_c \;\equiv\; \frac{1}{6\pi G M}. \tag{2.6}$$

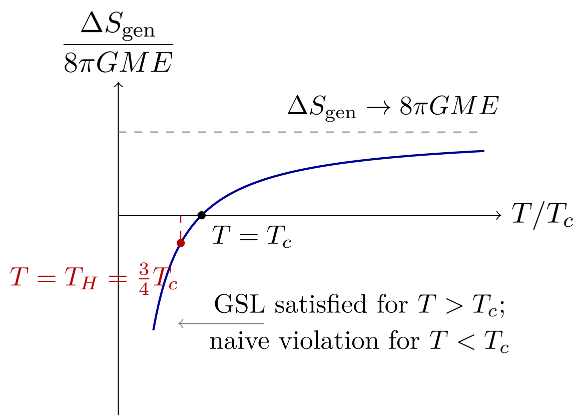


Figure 2.2: The entropy balance (2.6) for absorption of black-body radiation at temperature T, in units of $T_c = 1/6\pi GM$. For $T < T_c$ the naive generalized entropy decreases. The Hawking temperature $T_H = \frac{3}{4}T_c$ sits inside the problematic region: the resolution requires the black hole's own emission.

A thermal photon has wavelength $\lambda \sim 1/T$, so the condition is $\lambda \lesssim r_S$ up to order-one factors: the GSL passes whenever the absorbed quanta are small compared with the hole, and passes with room to spare for the tea cup, whose wavelengths are negligible against the kilometres of a stellar horizon (Figure 2.2).

2.5 The long-wavelength problem, and what it is telling us

- *The problem.* For $T < T_c$, radiation with wavelength larger than the hole, (2.6) is negative. Absorption of such quanta is inefficient, with a cross section suppressed, as stated in Chapter 5, but not zero, and the second law does not accept "rarely" as an excuse.

- *No resolution within Bekenstein's framework.* His one hypothesis was that nothing comes out. A body that absorbs but never emits has $T = 0$, and then $\mathrm{d}E = T\,\mathrm{d}S$ with $\mathrm{d}E \neq 0$ requires infinite $\mathrm{d}S$: finite entropy, finite energy and zero temperature cannot coexist for a system that exchanges energy.

- *The prediction.* The hypothesis is wrong: a black hole must emit, and precisely for quanta of wavelength of order r_S or longer, $\omega \lesssim 1/2GM$, where the naive GSL fails. This is what Hawking found in the collapse geometry (Chapters 3 and 4). With emission included the GSL holds for all bath temperatures: for $T > T_H$ the emitted quanta add entropy to S_{out}; for $T < T_H$ the hole loses mass, the area decreases, and the radiated entropy overcompensates. Chapter 5 verifies this mode by mode.

2.6 A temperature for the black hole

If the black hole is a thermodynamic system with entropy (2.2), its temperature is fixed by $\mathrm{d}E = T\,\mathrm{d}S$ with $E = M$: $1/T = \mathrm{d}S_{\text{BH}}/\mathrm{d}M = 8\pi GM$.

The temperature implied by $S = A/4G$. A Schwarzschild black hole of mass M must have temperature

$$T_H = \frac{1}{8\pi GM} = \frac{\kappa}{2\pi}, \tag{2.7}$$

with $\kappa = 1/4GM$ the surface gravity (1.10). In full units $T_H = \hbar c^3/8\pi GMk_B \approx 6 \times 10^{-8}\,\mathrm{K} \times (M_\odot/M)$.

- *The logic is reversed.* Here the entropy is assumed and the temperature deduced. Chapter 1 fixed (T, S) up to α; the coefficient $1/4$ in (2.1) is the choice $\alpha = 1/2\pi$. Hawking's calculation delivers $T = \kappa/2\pi$ outright and thereby fixes Bekenstein's coefficient.

- *The two puzzles cancel.* A quantum at temperature (2.7) has $\lambda \sim 1/T_H \sim r_S$: the hole radiates exactly in the regime where the absorption test failed.

- *Negligible for astrophysics, decisive conceptually.* 60 nK for a solar mass, and less for heavier holes.

- *With rotation and charge*, the first law of Chapter 1 becomes

$$\mathrm{d}M \;=\; T_H\,\mathrm{d}S_{\mathrm{BH}} \;+\; \Omega_H\,\mathrm{d}J \;+\; \Phi\,\mathrm{d}Q, \tag{2.8}$$

  now an honest thermodynamic first law.

2.7 $\mathrm{e}^{A/4G}$ microstates and no hair

- *The count.* Read as a Boltzmann entropy, (2.1) says a macroscopic black hole corresponds to

$$\mathcal{N} \;\sim\; \mathrm{e}^{A/4G} \tag{2.9}$$

  states, $\mathrm{e}^{10^{77}}$ for a solar mass. Bekenstein read S_{BH} as the number of ways the hole could have been formed: many initial configurations end in the same exterior state, labelled by M, J, Q.

- *What the states are* is, for a generic black hole, still unknown. The entropy is inferred from thermodynamic consistency and from the Euclidean path integral (Chapter 7), both indirect; Part III presents the partial answers that holography gives.

- *The no-hair theorem* (Israel, Carter and others; see [3]) states that a stationary black hole of Einstein–Maxwell theory is characterized by M, J, Q alone: once the quasinormal ringdown has died away, no record of the initial state remains in the exterior geometry. How can three numbers describe $\mathrm{e}^{10^{77}}$ states?

- *No hair is thermalization.* Water poured into a glass sloshes, then settles into a state characterized by its conserved quantities alone. The memory of the pouring is gone from the macroscopic description, not from the world: the evolution is unitary and the microstate encodes it in scrambled form. The settling raises the thermodynamic entropy without raising the microscopic entropy, a distinction made precise in Chapter 19. Ringdown is the sloshing, the Kerr–Newman geometry is the equilibrium state, (M, J, Q) are its thermodynamic coordinates, and the microstate is hidden from exterior observations as the water's is from a thermometer. The no-hair theorem is the zeroth piece of evidence that black holes are thermodynamic systems. Whether the information is recoverable in principle, as it is from the water, is the information problem of Part III.

Remark 2.2 (The supersymmetric exception)**.** For certain extremal charged black holes of string theory, which preserve supersymmetry and sit at the $\kappa = 0$ boundary of the third law, Strominger and Vafa [19] identified the microstates as states of wrapped branes and counted them at weak coupling; the logarithm reproduces $A/4G$ including the $1/4$. The count is protected as the coupling varies because of supersymmetry, which is what makes the comparison possible, and it is the strongest direct evidence that $\mathrm{e}^{A/4G}$ is the dimension of a Hilbert space. Branes are introduced in Chapter 15.

Chapter 3

Quantum Fields in Curved Spacetime

To find out how black holes radiate we need quantum field theory in curved spacetime, and this chapter assembles the toolkit: canonical quantization on a globally hyperbolic background, the Klein–Gordon inner product, the absence of a preferred vacuum, and Bogoliubov transformations, whose β coefficients measure particle creation. We then prepare Hawking's calculation: the Regge–Wheeler equation for a scalar in the Schwarzschild geometry, and the question of which state the field is in.

3.1 A free scalar field on a curved background

Throughout, gravity is a fixed classical background: we quantize a field ϕ on a given spacetime $(\mathcal{M}, g_{\mu\nu})$ and neglect its back-reaction, an excellent approximation whenever curvature radii are large compared to the Planck length [1].

- *Global hyperbolicity.* A *Cauchy surface* Σ is a hypersurface that every inextendible causal curve meets exactly once; a spacetime admitting one is *globally hyperbolic* and is diffeomorphic to $\mathbb{R} \times \Sigma$. On such a spacetime the Klein–Gordon equation has a well-posed initial value problem: data $(\phi, \partial_t \phi)$ on any Cauchy surface determine the solution everywhere. Minkowski space, the Kruskal extension (Cauchy surface: the $T = 0$ slice through the bifurcation sphere) and the collapse spacetime of Figure 1.2 are globally hyperbolic; the maximally extended Reissner–Nordström solution is not, predictability failing at the inner horizon r_-.

- *Action and field equation.* For a real scalar,

$$\begin{aligned} S[\phi] \;&=\; -\frac{1}{2}\int_{\mathcal{M}} \mathrm{d}^4x\,\sqrt{-g}\,\Big(g^{\mu\nu}\,\partial_\mu\phi\,\partial_\nu\phi \;+\; m^2\phi^2 \;+\; \xi R\,\phi^2\Big), \\ \Big(\Box - m^2 - \xi R\Big)\,\phi \;&=\; 0, \end{aligned} \tag{3.1}$$

  with $\Box\phi = (-g)^{-1/2}\partial_\mu(\sqrt{-g}\,g^{\mu\nu}\partial_\nu\phi)$. The coupling $\xi R\phi^2$ is the only other term of dimension four; $\xi = 0$ is minimal coupling, $\xi = \frac{1}{6}$ makes the massless theory conformally invariant in four dimensions.

- *Canonical structure.* For a foliation by Cauchy surfaces with unit normal n^μ and induced metric h_{ij}, the momentum is $\pi = \sqrt{h}\,n^\mu\partial_\mu\phi$ and

$$[\phi(t,\mathbf{x}),\, \pi(t,\mathbf{y})] = \mathrm{i}\,\delta^{(3)}(\mathbf{x}-\mathbf{y}), \qquad [\phi,\phi] = [\pi,\pi] = 0, \tag{3.2}$$

  as in flat space. Everything distinctive about curved spacetime enters through the choice of *states*, not through the algebra.

3.2 The Klein–Gordon inner product

Definition 3.1 (Klein–Gordon inner product)**.** For complex solutions f, g of (3.1) and a Cauchy surface Σ with future-directed unit normal n^μ,

$$(f, g) \;=\; -\mathrm{i}\int_\Sigma \mathrm{d}^3x\,\sqrt{h}\,n^\mu\,(f\,\partial_\mu g^* \;-\; g^*\,\partial_\mu f)\,. \tag{3.3}$$

Proposition 3.1 (Conservation)**.** *The inner product* (3.3) *is independent of the choice of Cauchy surface.*

Proof. (f, g) is the flux through Σ of $j^\mu = -\mathrm{i}(f\nabla^\mu g^* - g^*\nabla^\mu f)$, and

$$\nabla_\mu j^\mu \;=\; -\mathrm{i}\,(f\,\Box g^* - g^*\Box f) \;=\; -\mathrm{i}\Big(f\,(m^2+\xi R)\,g^* - g^*(m^2+\xi R)f\Big) \;=\; 0 \tag{3.4}$$

by the field equation for f and g^* (the conjugate of a solution is a solution, the equation being real). Gauss's theorem between two Cauchy surfaces equates the fluxes, for solutions falling off at spatial infinity. □

Remark 3.1. The proof uses only the field equation, so the pairing (ϕ, f) of the field *operator* with a classical solution, by which mode operators are extracted in (3.8), is likewise independent of the surface. Chapter 5 uses this, in (5.7), to evaluate a late-time observable on an early slice.

- *Indefiniteness.* The product is linear in the first slot, antilinear in the second, and not positive definite:
$$(g, f) \;=\; (f, g)^*, \qquad (f^*, g^*) \;=\; -\,(f, g)^*. \tag{3.5}$$
If f has positive norm, f^* has negative norm. This is the point: the split into positive- and negative-norm halves is what defines particles and antiparticles.

- *Minkowski space.* For $f_{\mathbf{k}} = N_{\mathbf{k}}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}-\mathrm{i}\omega_{\mathbf{k}}t}$ with $\omega_{\mathbf{k}} = +\sqrt{\mathbf{k}^2+m^2}$, a constant-$t$ slice has $n^\mu\partial_\mu = \partial_t$ and (3.3) gives
$$(f_{\mathbf{k}}, f_{\mathbf{k}'}) = 2\,\omega_{\mathbf{k}}\,|N_{\mathbf{k}}|^2\,(2\pi)^3\delta^{(3)}(\mathbf{k}-\mathbf{k}') \implies N_{\mathbf{k}} = \frac{1}{\sqrt{(2\pi)^3\,2\omega_{\mathbf{k}}}} \tag{3.6}$$
for the normalization $(f_{\mathbf{k}}, f_{\mathbf{k}'}) = \delta^{(3)}(\mathbf{k}-\mathbf{k}')$. Positive-frequency solutions, $\propto \mathrm{e}^{-\mathrm{i}\omega t}$ with $\omega > 0$, have positive norm; their conjugates have negative norm; the two sets are orthogonal.

3.3 Mode expansions and the ambiguity of the vacuum

- *Quantization from a choice of modes.* Let $\{f_i\}$ be a complete set of solutions, orthonormal in the Klein–Gordon sense,
$$(f_i, f_j) = \delta_{ij}, \qquad (f_i, f_j^*) = 0, \qquad (f_i^*, f_j^*) = -\,\delta_{ij}. \tag{3.7}$$
Expanding the field and projecting with the inner product,
$$\phi(x) \;=\; \sum_i \Big(a_i\,f_i(x) \;+\; a_i^\dagger\,f_i^*(x)\Big)\,, \qquad a_i \;=\; (\phi, f_i), \qquad a_i^\dagger \;=\; -\,(\phi, f_i^*), \tag{3.8}$$
and the canonical commutators (3.2) in the surface integrals give
$$[a_i, a_j^\dagger] \;=\; (f_i, f_j) \;=\; \delta_{ij}, \qquad [a_i, a_j] \;=\; -\,(f_i, f_j^*) \;=\; 0\,: \tag{3.9}$$
orthonormality converts field commutators into oscillator commutators. The vacuum and Fock space adapted to the modes are *defined* by $a_i|0_f\rangle = 0$ and $|i_1\cdots i_n\rangle \propto a_{i_1}^\dagger\cdots a_{i_n}^\dagger|0_f\rangle$.

- *Minkowski space is canonical.* Positive frequency with respect to inertial time is Poincaré invariant, so all inertial observers agree on $|0\rangle$ and on what a particle is.
- *The ambiguity.* In a general spacetime there is no preferred time, hence no preferred notion of positive frequency, hence no preferred vacuum. Any basis obeying (3.7) gives a consistent Fock quantization, and different choices are inequivalent: the vacuum of one is an excited state of another. "Particle" is a property of a state relative to a mode decomposition; the invariant object is the field and its correlation functions.
- *Stationary spacetimes* restore a preferred choice: modes with $\mathcal{L}_K f_\omega = -\mathrm{i}\omega f_\omega$, $\omega > 0$, for the timelike Killing vector K (for Schwarzschild, $f \propto \mathrm{e}^{-\mathrm{i}\omega t}$). For a black hole ∂_t is timelike only outside the horizon, and the vacuum it defines, the Boulware vacuum of Section 3.5.1, is pathological at the horizon.
- *Asymptotically static spacetimes* have two natural bases: the *in* modes, positive frequency with respect to the early-time clock, and the *out* modes, positive frequency at late times. The in vacuum $|0_{\mathrm{in}}\rangle$ has no particles for early observers, and whether the evolution creates particles is the question of how the in modes decompose into out modes.

3.4 Bogoliubov transformations

Let $\{f_i\}$ and $\{g_j\}$ be two orthonormal bases, say in and out.

- *Definition and projection.*

$$g_j \;=\; \sum_i \left(\alpha_{ji}\, f_i \;+\; \beta_{ji}\, f_i^*\right), \qquad \alpha_{ji} = (g_j, f_i), \qquad \beta_{ji} = -\,(g_j, f_i^*). \tag{3.10}$$

- *Constraints.* Orthonormality of the g_j, with (3.7) and (3.5), gives

$$\alpha\,\alpha^\dagger \;-\; \beta\,\beta^\dagger \;=\; \mathbf{1}, \qquad \alpha\,\beta^T \;-\; \beta\,\alpha^T \;=\; 0, \tag{3.11}$$

  the minus sign coming from the negative norm of the f_k^*.

- *Operators.* Expanding the same field in both bases, $\phi = \sum_i (a_i f_i + a_i^\dagger f_i^*) = \sum_j (b_j g_j + b_j^\dagger g_j^*)$, and projecting, $b_j = (\phi, g_j)$, with $(f_i, g_j) = \alpha_{ji}^*$ and $(f_i^*, g_j) = -\beta_{ji}^*$:

$$b_j \;=\; \sum_i \Big(\alpha_{ji}^*\, a_i \;-\; \beta_{ji}^*\, a_i^\dagger\Big)\,, \qquad b_j^\dagger \;=\; \sum_i \Big(\alpha_{ji}\, a_i^\dagger \;-\; \beta_{ji}\, a_i\Big)\,. \tag{3.12}$$

  If $\beta \neq 0$ the annihilation operators of one quantization mix annihilation and creation operators of the other, and the vacua differ.

- *Particle number.* In the state $|0_f\rangle$, using (3.12) and $a_i|0_f\rangle = 0$,

$$\langle N_j\rangle \;=\; \langle 0_f|\, b_j^\dagger b_j\, |0_f\rangle \;=\; \sum_i |\beta_{ji}|^2\,. \tag{3.13}$$

Particle creation from mode mixing. If the in modes, expanded in the out modes, acquire negative frequency components ($\beta \neq 0$), then the in vacuum contains out particles:

$$\langle N_j\rangle \;=\; \sum_i |\beta_{ji}|^2, \qquad \langle N_{\mathrm{tot}}\rangle \;=\; \mathrm{Tr}\,\beta\beta^\dagger. \tag{3.14}$$

The Hawking effect of Chapter 4 is a computation of β coefficients for a suitable pair of mode bases.

Remark 3.2 (Pair creation and squeezed states)**.** The old vacuum can be written explicitly in the new Fock space.

- *Pairs.* On a translation-invariant background modes carry momentum, $f_k, g_k \propto \mathrm{e}^{\mathrm{i}kx}$, and matching the x-dependence in (3.10) forces $\alpha_{ki} = \alpha_k\delta_{ik}$, $\beta_{ki} = \beta_k\delta_{i,-k}$, so (3.12) collapses to $b_k = \alpha_k^* a_k - \beta_k^* a_{-k}^\dagger$: the creation operator mixed into b_k is that of the opposite-momentum mode, and each pair $(k,-k)$ decouples. For a black hole the partner is the mode behind the horizon (Chapter 4).

- *One pair.* Writing $\tilde{a} \equiv a_{-k}$, $\tilde{b} \equiv b_{-k}$ and dropping k,

$$b \;=\; \alpha^* a - \beta^* \tilde{a}^\dagger, \qquad \tilde{b} \;=\; \alpha^* \tilde{a} - \beta^* a^\dagger, \qquad |\alpha|^2 - |\beta|^2 = 1, \tag{3.15}$$

  with $b|0_g\rangle = \tilde{b}|0_g\rangle = 0$ and $a|0_f\rangle = \tilde{a}|0_f\rangle = 0$. Inverting, $a = \alpha b + \beta^* \tilde{b}^\dagger$, and since $b\,\mathrm{e}^{\lambda b^\dagger \tilde{b}^\dagger}|0_g\rangle = \lambda \tilde{b}^\dagger \mathrm{e}^{\lambda b^\dagger \tilde{b}^\dagger}|0_g\rangle$, the condition $a|0_f\rangle = 0$ is solved by

$$|0_f\rangle \;\propto\; \exp\!\Big(\lambda\, b^\dagger \tilde{b}^\dagger\Big)|0_g\rangle \;\propto\; \sum_{n=0}^{\infty} \lambda^n\, |n\rangle_b \otimes |n\rangle_{\tilde{b}}, \qquad \lambda \;=\; -\frac{\beta^*}{\alpha}, \tag{3.16}$$

  and $\tilde{a}|0_f\rangle = 0$ gives the same λ.

- *Lesson.* The old vacuum contains new particles only in perfectly correlated pairs, n quanta in a mode with n in its partner: a *two-mode squeezed state.* These pair correlations, and the entanglement they embody, are at the heart of Chapter 6 and of Part III.

3.5 The massless scalar in the Schwarzschild geometry

We now set up the wave equation for a massless scalar, a stand-in for the photon, in the Schwarzschild metric (1.1), for which $\sqrt{-g} = r^2 \sin\theta$ and

$$\Box\phi \;=\; -\frac{1}{f}\,\partial_t^2\phi \;+\; \frac{1}{r^2}\,\partial_r\Big(r^2 f\,\partial_r\phi\Big) \;+\; \frac{1}{r^2}\,\Delta_{S^2}\phi \;=\; 0. \tag{3.17}$$

1. *Separation.* Staticity and spherical symmetry suggest

$$\phi(t,r,\theta,\varphi) \;=\; \mathrm{e}^{-\mathrm{i}\omega t}\, Y_{\ell m}(\theta,\varphi)\, \frac{u_{\omega\ell}(r)}{r}, \tag{3.18}$$

   with $\Delta_{S^2} Y_{\ell m} = -\ell(\ell+1) Y_{\ell m}$; the $1/r$ eliminates first-derivative terms. Substituting and multiplying by r,

$$\frac{\omega^2}{f}\,u \;+\; \frac{1}{r}\,\frac{\mathrm{d}}{\mathrm{d}r}\Big(r^2 f\,\frac{\mathrm{d}}{\mathrm{d}r}\frac{u}{r}\Big) \;-\; \frac{\ell(\ell+1)}{r^2}\,u \;=\; 0. \tag{3.19}$$

2. *The radial term.*

$$r^2 f\left(\frac{u}{r}\right)' = f\,(ru' - u)\,, \qquad \frac{1}{r}\,\frac{\mathrm{d}}{\mathrm{d}r}\Big[f(ru' - u)\Big] = \frac{f'(ru' - u)}{r} + f u'', \tag{3.20}$$

   the $f(ru'' + u' - u')$ term collapsing to fu''. Multiplying (3.19) by f,

$$\omega^2 u + f^2 u'' + f f' u' - f\left(\frac{f'}{r} + \frac{\ell(\ell+1)}{r^2}\right) u = 0. \tag{3.21}$$

3. *Tortoise coordinate.* With $\mathrm{d}r_*/\mathrm{d}r = 1/f$ from (1.2),

$$\frac{\mathrm{d}^2u}{\mathrm{d}r_*^2} = f\,\frac{\mathrm{d}}{\mathrm{d}r}\left(f\,\frac{\mathrm{d}u}{\mathrm{d}r}\right) = f^2u'' + ff'u', \tag{3.22}$$

exactly the combination in (3.21).

The Regge–Wheeler equation [20]. Each partial wave of a massless scalar in Schwarzschild obeys a flat-space Schrödinger-like scattering equation in the tortoise coordinate,

$$\frac{\mathrm{d}^2u_{\omega\ell}}{\mathrm{d}r_*^2} + \Big(\omega^2 - V_\ell(r)\Big)\,u_{\omega\ell} = 0, \qquad V_\ell(r) = \left(1 - \frac{2GM}{r}\right)\left(\frac{\ell(\ell+1)}{r^2} + \frac{2GM}{r^3}\right), \tag{3.23}$$

with $r = r(r_*)$. The centrifugal term is inherited from flat space; the term $2GM/r^3 = f'/r$ is a gravitational barrier present even for $\ell = 0$; the factor f shuts the potential off at the horizon. For spin s the same derivation gives $V = f\,[\ell(\ell+1)/r^2 + 2GM(1-s^2)/r^3]$.

- *Asymptotic plane waves.* As $r \to \infty$, $V_\ell \to 0$ as a power; as $r \to 2GM$, i.e. $r_* \to -\infty$, $f \sim \mathrm{e}^{r_*/2GM}$ kills the potential exponentially. At both ends the equation is free,

$$u_{\omega\ell} \;\sim\; \mathrm{e}^{\pm \mathrm{i}\omega r_*} \qquad (r_* \to \pm\infty), \tag{3.24}$$

  so in the null coordinates $u = t - r_*$, $v = t + r_*$ each partial wave is a free massless field in 1+1 dimensions near the horizon and near infinity, with the barrier, peaked near $r \sim 3GM$, acting in between (Figure 5.1).

- *A scattering problem.* A wave climbing out of the horizon region is partly transmitted to infinity, with probability $\Gamma_\ell(\omega)$, and partly reflected. These gray-body factors filter the thermal emission of the horizon and are the subject of Chapter 5; for $\omega r_S \ll 1$ the barrier is nearly opaque, which is the inefficient absorption of long-wavelength quanta met in Chapter 2.

- *Hawking's problem is two-dimensional.* The quantum mechanics of black hole radiance involves only the free 1+1 fields near the horizon and near infinity, connected by ray tracing through the collapse geometry; the four-dimensional details enter only through the gray-body filtering. This is the structure exploited in Chapter 4 [1].

3.5.1 Which state?

The mode analysis fixes the field equation, not the state: by Section 3.3, a state on the Schwarzschild background is a choice, and the physics of the next chapters depends on it.

- *The Boulware vacuum [21] fails at the horizon.* This is the choice singled out in Section 3.3 for a stationary spacetime: positive frequency with respect to the Killing time t, i.e. the modes $\mathrm{e}^{-\mathrm{i}\omega u}$ near the horizon. By (1.4), $\mathrm{e}^{-\mathrm{i}\omega u} = (-U)^{\mathrm{i}\omega/\kappa}$ oscillates infinitely fast in the Kruskal coordinate U as $U \to 0$: to a freely falling observer crossing the horizon the state does not look like the Minkowski vacuum at short distances, and its renormalized stress tensor diverges there [22].

- *Admissible states are Hadamard.* Their short-distance correlations must agree with those of the Minkowski vacuum in every freely falling frame [5]. Two such states will be constructed.

- *The Unruh state [23]* is the state produced by gravitational collapse, which Hawking's calculation of Chapter 4 determines; it is named in Section 5.3.
- *The Hartle–Hawking state [24]* is the equilibrium state of the eternal black hole, constructed in Section 7.8 by a Euclidean path integral.

The lesson to carry forward is that black hole radiance is a property of the geometry together with a state, and that regularity at the horizon is what forces thermal behaviour.

Chapter 4

Hawking Radiation

This chapter contains the central result of Part I: a black hole formed in gravitational collapse radiates thermally at $T_H = \kappa/2\pi$. The key geometric fact is the exponential map $u \simeq -4GM\ln(v_0 - v)$ relating late-time outgoing rays to their ingoing ancestors; fed into the geometric-optics propagation of modes it gives Bogoliubov coefficients with $|\alpha_\omega|^2 = \mathrm{e}^{8\pi GM\omega}|\beta_\omega|^2$, hence thermal occupation at $T_H = 1/8\pi GM$, confirmed by a near-horizon two-point-function shortcut. The first law then fixes the coefficient in Bekenstein's entropy, $S = A/4G$. We end with the entanglement of the emitted pairs, the evaporation rate, and the information problem.

4.1 The collapse spacetime and the strategy

We work in the collapse spacetime of Chapter 1 (Figure 4.1), with a free massless minimally coupled scalar ϕ on the fixed background, an approximation that is superb when r_S is large compared with the Planck length. The question is operational: what does a distant observer near $\mathscr{I}^+$ measure at late times, after the transients of the collapse have died away?

- *Trace observables backwards* [25, 26]. Any observable at $\mathscr{I}^+$ can be evolved back to data on an earlier Cauchy surface: all the way to $\mathscr{I}^-$, before the collapse, where the state is the Minkowski vacuum; or to a slice $\mathcal{S}$ crossing the horizon at late times, outside the star.

- *Late-time signals come from exponentially close to the horizon.* In the Kruskal metric (1.6) outgoing rays are lines of constant U. For any smooth transverse coordinate u vanishing on the horizon and positive outside, $u = -cU + \mathcal{O}(U^2)$ with $c > 0$, so by (1.4) the ray labelled U has retarded time $-4GM\ln(-U) = 4GM\ln(1/u) + \text{const}$, and the observer at fixed radius receives it at

$$t \;=\; 4GM\,\ln\frac{1}{u} \;+\; C, \qquad \text{i.e.} \qquad u \;=\; \mathrm{e}^{C/4GM}\,\mathrm{e}^{-t/4GM}. \tag{4.1}$$

  Waiting one more time $4GM$, about 2×10^{-5} s for a solar mass, multiplies u by $1/\mathrm{e}$.

- *Universality.* Every reasonable state looks like the vacuum at short distances, and late-time observations probe exponentially short distances from the horizon, so they are insensitive to how the black hole formed: whatever is seen at late times is a property of the geometry, characterized by M alone.

- *Exponential redshift.* A quantum received with energy ω climbed out of the near-horizon region from an exponentially larger proper energy, so it propagated by geometric optics, which lets us push the modes through the collapse region.

- *No trans-Planckian input is needed* [1]. It suffices to trace the signal back to a slice $\mathcal{S}$ where it originates at a distance from the horizon tiny compared with r_S but huge

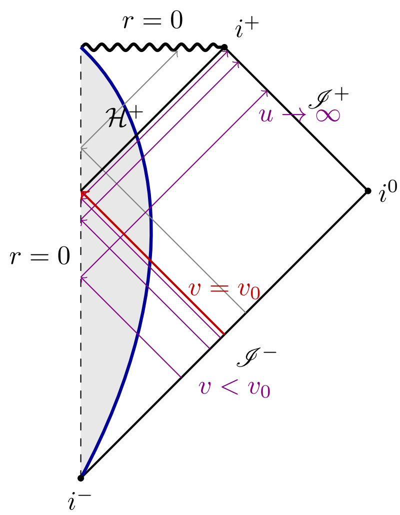


Figure 4.1: Ray tracing through the collapsing star. Ingoing radial null rays labelled by advanced time v pass through the centre $r=0$, emerge as outgoing rays, and reach $\mathscr{I}^+$ at retarded time $u(v)$. The critical ray $v=v_0$ (red) becomes the horizon generator. Rays with v slightly below v_0 (violet) hug the horizon and escape only at very late retarded time: equal spacings in v map to exponentially growing spacings in u. Rays with $v>v_0$ (gray) are trapped.

compared with any scale of unknown physics; for a kilometre-sized hole a millimetre already suffices for the vacuum approximation.

4.2 Ray tracing: the exponential map between $\mathscr{I}^-$ and $\mathscr{I}^+$

- *Two labellings of the same rays.* With $u=t-r_*$, $v=t+r_*$ and the Kruskal coordinates

$$U \;=\; -\,\mathrm{e}^{-\kappa u}, \qquad V \;=\; \mathrm{e}^{\kappa v}, \qquad \kappa \;=\; \frac{1}{4GM}, \tag{4.2}$$

  radial null geodesics are lines of constant U or V. At $\mathscr{I}^+$ the physical label of an outgoing ray is the retarded time u read by the observer's clock, and the whole future $u\to\infty$ is crammed into $U\to 0^-$; near the horizon it is U that is regular while u diverges. The exponential relation between the two, with rate κ, is the geometric origin of the thermal spectrum; (4.1) is (4.2) with $u_{\text{transverse}}=-U$ and $t\simeq u+\text{const}$.

- *Through the star.* Follow the ray reaching $\mathscr{I}^+$ at retarded time u backwards: it plunges into the star, passes through the regular centre $r=0$, re-emerges as an ingoing ray and arrives at $\mathscr{I}^-$ at advanced time v. This defines a map $v\mapsto u(v)$. The ingoing ray at v_0 that hovers forever at $r=2GM$ becomes the horizon generator; rays with $v>v_0$ are trapped; rays with v_0-v small escape at enormous u.

- *Smoothness is all that is needed.* Between $\mathscr{I}^-$ and any late-time surface near the horizon the backward ray crosses only smooth regions, so the map between the two smooth labellings, U near the horizon and v at $\mathscr{I}^-$, is smooth with smooth inverse near the critical ray $U=0\leftrightarrow v=v_0$:

$$v \;=\; v_0 \;+\; a\,U \;+\; \mathcal{O}(U^2), \qquad a>0. \tag{4.3}$$

  Inserting $U=-\mathrm{e}^{-\kappa u}$,

$$v_0-v \;=\; a\,\mathrm{e}^{-\kappa u} \qquad \Longleftrightarrow \qquad u \;=\; -\,\frac{1}{\kappa}\,\ln\frac{v_0-v}{a} \;=\; -\,4GM\,\ln\frac{v_0-v}{a}. \tag{4.4}$$

- *What is robust.* The rate κ is fixed by the exterior geometry alone. The collapse details enter only through a and v_0, shifts and rescalings of the null coordinates, which will contribute only phases to the Bogoliubov coefficients: this is the precise sense in which Hawking radiation is independent of how the black hole formed.

- *A redshift that compounds.* By (4.4), $\mathrm{d}u/\mathrm{d}v = \mathrm{e}^{\kappa u}/a\kappa$: a wave crest leaving $\mathscr{I}^-$ just before v_0 is received stretched by a factor growing exponentially with the reception time, unlike the fixed static redshift between hovering observers. Exponential redshift at a constant rate is what turns vacuum fluctuations into steady thermal radiation.

4.3 Propagation of modes: geometric optics

- *Partial waves.* Expanding $\phi = \sum_{lm} \sigma_{lm}(t, r_*) Y_{lm}/r$, each σ_{lm} obeys the Regge–Wheeler equation (3.23),
$$\Big(\partial_t^2 - \partial_{r_*}^2 + V_l(r)\Big)\,\sigma_{lm} \;=\; 0. \tag{4.5}$$
In this chapter we neglect V_l, so that each partial wave is a free two-dimensional field, $\sigma \simeq \sigma_{\rm out}(u) + \sigma_{\rm in}(v)$; the barrier multiplies the flux in each mode by a transmission probability $\Gamma_l(\omega) \le 1$ (Chapter 5) and does not change the temperature.

- *Out-modes.* The modes relevant to the distant observer behave near $\mathscr{I}^+$ as
$$p_\omega \;\sim\; \frac{1}{\sqrt{4\pi\omega}}\,\mathrm{e}^{-\mathrm{i}\omega u}, \qquad \omega > 0, \tag{4.6}$$
with the standard two-dimensional Klein–Gordon normalization.

- *Geometric optics.* A solution $A\mathrm{e}^{\mathrm{i}S}$ with rapidly varying phase satisfies the wave equation to leading order iff $g^{\mu\nu}\partial_\mu S \partial_\nu S = 0$, i.e. the phase is constant along null geodesics, valid when the local wavelength is small compared with the curvature radius and the star. For late-time modes this is exponentially well satisfied: a mode received with $\omega \sim 1/GM$ at retarded time u had, on its way through the star, a local frequency blueshifted by $\mathrm{e}^{\kappa u}$.

- *The traced-back mode.* Transporting the phase $\mathrm{e}^{-\mathrm{i}\omega u}$ along the rays to $\mathscr{I}^-$ and inserting (4.4),
$$p_\omega\Big|_{\mathscr{I}^-} \;\simeq\; \frac{1}{\sqrt{4\pi\omega}}\,P_\omega(v), \qquad P_\omega(v) \;=\; \begin{cases} \exp\left[\mathrm{i}\,\dfrac{\omega}{\kappa}\,\ln\dfrac{v_0 - v}{a}\right], & v < v_0, \\ 0, & v > v_0, \end{cases} \tag{4.7}$$
the zero for $v > v_0$ expressing that those rays never reach $\mathscr{I}^+$. P_ω oscillates infinitely fast as $v \to v_0^-$ and vanishes beyond: as far from a positive-frequency plane wave in v as can be, which is why particles are created.

4.4 The Bogoliubov coefficients

1. *Setup.* On $\mathscr{I}^-$ the positive-frequency modes are $f_{\omega'} = (4\pi\omega')^{-1/2}\mathrm{e}^{-\mathrm{i}\omega' v}$, $\omega' > 0$, and the in-state is their vacuum $|0_{\rm in}\rangle$. Expand the traced-back out-mode as in (3.10),
$$p_\omega \;=\; \int_0^\infty \mathrm{d}\omega' \,\Big(\alpha_{\omega\omega'}\, f_{\omega'} + \beta_{\omega\omega'}\, \bar{f}_{\omega'}\Big). \tag{4.8}$$
The coefficients are Fourier coefficients: with $\widetilde{P}_\omega(\omega') = \int \mathrm{d}v\, \mathrm{e}^{\mathrm{i}\omega' v} P_\omega(v)$,
$$\alpha_{\omega\omega'} \;=\; \frac{1}{2\pi}\sqrt{\frac{\omega'}{\omega}}\;\widetilde{P}_\omega(\omega'), \qquad \beta_{\omega\omega'} \;=\; \frac{1}{2\pi}\sqrt{\frac{\omega'}{\omega}}\;\widetilde{P}_\omega(-\omega'), \tag{4.9}$$

and by (3.13) the in-vacuum contains $\int_0^\infty \mathrm{d}\omega' |\beta_{\omega\omega'}|^2$ quanta in the out-mode ω.

2. *The master integral.* With $x = v_0 - v$ in (4.7),

$$\widetilde{P}_\omega(\pm\omega') \;=\; \mathrm{e}^{\pm \mathrm{i}\omega' v_0}\, a^{-\mathrm{i}\omega/\kappa} \int_0^\infty \mathrm{d}x\; \mathrm{e}^{\mp \mathrm{i}\omega' x}\, x^{\mathrm{i}\omega/\kappa}, \qquad \omega' > 0, \tag{4.10}$$

where v_0 and a sit only in phases.[1]

3. *Contour rotation.* For $\widetilde{P}_\omega(+\omega')$ the exponential $\mathrm{e}^{-\mathrm{i}\omega' x}$ decays in the lower half-plane; rotate $x = \mathrm{e}^{-\mathrm{i}\pi/2} s$, on which the principal branch gives $x^{\mathrm{i}\omega/\kappa} = s^{\mathrm{i}\omega/\kappa}\mathrm{e}^{+\pi\omega/2\kappa}$, so

$$\int_0^\infty \mathrm{d}x\, \mathrm{e}^{-\mathrm{i}\omega' x} x^{\mathrm{i}\omega/\kappa} \;=\; \mathrm{e}^{-\mathrm{i}\pi/2}\, \mathrm{e}^{+\pi\omega/2\kappa} \int_0^\infty \mathrm{d}s\; \mathrm{e}^{-\omega' s}\, s^{\mathrm{i}\omega/\kappa} \;=\; \mathrm{e}^{-\mathrm{i}\pi/2}\, \mathrm{e}^{+\pi\omega/2\kappa}\, \frac{\Gamma(1+\mathrm{i}\omega/\kappa)}{(\omega')^{1+\mathrm{i}\omega/\kappa}}. \tag{4.11}$$

For $\widetilde{P}_\omega(-\omega')$ the exponential $\mathrm{e}^{+\mathrm{i}\omega' x}$ decays in the upper half-plane; rotating $x = \mathrm{e}^{+\mathrm{i}\pi/2} s$ gives $x^{\mathrm{i}\omega/\kappa} = s^{\mathrm{i}\omega/\kappa}\mathrm{e}^{-\pi\omega/2\kappa}$ and

$$\int_0^\infty \mathrm{d}x\, \mathrm{e}^{+\mathrm{i}\omega' x} x^{\mathrm{i}\omega/\kappa} \;=\; \mathrm{e}^{+\mathrm{i}\pi/2}\, \mathrm{e}^{-\pi\omega/2\kappa}\, \frac{\Gamma(1+\mathrm{i}\omega/\kappa)}{(\omega')^{1+\mathrm{i}\omega/\kappa}}. \tag{4.12}$$

The two rotations approach the branch point $x = 0$ of $x^{\mathrm{i}\omega/\kappa}$ from opposite sides of the cut, and the entire difference between (4.11) and (4.12) is the factor $\mathrm{e}^{\pi\omega/\kappa}$.

4. *The Boltzmann ratio.* Since $|(\omega')^{-\mathrm{i}\omega/\kappa}| = 1$,

$$\boxed{\left|\alpha_{\omega\omega'}\right|^2 \;=\; \mathrm{e}^{2\pi\omega/\kappa} \left|\beta_{\omega\omega'}\right|^2 \;=\; \mathrm{e}^{8\pi G M \omega} \left|\beta_{\omega\omega'}\right|^2} \tag{4.13}$$

for every ω': a Boltzmann factor $\mathrm{e}^{-E/T}$ with $E = \omega$ and $1/T = 2\pi/\kappa = 8\pi G M$, with all reference to a, v_0 and ω' gone. Equivalently, $\beta(\omega')$ is the continuation of $\alpha(\omega')$ through the upper half ω'-plane, $\omega' \to \mathrm{e}^{\mathrm{i}\pi}\omega'$: the thermal factor comes from the branch cut, that is, from the logarithm in (4.4). Exponential kinematics in, thermal physics out. The normalization, using $|\Gamma(1+\mathrm{i}y)|^2 = \pi y/\sinh \pi y$, is

$$\left|\beta_{\omega\omega'}\right|^2 \;=\; \frac{1}{2\pi\kappa\,\omega'}\; \frac{1}{\mathrm{e}^{2\pi\omega/\kappa} - 1}\,. \tag{4.14}$$

5. *A steady flux.* $\int \mathrm{d}\omega' |\beta_{\omega\omega'}|^2$ diverges logarithmically at small ω': each logarithmic interval of ω' contributes equally, and a shift $u \to u + \Delta$ rescales $\omega' \to \mathrm{e}^{\kappa\Delta}\omega'$ by (4.4), so equal intervals of emission time are equal logarithmic intervals of ω'. The divergence says that the hole radiates at a steady rate for all time, and the plane-wave mode collects all of it.

6. *Wavepackets and occupation.* Hawking's wavepackets

$$p_{jn} \;=\; \frac{1}{\sqrt{\varepsilon}} \int_{j\varepsilon}^{(j+1)\varepsilon} \mathrm{d}\omega\; \mathrm{e}^{2\pi \mathrm{i} n\omega/\varepsilon}\; p_\omega \tag{4.15}$$

are unit-norm modes centred on $\omega_j \simeq (j + \frac{1}{2})\varepsilon$ and on retarded time $u_n \simeq 2\pi n/\varepsilon$; for late packets (4.13) holds with $\omega = \omega_j$. The first identity in (3.11), for the normalized mode, reads $\int \mathrm{d}\omega' (|\alpha_{jn,\omega'}|^2 - |\beta_{jn,\omega'}|^2) = 1$, and combined with the Boltzmann ratio,

$$\langle 0_{\mathrm{in}} |\, N_{jn}\, | 0_{\mathrm{in}} \rangle \;=\; \int_0^\infty \mathrm{d}\omega' \left|\beta_{jn,\omega'}\right|^2 \;=\; \frac{1}{\mathrm{e}^{2\pi\omega_j/\kappa} - 1}\,, \tag{4.16}$$

independent of the time label n: a steady flux with the Bose–Einstein occupation of a thermal state.

[1] The integrals are conditionally convergent and are defined as $\lim_{\epsilon\to 0^+} \int_0^\infty \mathrm{d}x\, \mathrm{e}^{-(\epsilon \pm \mathrm{i}\omega')x} x^{\mathrm{i}\omega/\kappa}$. The regulated integrand decays exponentially, so the integral converges absolutely and the contour rotations below are justified by Cauchy's theorem: the rotated contours keep $\mathrm{Re}\,[(\epsilon \pm \mathrm{i}\omega')x] > 0$, and the arcs at infinity do not contribute.

Hawking radiation. A black hole formed by collapse settles into a state in which a distant observer detects, in each outgoing mode of frequency ω and angular momentum l, a mean number of quanta

$$\langle N_\omega \rangle \;=\; \frac{\Gamma_l(\omega)}{\mathrm{e}^{\omega/T_H} - 1}\,, \qquad T_H \;=\; \frac{\kappa}{2\pi} \;=\; \frac{1}{8\pi G M}\,, \tag{4.17}$$

where $\Gamma_l(\omega)$ is the gray-body factor defined in (5.3), equal to 1 when the barrier is neglected. Restoring units, $T_H = \hbar c^3/8\pi G M k_B$: the temperature is proportional to $\hbar$ and invisible in classical general relativity.

For fermions the Bogoliubov identity carries a relative plus sign and the same computation gives the Fermi–Dirac occupation $\Gamma/(\mathrm{e}^{\omega/T_H} + 1)$ at the same temperature, as a genuine temperature requires.

4.5 The Hawking temperature: magnitudes and sanity checks

- *Generality.* The derivation used only the exponential relation $U = -\mathrm{e}^{-\kappa u}$ between the regular horizon coordinate and the asymptotic clock, so $T_H = \kappa/2\pi$ holds for any black hole with surface gravity κ, Kerr and Reissner–Nordström included.

- *Wavelengths.* For Schwarzschild

$$T_H \;=\; \frac{1}{8\pi G M} \;=\; \frac{1}{4\pi r_S}\,, \tag{4.18}$$

  so a typical quantum has wavelength of order r_S: the emission is a coherent property of the whole near-horizon region, not of a point on the horizon (Section 4.8).

- *Numbers.*

$$T_H \;\approx\; 6.2 \times 10^{-8} \left(\frac{M_\odot}{M}\right) \mathrm{K}, \tag{4.19}$$

  colder than the microwave background by 4×10^7 for a solar mass. The temperature grows as the hole shrinks, the negative specific heat of Chapter 5; holes heavier than about $10^{-7} M_\odot$ are colder than the background today and grow rather than evaporate.

4.6 A shortcut: the near-horizon two-point function

Witten [1] extracts the temperature from correlation functions, bypassing Bogoliubov coefficients; the route recurs in Chapters 6 and 7.

1. The observer's two-point function at fixed large radius, traced back to the slice $\mathcal{S}$ by (4.1), probes the field at transverse positions $u = \mathrm{e}^{C/4GM}\mathrm{e}^{-t/4GM}$ and u', exponentially small and close together, where every state has the vacuum form. The simplest case, realized by the lowest partial wave of a charged fermion around a magnetically charged hole, is a two-dimensional chiral fermion with vacuum correlator

$$\langle \psi(u)\,\psi(u') \rangle \;=\; \frac{(\mathrm{d}u\,\mathrm{d}u')^{1/2}}{u - u'}\,, \tag{4.20}$$

   the half densities supplying, under $u \to t(u)$, the Jacobians $(\mathrm{d}u/\mathrm{d}t)^{1/2}$ required for a field of dimension $\frac{1}{2}$.

2. Change variables to the observer's clock, $u = \mathrm{e}^{C/4GM}\mathrm{e}^{-t/4GM}$:

$$u - u' = \mathrm{e}^{C/4GM}(\mathrm{e}^{-t/4GM} - \mathrm{e}^{-t'/4GM}), \qquad (\mathrm{d}u\,\mathrm{d}u')^{1/2} = \frac{\mathrm{e}^{C/4GM}}{4GM}\,\mathrm{e}^{-(t+t')/8GM}\,(\mathrm{d}t\,\mathrm{d}t')^{1/2}, \tag{4.21}$$

so that

$$\langle\psi(t)\,\psi(t')\rangle \;=\; \frac{1}{8GM}\,\frac{(\mathrm{d}t\,\mathrm{d}t')^{1/2}}{\sinh(\frac{t-t'}{8GM})}\;: \tag{4.22}$$

C has cancelled, the result depends only on $t - t'$, and it is antiperiodic under $t \to t + 8\pi GM\,\mathrm{i}$.

Antiperiodicity in imaginary time is the fingerprint of thermal equilibrium.

The Kubo–Martin–Schwinger (KMS) condition [27, 28]**.** Let $\rho_\beta = \mathrm{e}^{-\beta H}/Z$ be the thermal state of a system with Hamiltonian H, and $A(t) = \mathrm{e}^{\mathrm{i}Ht}A\,\mathrm{e}^{-\mathrm{i}Ht}$. For any two operators A and B,

$$\langle A(t)\,B\rangle_\beta \;=\; \langle B\,A(t+\mathrm{i}\beta)\rangle_\beta, \qquad \langle\,\cdot\,\rangle_\beta \;\equiv\; \mathrm{Tr}\,[\rho_\beta\;\cdot\;]\;: \tag{4.23}$$

moving an operator once around the thermal circle $t \to t + \mathrm{i}\beta$ exchanges the operator ordering. Consequently, Euclidean (time-ordered) correlators are periodic in imaginary time with period β for bosonic operators and antiperiodic for fermionic ones. Conversely, a state whose correlators obey (4.23) is thermal at inverse temperature β with respect to H; this characterization involves only correlation functions, and survives when ρ_β itself is ill defined.

The proof is cyclicity of the trace: with $\mathrm{e}^{\beta H}B\,\mathrm{e}^{-\beta H} = B(-\mathrm{i}\beta)$,

$$\mathrm{Tr}\,[\mathrm{e}^{-\beta H}A(t)\,B] \;=\; \mathrm{Tr}\,[B\,\mathrm{e}^{-\beta H}A(t)] \;=\; \mathrm{Tr}\,[\mathrm{e}^{-\beta H}\,B(-\mathrm{i}\beta)\,A(t)],$$

and time-translation invariance by $\mathrm{i}\beta$ gives the right-hand side; for fermions the reordering inside a time-ordered correlator costs a sign. The correlator (4.22) is the thermal two-point function of a chiral fermion at $\beta = 8\pi GM$, fixed uniquely by antiperiodicity and a simple pole of unit residue at $t = t'$; a bosonic chiral current with correlator $\mathrm{d}u\,\mathrm{d}u'/(u-u')^2$ gives a periodic correlator at the same β. The thermal nature of the black hole is the exponential map (4.1) plus the universal short-distance structure of quantum fields, nothing else.

4.7 The first law revisited: $S = A/4G$

Bekenstein's entropy was $\propto A/G$ with an undetermined coefficient, and the four laws fixed $(T, S) = (\alpha\kappa, A/8\pi G\alpha)$ up to one constant. If the black hole is a thermal system at $T_H = 1/8\pi GM$, the first law gives

$$\mathrm{d}S \;=\; \frac{\mathrm{d}M}{T_H} \;=\; 8\pi GM\,\mathrm{d}M \qquad\Longrightarrow\qquad S \;=\; 4\pi GM^2, \tag{4.24}$$

with $S = 0$ at $M = 0$. Since $A = 16\pi G^2M^2$:

The Bekenstein–Hawking entropy. Consistency of the Hawking temperature with the first law $\mathrm{d}M = T\,\mathrm{d}S$ fixes

$$S_{\mathrm{BH}} \;=\; 4\pi GM^2 \;=\; \frac{A}{4G} \;=\; \frac{A}{4\,\ell_P^2} \qquad (\ell_P^2 = G\hbar \text{ restored}), \tag{4.25}$$

one quarter of the horizon area in Planck units: Bekenstein's proposal with the coefficient determined [26].

- *Rotation and charge.* The same identifications turn the classical first law (1.20) into $\mathrm{d}M = T_H\,\mathrm{d}S + \Omega_H \mathrm{d}J + \Phi\,\mathrm{d}Q$.

- *What the entropy counts* is not answered by this calculation, which never exhibits e^S microstates; the question stays with us for the rest of the course.

- *The long-wavelength puzzle of Chapter 2 dissolves.* By (4.17), modes with $\omega \ll T_H$ have occupation $\approx \Gamma T_H/\omega \gg \Gamma$: the hole strongly emits exactly the quanta Bekenstein could not absorb consistently, and the entropy they carry rescues the GSL, as Chapter 5 shows mode by mode.

4.8 Where are the quanta created? Pairs and entanglement

The derivation is global, mode functions in and occupation numbers out, and local pictures of "what happens at the horizon" should be resisted. Three statements are sharp.

- *The emission region.* A Hawking quantum has wavelength $\sim r_S$, so it materializes out of a region of that size around the horizon, the quantum atmosphere. A freely falling observer crossing the horizon sees the vacuum, as the equivalence principle requires; only the long-wavelength correlations, read out by the exponentially accelerating clock (4.1), are thermal.

- *Pairs and negative energy.* The mixing (4.13) creates quanta in pairs: each escaping wavepacket has a partner whose support lies behind the horizon ($v > v_0$, the gray ray of Figure 4.1). The partner carries negative Killing energy, allowed because ∂_t is spacelike inside the horizon, and falls into the singularity: this is how the mass decreases by ω per escaped quantum, and how the quantum theory evades the area theorem, whose null-energy hypothesis $\langle T_{\mu\nu}\rangle$ violates near the horizon.

- *Entanglement.* For each mode pair the state has the two-mode-squeezed form (3.16) with $|\lambda| = |\beta/\alpha| = \mathrm{e}^{-\pi\omega/\kappa}$:

$$|0_{\mathrm{in}}\rangle \;\sim\; \prod_{\omega} \left(\sum_{n=0}^{\infty} \mathrm{e}^{-\pi n\omega/\kappa}\; |n\rangle_{\mathrm{inside}} \otimes |n\rangle_{\mathrm{outside}} \right), \tag{4.26}$$

  a pure state whose restriction to the exterior is the thermal density matrix at $\kappa/2\pi$, the probabilities being $\mathrm{e}^{-2\pi n\omega/\kappa} = \mathrm{e}^{-n\omega/T_H}$. Chapter 6 derives this as the thermofield double. The radiation is thermal not because the hole forgot anything but because each quantum has a partner behind the horizon, and tracing out the partners of a pure entangled state gives a mixed state (Figure 4.2).

4.9 Evaporation, and the information problem

- *Mass loss.* A body of area A at temperature T radiates a luminosity $\sim AT^4$; with $A \sim G^2M^2$ and $T_H \sim 1/GM$,

$$\frac{\mathrm{d}M}{\mathrm{d}t} \;\sim\; -\,A\,T_H^4 \;\sim\; -\,\frac{b}{G^2M^2}\,, \tag{4.27}$$

  with $b \sim 10^{-4}$ a pure number summing the species weighted by their gray-body factors [29]. The lighter the hole, the faster it radiates.

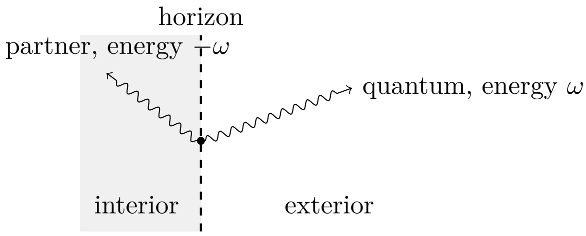


Figure 4.2: The pair structure of Hawking emission. Each escaping quantum of Killing energy ω is entangled with a partner of Killing energy $-\omega$ inside the horizon, where ∂_t is spacelike. The partner's infall decreases the mass. Both quanta have wavelength of order r_S, so the cartoon is not to be taken literally.

- *Lifetime.* Integrating,

$$M(t)^3 \;=\; M_0^3 - \frac{3b\,t}{G^2} \qquad \Longrightarrow \qquad \tau \;\sim\; \frac{G^2 M_0^3}{3b} \;\sim\; 10^{67}\text{ years for } M_0 = M_\odot, \tag{4.28}$$

  some 10^{57} ages of the universe. A primordial hole of initial mass 5×10^{11} kg would be finishing its evaporation now; no such burst has been seen. The number of emitted quanta is $N \sim M/T_H \sim GM^2 \sim S_{\rm BH}$, consistent with the radiation carrying off entropy of order the hole's.

- *Validity.* The fractional mass loss per light-crossing time is $\sim 1/GM^2 \lll 1$, so quantum fields on a quasi-static background are reliable until M approaches the Planck mass, in the last 10^{-43} s.

- *The information problem.* If the collapsing star is in a pure state and quantum mechanics applies throughout, the final radiation must be pure. The radiation computed here is thermal, mode by mode mixed, which within the present framework is no contradiction: the exterior is entangled with the partners behind the horizon and the global state is pure. The problem bites at the end of evaporation, when no interior remains for the partners: either unitarity fails, Hawking's original conclusion [30], or the radiation is only approximately thermal, with correlations among the quanta, invisible semiclassically, that encode the initial state. Making this quantitative requires a measure of entanglement between radiation and black hole, and is the subject of Part III.

Chapter 5

Gray Body Factors and Thermodynamic Instability

The thermal flux derived in Chapter 4 must still traverse the potential barrier surrounding the hole, so what reaches infinity is *gray-body* radiation, each mode weighted by a transmission probability $\Gamma_l(\omega)$. We state what the barrier does to the flux, quote the low-frequency result that the s-wave absorption cross section equals the horizon area, and use it to settle the long-wavelength puzzle of Chapter 2. Finally we confront the negative specific heat of the Schwarzschild black hole and preview the cure, a negative cosmological constant (Chapter 18). The gray-body factors themselves are derived in [1] and computed in [29].

5.1 Why the flux is not exactly Planckian

Each partial wave of a massless field in the Schwarzschild background is a two-dimensional field $\sigma(t, r_*)$ subject to a potential $V_l(r_*)$ which vanishes at the horizon ($r_* \to -\infty$) and at spatial infinity ($r_* \to +\infty$), with a barrier in between. The Hawking flux of Chapter 4 was derived by neglecting V_l. This fails in two ways.

1. Even for $l = 0$ the potential is nonzero, and T_H is of order the barrier height, so a typical Hawking quantum is reflected back into the hole with order-one probability.

2. For angular momentum l the barrier height grows like $l^2/(GM)^2$, so high partial waves are almost entirely suppressed. Without this suppression, equal thermal emission in infinitely many partial waves would give the black hole an *infinite* luminosity.

(The one semirealistic case with no barrier at all is the lowest partial wave of a massless charged fermion around a magnetically charged black hole, which is a free 1+1 dimensional fermion [1]; generically the barrier is there.)

5.2 The barrier and the gray-body factors

- *The barrier.* Inserting the Schwarzschild metric and a spherical harmonic into the action of a minimally coupled massless scalar, passing to the tortoise coordinate and substituting $\phi = \sigma/r$ gives, after an integration by parts,

$$\begin{aligned} I &= \int \mathrm{d}t\, \mathrm{d}r_* \left(\frac{1}{2}\,\dot\sigma^{\,2} - \frac{1}{2}\Big(\frac{\partial \sigma}{\partial r_*}\Big)^2 - \frac{1}{2}\, V_l(r)\, \sigma^2 \right), \\ V_l &= \Big(1 - \frac{2GM}{r}\Big) \left(\frac{l(l+1)}{r^2} + \frac{2GM}{r^3} \right), \end{aligned} \tag{5.1}$$

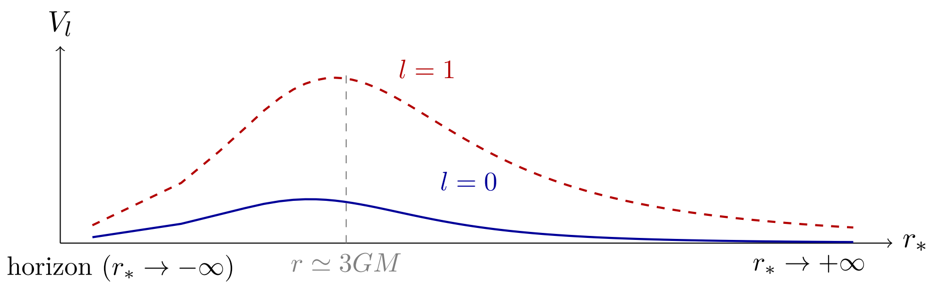


Figure 5.1: The Regge–Wheeler potential (5.1) for a massless scalar in the Schwarzschild geometry, drawn to scale in units $GM = 1$. Even the s-wave sees a barrier; for higher l the barrier peaks near the photon sphere $r = 3GM$.

the Regge–Wheeler potential (3.23) (Figure 5.1). The term $2GM/r^3$ survives at $l = 0$, where the barrier peaks at $r = \frac{8}{3}GM$ with $V_{\max} = \frac{27}{1024}(GM)^{-2}$; for large l it peaks at the photon sphere $r = 3GM$ with $V_{\max} \simeq l(l+1)/27(GM)^2$.

- *Thermal quanta sit at the top of the $l = 0$ barrier.*
$$\omega \;\sim\; T_H \;=\; \frac{1}{8\pi GM} \;\approx\; \frac{0.04}{GM} \;\sim\; \sqrt{V_{\max}(l=0)}\,, \tag{5.2}$$
and far below the barrier for every $l \geq 1$.

- *Transmission.* With $\sigma = \lambda(r_*)\mathrm{e}^{-\mathrm{i}\omega t}$ the mode equation is $(-\mathrm{d}^2/\mathrm{d}r_*^2 + V_l)\lambda = \omega^2\lambda$; a wave incident from the horizon is transmitted with amplitude $T(\omega)$ and reflected with amplitude $R(\omega)$, $|T|^2 + |R|^2 = 1$, and
$$\Gamma_l(\omega) \;\equiv\; |T(\omega)|^2 \tag{5.3}$$
is the probability that an outgoing quantum escapes.

- *Reciprocity.* The transmission amplitude for a wave sent in from infinity is the same $T(\omega)$, although V_l is not symmetric: the Wronskian of the two scattering solutions is constant, and evaluating it at both ends gives $T = \widetilde{T}$. The same $\Gamma_l(\omega)$ governs emission and absorption.

- *Detailed balance* [1]. Grant that the black hole is a thermal body at T_H; then it can sit in equilibrium with radiation at T_H [31], with emission and absorption balancing mode by mode. Absorption of the incident thermal flux is reduced by Γ_l, so emission must be too:

Gray-body emission. A Schwarzschild black hole loses energy to each massless bosonic field at the rate
$$\frac{\mathrm{d}E}{\mathrm{d}t} \;=\; \sum_{l=0}^{\infty}(2l+1)\int_0^{\infty}\frac{\mathrm{d}\omega}{2\pi}\,\Gamma_l(\omega)\,\frac{\omega}{\mathrm{e}^{\omega/T_H}-1}\,, \qquad T_H = \frac{1}{8\pi GM}, \tag{5.4}$$
where $\Gamma_l(\omega) = |T(\omega)|^2$ is the transmission probability through the Regge–Wheeler barrier and $(2l+1)$ counts the degenerate m modes.

Two consequences.

1. *Finiteness.* For $\omega \sim T_H$ and large l the barrier height $\sim l^2/(GM)^2$ far exceeds ω^2, and the WKB tunnelling factor $\Gamma_l \approx \exp[-2\int \mathrm{d}r_*\sqrt{V_l - \omega^2}]$ vanishes exponentially in l. The sum in (5.4) is dominated by $l = 0, 1$: gray-body factors are what makes the luminosity finite.

2. *Magnitude.* Each open channel radiates $\int_0^\infty \frac{\mathrm{d}\omega}{2\pi}\frac{\omega}{\mathrm{e}^{\omega/T_H}-1} = \frac{\pi}{12}T_H^2$, the one-dimensional Stefan–Boltzmann law, so $\mathrm{d}E/\mathrm{d}t \sim T_H^2 \sim 1/(GM)^2$, confirming (4.27) with a small coefficient computed by Page [29].

3. *Low and high frequencies.* The $\Gamma_l(\omega)$ are not elementary, but the s-wave can be done for $\omega r_S \ll 1$. Write the radial function as $R(r) = u_{\omega 0}/r$, so that (3.17) with the ansatz (3.18) at $\ell = 0$ reads
$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 f\,\frac{\mathrm{d}R}{\mathrm{d}r}\right) + \frac{\omega^2}{f}\,R \;=\; 0, \qquad f = 1 - \frac{r_S}{r}\,. \tag{5.5}$$
For $\omega r_S \ll 1$ there is an overlap region $r_S \ll r \ll 1/\omega$ in which two approximations to (5.5) hold at once: for $r \ll 1/\omega$ the ω^2 term is negligible and the equation is static, $(r^2 f R')' = 0$; for $r \gg r_S$ one may set $f = 1$ and the equation is the flat-space s-wave equation. Matching the two solutions in the overlap, with infall at the horizon, gives [32, 29, 23]
$$\Gamma_0(\omega) \;=\; 4\,\omega^2 r_S^2 \;=\; \frac{A\,\omega^2}{\pi}, \qquad \sigma_{\mathrm{abs}}(\omega) \;=\; \frac{\pi}{\omega^2}\,\Gamma_0(\omega) \xrightarrow[\omega\to 0]{} A, \qquad \Gamma_l \;\propto\; (\omega r_S)^{2l+2}, \tag{5.6}$$
with $A = 4\pi r_S^2$: soft quanta almost always bounce off the barrier, in either direction, and the low-frequency absorption cross section of a black hole is exactly its horizon area, for any stationary black hole. At high frequency geometric optics captures rays with impact parameter below $b_c = 3\sqrt{3}\,GM$, set by the photon sphere, so $\sigma_{\mathrm{abs}} \to \pi b_c^2 = \frac{27}{16}A$. Integrating (5.4) over the actual gray-body factors, a Schwarzschild black hole radiates a few percent of what a black body of area A at T_H would [29].

5.3 The more detailed argument

Detailed balance assumed what we want to prove, that the black hole is a thermal body. Hawking [26] and Witten [1] derive (5.4) without that assumption, by propagating the *observable* backwards in time rather than the state forwards.

1. The late-time detector measures $W^\dagger W$ with
$$W \;=\; \int_{\mathcal{S}'} \mathrm{d}\Sigma^\mu\, f\overleftrightarrow{\partial_\mu}\sigma, \tag{5.7}$$
the Klein–Gordon pairing of the field with a classical solution f localized at the detector, of frequency $\approx \omega$. The pairing is conserved (Proposition 3.1, Remark 3.1), so W may be evaluated on the early slice $\mathcal{S}$ of Chapter 4.

2. Propagated backwards, f meets the barrier and splits,
$$W \;=\; T'\,W_{\mathrm{hor}} \;+\; R'\,W_\infty, \tag{5.8}$$
with W_{hor} the near-horizon operator whose expectation values Chapter 4 found thermal, and W_∞ supported at large radius at early times.

3. The barrier is static, so reflection preserves the frequency: W_∞ is a superposition of in-mode annihilation operators, by the projection formula (3.8), and annihilates $|0_{\mathrm{in}}\rangle$. Hence $\langle W^\dagger W\rangle = |T'|^2\langle W_{\mathrm{hor}}^\dagger W_{\mathrm{hor}}\rangle$, and time reversal gives $|T'| = |T|$: the measured flux is Γ_l times the free thermal flux, now derived rather than inferred.

The argument also shows that at late times the fields outside the horizon settle into a universal state, thermal outgoing flux filtered by the barrier and vacuum inflow from $\mathscr{I}^-$, independent of the details of the collapse: the *Unruh state* [23] of Section 3.5.1, which describes a black hole formed by collapse and evaporating into empty space, whereas the Hartle–Hawking state of Section 7.8 describes the black hole in equilibrium in a thermal box.

5.4 Resolution of the long-wavelength puzzle

We can now pay the debt of Chapter 2. There the horizon entropy gained $E/T_H = 8\pi GME$ while the radiation lost $4E/3T$, so

$$\Delta S_{\rm gen} \;=\; \Big(\frac{1}{T_H} - \frac{4}{3T}\Big)E, \tag{5.9}$$

negative for $T \lesssim T_H$, i.e. for photons of wavelength longer than r_S. The resolution is that a black hole bathed in radiation colder than itself does not absorb net energy at all. The bookkeeping, mode by mode:

1. *Occupations.* In the mode (l, m, ω) the beam has $n_{\rm in} = 1/(e^{\omega/T} - 1)$. The outgoing mode is fed by the reflected beam (weight $1 - \Gamma$) and the transmitted near-horizon thermal flux (weight Γ),
$$n_{\rm out} \;=\; (1-\Gamma)\, n_{\rm in} \;+\; \Gamma\, \bar n_H, \qquad \bar n_H \;=\; \frac{1}{e^{\omega/T_H} - 1}\,, \tag{5.10}$$
and since two thermal sources combined on a beam-splitter-like barrier give a thermal mode, the outgoing mode is thermal with mean $n_{\rm out}$. (The stimulated emission of the black hole, with its $(n+1)$ factors, is contained in this description; for the entropy budget the thermal characterization suffices.) The mass changes by $\Delta M = \omega(n_{\rm in} - n_{\rm out}) = \Gamma\omega(n_{\rm in} - \bar n_H)$: the hole gains energy only from a beam hotter than itself.

2. *Entropies.* A thermal mode with mean occupation n has entropy $s(n) = (n+1)\ln(n+1) - n\ln n$, and
$$\begin{aligned}\Delta S_{\rm gen} \;&=\; \underbrace{\frac{\omega\,(n_{\rm in} - n_{\rm out})}{T_H}}_{\Delta S_{\rm BH}} \;+\; \underbrace{s(n_{\rm out}) - s(n_{\rm in})}_{\Delta S_{\rm out}} \;=\; \frac{F(n_{\rm in}) - F(n_{\rm out})}{T_H}\,,\\ F(n) \;&\equiv\; \omega\, n - T_H\, s(n),\end{aligned} \tag{5.11}$$
the mode free energy at the black hole's temperature.

3. *Convexity.*
$$F'(n) \;=\; \omega - T_H \ln\frac{n+1}{n}\,, \qquad F''(n) \;=\; \frac{T_H}{n(n+1)} \;>\; 0\,: \tag{5.12}$$
F is strictly convex with unique minimum at $n = \bar n_H$, the thermal occupation at the Hawking temperature. By (5.10), $n_{\rm out}$ lies on the segment between $n_{\rm in}$ and the minimizer, so
$$F(n_{\rm out}) \;\le\; (1-\Gamma)F(n_{\rm in}) + \Gamma F(\bar n_H) \;\le\; F(n_{\rm in}) \qquad\Longrightarrow\qquad \Delta S_{\rm gen} \;\ge\; 0, \tag{5.13}$$
with equality only if $\Gamma = 0$ or $n_{\rm in} = \bar n_H$, true equilibrium. The Generalized Second Law holds mode by mode, at every frequency and temperature.

The two regimes:

- *Hot beam, $T > T_H$* (Bekenstein's safe regime): then $n_{\rm in} > \bar n_H$ at every ω, the black hole absorbs net energy, and the area term dominates the budget, as in Chapter 2.
- *Cold beam, $T < T_H$* (the puzzle regime): then $n_{\rm in} < \bar n_H$ at every ω, so the hole is a net *emitter* in every single mode: $\Delta M < 0$ and the horizon *shrinks*. The area term in $S_{\rm gen}$ decreases—something Bekenstein, without Hawking radiation, could never allow—but the

radiation entropy increases by more. For pure spontaneous emission into an empty mode ($n_{\text{in}} = 0$, $\Gamma = 1$) the budget (5.11) evaluates to the tidy exact expression

$$\Delta S_{\text{gen}} \;=\; s(\bar n_H) - \frac{\omega\,\bar n_H}{T_H} \;=\; \ln(1+\bar n_H) \;>\; 0, \tag{5.14}$$

which for soft quanta $\omega \ll T_H$ grows as $\ln(T_H/\omega)$: exactly the long-wavelength modes that caused the paradox are the ones whose emission produces the most entropy per quantum.

The puzzle of Chapter 2 was thus not a flaw in the Generalized Second Law but a signal, correctly read by Hawking, that the black hole *must* radiate: assign the black hole the temperature T_H and the emission that goes with it, and the second law takes care of itself.

Remark 5.1 (Detailed balance and e^S)**.** There is an Einstein-coefficient way to see the equilibrium occupation that connects emission directly to state counting. Model the black hole as a system with $e^{S(M)}$ states at mass M, coupled to a radiation mode with n quanta through an interaction linear in the mode, $V = (a + a^\dagger) \otimes \mathcal{O}$, with $\mathcal{O}$ acting on the black hole. Compare the emission transition $M \to M - \omega$ with its reverse. Fermi's golden rule[1] gives

$$\frac{\text{rate}(M \to M-\omega,\; n \to n+1)}{\text{rate}(M-\omega \to M,\; n \to n-1)} \;=\; \frac{(n+1)\,e^{S(M-\omega)}}{n\,e^{S(M)}} \;=\; \frac{n+1}{n}\,e^{-\omega/T_H}, \tag{5.15}$$

using $S(M) - S(M-\omega) \simeq \omega\,\mathrm{d}S/\mathrm{d}M = \omega/T_H$ by the first law. In equilibrium the two mass levels are equally populated (the canonical weight $e^{S(M)-M/T}$ is stationary in M at $T = T_H$), so the mode occupation is stationary when the two rates are equal, i.e. precisely at $n = \bar n_H$: the Bose factors $(n+1)$ and n are the stimulated/spontaneous structure, and the Boltzmann factor is the *ratio of numbers of black hole microstates.* Read this way, Hawking emission is what any system with entropy $S(M) = A/4G$ and the first law is statistically obliged to do.

5.5 Thermodynamic instability

Equilibrium between a black hole and a thermal gas played a central role above, so we should ask whether such equilibrium is actually stable. In asymptotically flat space the answer is no, for two distinct reasons [1, 31].

- *Negative specific heat.* For Schwarzschild $E = M$ and $T = T_H = 1/8\pi GM$ decreases with the energy, so
$$C \;=\; \frac{\mathrm{d}E}{\mathrm{d}T} \;=\; -\,\frac{1}{8\pi G\,T^2} \;=\; -\,8\pi G M^2 \;<\; 0. \tag{5.16}$$
In an infinite bath at $T = T_H(M)$ both fluctuations run away: emitting more than it absorbs, the hole gets lighter and hotter and evaporates; absorbing more, it gets colder and grows without bound. The equilibrium is a saddle of the free energy.

- *No canonical ensemble.* Canonical energy fluctuations satisfy $\langle(\Delta E)^2\rangle = T^2 C$, which cannot be negative, and indeed
$$Z(\beta) \;\sim\; \int \mathrm{d}M\; e^{\,S(M)-\beta M} \;=\; \int \mathrm{d}M\; e^{\,4\pi G M^2 - \beta M} \;=\; \infty, \tag{5.17}$$

[1]The rate from an initial microstate $|i\rangle$ into a set of final states is $\Gamma = 2\pi \sum_f |\langle f|V|i\rangle|^2\,\delta(E_f - E_i)$. For $V = (a + a^\dagger) \otimes \mathcal{O}$ the matrix element factorizes: the mode contributes $|\langle n+1|a^\dagger|n\rangle|^2 = n+1$ for emission and $|\langle n-1|a|n\rangle|^2 = n$ for absorption, while the sum over final black hole microstates contributes their number, $e^{S(M-\omega)}$ for emission from mass M and $e^{S(M)}$ for absorption from mass $M-\omega$, times the squared matrix element of $\mathcal{O}$ between individual microstates, which is the same in both directions since $\mathcal{O}$ is hermitian (microreversibility).

the entropy growing quadratically with the energy. The microcanonical ensemble is healthy, and negative specific heat is shared by all self-gravitating systems; what is not sensible is “a black hole at temperature T in an infinite bath”.

- *No thermal gas in infinite flat space.* A ball of radiation of radius R at temperature T has mass $\sim T^4 R^3$ and lies inside its own Schwarzschild radius once
$$R \gtrsim \frac{1}{\sqrt{G}\,T^2}, \tag{5.18}$$
a Jeans-type bound: the infinite bath cannot exist in the first place.

- *The $G \to 0$ sense of equilibrium* [1]. The arguments of Section 5.2 are exact as $G \to 0$ at fixed r_S, the limit in which quantum field theory on a fixed background is exact: $T_H = 1/4\pi r_S$ stays fixed, the Jeans bound allows a bath of size $R \sim r_S^2/\sqrt{G}$ with $R/r_S \to \infty$, and a fluctuation of M by $\mathcal{O}(1)$ changes T_H by $\mathcal{O}(G)$, so the runaway takes a time of order $1/G^2$. A black hole cannot be in perfect equilibrium with radiation in flat space, but it can be arbitrarily close to it.

- *Outlook: boxes, and anti-de Sitter space.* Both instabilities are cured by a finite box, so that the total energy is fixed and the bath cannot exceed its Jeans size. One can enclose the hole in a reflecting cavity by hand [33], but nature provides a better box: a negative cosmological constant. Anti-de Sitter space confines the thermal gas, a large AdS black hole has positive specific heat and sits in true equilibrium with its radiation, and the canonical ensemble, meaningless in flat space by (5.17), is well defined. All of this is derived in Chapter 18.

Chapter 6

Rindler Space and the Unruh Effect

This chapter uncovers Hawking's physics in its simplest habitat, flat space. An observer confined to a Rindler wedge perceives the Minkowski vacuum as thermal at the Unruh temperature $T = a/2\pi$. We derive this from the Euclidean path integral by "making a cut", identifying the reduced density matrix of the wedge as $\mathrm{e}^{-2\pi K}$ with K the boost generator, a derivation valid for any interacting theory. We introduce the thermofield double, show that the Minkowski vacuum is the thermofield double of the two wedges at inverse temperature 2π, and transplant everything to the near-horizon region of a black hole, preparing the Euclidean technology of Chapter 7.

6.1 Rindler wedges and accelerated observers

In Minkowski space with coordinates $(t, x, \vec{y})$, the $\vec{y}$ being spectators throughout:

- *Wedges.* The right and left Rindler wedges are $\mathcal{R}_r = \{x > |t|\}$ and $\mathcal{R}_\ell = \{x < -|t|\}$, the domains of dependence of the two halves $\mathcal{S}_r = \{x \geq 0\}$ and $\mathcal{S}_\ell = \{x \leq 0\}$ of the $t = 0$ surface $\mathcal{S}$ (Figure 6.1). Every point of one wedge is spacelike to every point of the other.

- *Accelerated observers.* An observer who stays in $\mathcal{R}_r$ accelerates forever. The prototype is the hyperbola

$$t \;=\; L\,\sinh\frac{\tau}{L}, \qquad x \;=\; L\,\cosh\frac{\tau}{L}, \qquad a^\mu a_\mu \;=\; \frac{1}{L^2}, \tag{6.1}$$

  with τ the proper time and constant proper acceleration $a = 1/L$; L is the distance of closest approach to the origin.

- *Rindler coordinates.* With $t = \rho\sinh\eta$, $x = \rho\cosh\eta$,

$$\mathrm{d}s^2 \;=\; -\,\rho^2\,\mathrm{d}\eta^2 \;+\; \mathrm{d}\rho^2 \;+\; \mathrm{d}\vec{y}^{\,2}, \tag{6.2}$$

  flat space in disguise and exactly the near-horizon geometry (1.11), with η in the role of κt and ρ the proper distance from the horizon. The observer sits at $\rho = L$ with $\tau = L\eta$; her acceleration $1/\rho$ diverges as $\rho \to 0$, like that of an observer hovering at a horizon.

- *The boost and its horizons.* The translation of η is generated by the boost Killing vector

$$\xi \;=\; \partial_\eta \;=\; x\,\partial_t + t\,\partial_x, \tag{6.3}$$

  future-directed timelike in $\mathcal{R}_r$, past-directed in $\mathcal{R}_\ell$, and null on $x = \pm t$: the null planes are Killing horizons of the boost, meeting at the *bifurcation surface* $\Sigma = \{x = t = 0\}$, also called the entangling surface. This is the flat-space skeleton of the Kruskal diagram, $\mathcal{R}_r$ and $\mathcal{R}_\ell$ being regions I and III and Σ the bifurcation sphere; Section 6.6 makes the correspondence precise.

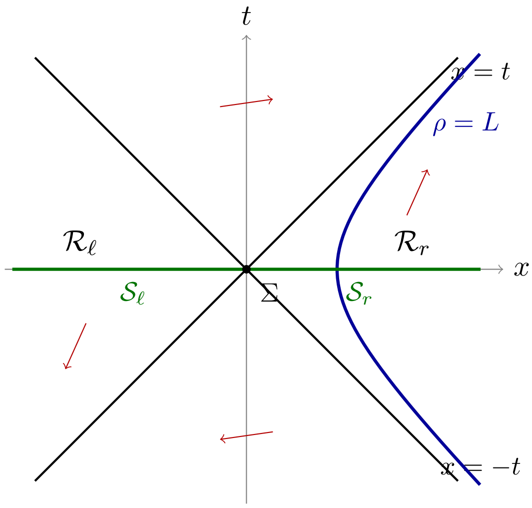


Figure 6.1: The left and right Rindler wedges $\mathcal{R}_\ell$ and $\mathcal{R}_r$ are the domains of dependence of the half-surfaces $\mathcal{S}_\ell$ and $\mathcal{S}_r$. The diagonal null lines are the past and future horizons of observers confined to one wedge; they intersect at the bifurcation surface Σ. The blue hyperbola is a uniformly accelerated worldline at $\rho = L$; the red arrows indicate the flow of the boost generator K, future-directed in $\mathcal{R}_r$ and past-directed in $\mathcal{R}_\ell$.

- *The question.* What does the Minkowski vacuum $|\Omega\rangle$ look like to an observer who measures only in $\mathcal{R}_r$? Her measurements are determined by the field on $\mathcal{S}_r$, so the answer is encoded in the *reduced density matrix* obtained from $|\Omega\rangle\langle\Omega|$ by tracing out the degrees of freedom on $\mathcal{S}_\ell$.

- *Density matrices*, for readers meeting them for the first time (Section 19.2 develops them systematically). A *density matrix* ρ is a positive operator with $\operatorname{Tr}\rho = 1$; expectation values in the state it describes are $\langle\mathcal{O}\rangle = \operatorname{Tr}(\rho\,\mathcal{O})$, and a pure state is the special case $\rho = |\Psi\rangle\langle\Psi|$, any other ρ being called *mixed.* If the Hilbert space factorizes, $\mathcal{H} = \mathcal{H}_r \otimes \mathcal{H}_\ell$, the reduced density matrix $\rho_r = \operatorname{Tr}_\ell |\Omega\rangle\langle\Omega|$ is the unique operator on $\mathcal{H}_r$ such that

$$\operatorname{Tr}\left(\rho_r\,\mathcal{O}_r\right) \;=\; \langle\Omega|\,\mathcal{O}_r \otimes \mathbf{1}_\ell\,|\Omega\rangle \qquad \text{for every operator } \mathcal{O}_r \text{ on } \mathcal{H}_r : \tag{6.4}$$

  it encodes everything an observer confined to $\mathcal{R}_r$ can measure, and it is in general mixed even though $|\Omega\rangle$ is pure. In field theory the factorization of $\mathcal{H}$ across Σ is only formal (Remark 6.1), but ρ_r acquires a precise meaning through the Euclidean path integral, which is also the most efficient way to compute it.

6.2 Making the cut: the vacuum as a Euclidean path integral

Consider a real scalar ϕ; nothing depends on this choice, and the argument applies to any quantum field theory. A state on $\mathcal{S} = \{t = 0\}$ is a wavefunctional $\Psi(\phi(\vec{x}))$.

- *Preparing the vacuum.* With $t = -\mathrm{i}t_E$, the Euclidean path integral over the half-space $t_E \le 0$ with the field fixed on $t_E = 0$ computes the vacuum wavefunctional,

$$\Omega(\phi(\vec{x})) \;=\; \int_{\substack{\phi(t_E,\vec{x})\,\to\,0,\;\; t_E\to-\infty \\ \phi(0,\vec{x})\,=\,\phi(\vec{x})}} D\phi\;\; \mathrm{e}^{-S_E[\phi]}, \tag{6.5}$$

  because cutting it at $t_E = -T$ shows it equals $\lim_{T\to\infty}\langle\phi|\mathrm{e}^{-TH}|\chi\rangle$ for any χ with $\langle\Omega|\chi\rangle \neq 0$, and e^{-TH} projects onto the ground state (Figure 6.2(a)).

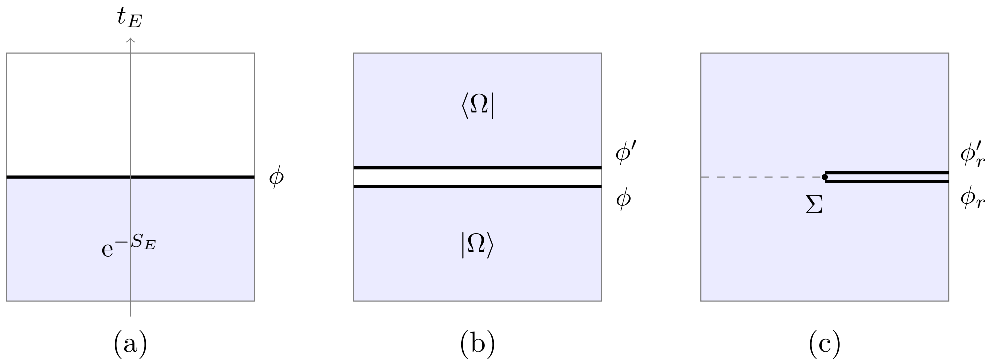


Figure 6.2: (a) A Euclidean path integral on the half-space $t_E \leq 0$ with fixed boundary values ϕ prepares the vacuum wavefunctional $\Omega(\phi)$. (b) Two such half-spaces, with independent boundary values ϕ and ϕ' on the two sides of the hyperplane $t_E = 0$ (drawn separated for visibility), represent the pure-state density matrix $|\Omega\rangle\langle\Omega|$. (c) Gluing the two copies of the negative x axis and integrating over the common boundary value traces out $\mathcal{H}_\ell$: the result is a path integral on the plane with a cut along $t_E = 0$, $x > 0$, which computes the reduced density matrix $\rho_r(\phi_r; \phi'_r)$ of the right wedge. The tip of the cut is the bifurcation surface Σ.

- *The pure-state density matrix.* The upper half-space $t_E \geq 0$ with boundary values ϕ' computes the bra $\langle\Omega|$. Together, the path integral on $\mathbb{R}^4$ with the hyperplane $t_E = 0$ cut open, and independent values ϕ below and ϕ' above the cut, is

$$\rho(\phi; \phi') \;=\; \Omega(\phi)\,\overline{\Omega(\phi')} \tag{6.6}$$

  (Figure 6.2(b)).

- *Tracing out the left half.* Split $\phi = (\phi_\ell, \phi_r)$ according to $\mathcal{S} = \mathcal{S}_\ell \cup \mathcal{S}_r$. The partial trace over $\mathcal{H}_\ell$ sets $\phi_\ell = \phi'_\ell$ and integrates,

$$\rho_r(\phi_r; \phi'_r) \;=\; \int D\phi_\ell \;\; \Omega(\phi_\ell, \phi_r)\,\overline{\Omega(\phi_\ell, \phi'_r)}, \tag{6.7}$$

  which *glues the cut back together* along $x < 0$. What remains is the path integral on $\mathbb{R}^4$ cut only along the half-hyperplane $t_E = 0$, $x > 0$, with independent values ϕ_r below and ϕ'_r above (Figure 6.2(c)): this cut-plane path integral is the reduced density matrix of the right wedge. Exchanging left and right gives ρ_ℓ from the cut along $x < 0$.

Remark 6.1 (How formal is the factorization?)**.** In continuum quantum field theory $\mathcal{H} = \mathcal{H}_\ell \otimes \mathcal{H}_r$ is not literally true: short-distance fluctuations near $x = 0$ obstruct it, and a lattice cutoff that restores it breaks the Lorentz invariance the derivation relies on. We proceed in the informal spirit of the physics literature; the results are correct and are in fact a theorem.[1] The divergences of entanglement entropy in Chapter 20 are the same issue.

[1]Bisognano and Wichmann [34] proved, for any quantum field theory obeying the Wightman axioms [35] and any observables A, B localized in $\mathcal{R}_r$, with $A(\eta) = e^{i\eta K} A\, e^{-i\eta K}$ the boosted observable, that

$$\langle\Omega|\, A(\eta)\, B\,|\Omega\rangle \;=\; \langle\Omega|\, B\, A(\eta + 2\pi i)\,|\Omega\rangle,$$

the right-hand side defined by analytic continuation in the strip $0 < \operatorname{Im}\eta < 2\pi$. By (4.23) this is the KMS condition at $\beta = 2\pi$ with the boost generator K as Hamiltonian: the content of (6.9) below, phrased without factorizing the Hilbert space, in terms of vacuum correlators of local operators only. In Tomita–Takesaki modular theory, every state and algebra of observables define a modular flow for which the state is KMS at inverse temperature 1, and the theorem says that for the vacuum and the wedge algebra the modular flow is the boost rescaled by 2π [36]. The theorem predates Unruh's paper by months; its extension to black hole horizons is due to Sewell [37].

6.3 Boosts, rotations, and the Rindler density matrix

The cut-plane path integral has a symmetry, rotations of the Euclidean (x, t_E) plane about the origin, and this identifies ρ_r as an operator.

- *The boost generator.* The conserved charge generating boosts of the (t, x) plane weights the energy density by x, and splits into wedge contributions because x changes sign at Σ:

$$\begin{aligned} K \;&=\; \int_{\mathcal{S}} \mathrm{d}x\, \mathrm{d}\vec{y}\;\, x\, T_{00}(x,\vec{y}) \;=\; K_r - K_\ell, \\ K_r \;&=\; \int_{\mathcal{S}_r} \mathrm{d}x\, \mathrm{d}\vec{y}\, x\, T_{00}, \qquad K_\ell \;=\; \int_{\mathcal{S}_\ell} \mathrm{d}x\, \mathrm{d}\vec{y}\, |x|\, T_{00}. \end{aligned} \tag{6.8}$$

  K_r is built from fields in $\mathcal{S}_r$, generates the translation of Rindler time η, i.e. forward evolution within $\mathcal{R}_r$, and commutes with everything in $\mathcal{R}_\ell$; the minus sign makes K_ℓ generate forward evolution in $\mathcal{R}_\ell$ too, since the boost flow is past-directed there. The additive ambiguity in T_{00} is fixed by $\langle T_{00}\rangle = 0$, without which the integrals diverge.

- *From boosts to rotations.* $\mathrm{e}^{-\mathrm{i}\eta K_r}$ boosts the right wedge by rapidity η; for imaginary rapidity $\eta = -\mathrm{i}\theta$ the boost is a rotation of the Euclidean (t_E, x) plane by θ and the operator is $\mathrm{e}^{-\theta K_r}$. Hence the Euclidean path integral on an angular wedge $\{0 \le \arg(x + \mathrm{i}t_E) \le \theta\}$ with fields fixed on its two edges computes the matrix elements of $\mathrm{e}^{-\theta K_r}$ (Figure 6.3(a)), as a Euclidean slab of height T computes those of e^{-TH}.

- *Opening the wedge to* 2π. The wedge of angle 2π is the plane cut along the positive x axis with independent values on the two lips of the cut, which is exactly the path integral for ρ_r (Figure 6.3(b)).

> The reduced density matrix of the Minkowski vacuum restricted to the right Rindler wedge is
> $$\rho_r \;=\; \frac{1}{Z}\,\mathrm{e}^{-2\pi K_r}, \qquad Z = \mathrm{Tr}\;\mathrm{e}^{-2\pi K_r}, \tag{6.9}$$
> a *thermal* state at inverse temperature 2π with respect to the boost Hamiltonian K_r. With the energy density renormalized so that $\langle T_{00}\rangle = 0$, one has $Z = 1$. The same holds for the left wedge: $\rho_\ell = \mathrm{e}^{-2\pi K_\ell}/Z$.

- *Normalization.* $\mathrm{Tr}\;\mathrm{e}^{-2\pi K_r}$ glues the two edges of the 2π wedge together: it is the path integral on the uncut plane, equal to 1 once the vacuum energy is subtracted.

- *Generality.* Nothing was assumed about the theory beyond the existence of a Euclidean path integral, its rotation symmetry and the relation between boosts and Euclidean angles. The result holds for interactions of any strength, in any dimension, in contrast to the free-field mode analysis of Chapter 4, and Chapter 7 inherits the same generality for the Euclidean black hole.

6.4 The Unruh temperature

For the observer at $\rho = L$, with $a = 1/L$ and $\tau = L\eta$, the boost acts on her worldline as $\partial_\eta = L\,\mathrm{d}/\mathrm{d}\tau$: K_r is L times her natural Hamiltonian. For observations localized near $x = L$, (6.8) gives

$$K_r \;=\; \int_{\mathcal{S}_r} \mathrm{d}x\,\mathrm{d}\vec{y}\;\, x\, T_{00} \;\approx\; L \int_{\mathcal{S}_r} \mathrm{d}x\,\mathrm{d}\vec{y}\;\, T_{00} \;=\; L\, H_{\mathrm{loc}}, \qquad \text{so} \qquad \rho_r \;\approx\; \frac{1}{Z}\,\mathrm{e}^{-2\pi L\, H_{\mathrm{loc}}}, \tag{6.10}$$

with H_{loc} the generator of her proper-time translations: she is immersed in a thermal bath at temperature $1/2\pi L$.

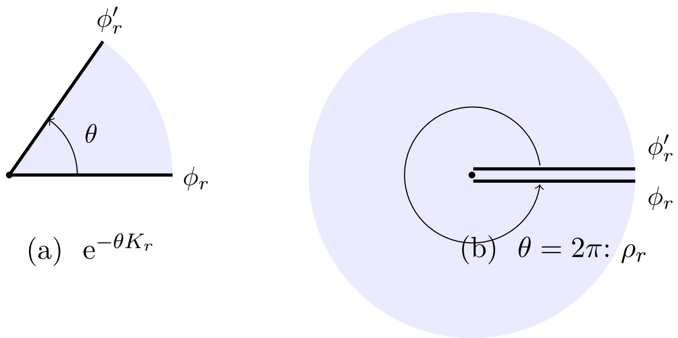


Figure 6.3: (a) The Euclidean path integral on an angular wedge of opening angle θ, with boundary values fixed on the two edges, computes the matrix elements of $\mathrm{e}^{-\theta K_r}$. (b) Opening the wedge all the way to $\theta = 2\pi$ reproduces the cut plane of Figure 6.2(c), i.e. the reduced density matrix ρ_r. Setting $\theta = \pi$ instead gives the half-plane of Figure 6.2(a), which prepares the vacuum, the key to Section 6.5.1.

An observer moving through the Minkowski vacuum with constant proper acceleration a observes thermal correlations at the *Unruh temperature* [23]

$$T_U \;=\; \frac{a}{2\pi}\,. \tag{6.11}$$

Restoring units, $T_U = \hbar a/(2\pi c\, k_B) \approx 4 \times 10^{-20}\,\mathrm{K}$ for $a = 9.8$ m/s^2.

- *Euclidean shortcut.* Continue the orbit (6.1) to Euclidean signature, $t = \mathrm{i}t_E$, $\tau = \mathrm{i}\tau_E$:

$$t_E \;=\; L\,\sin\frac{\tau_E}{L}, \qquad x \;=\; L\,\cos\frac{\tau_E}{L}, \tag{6.12}$$

  a circle of circumference $2\pi L$: the hyperbola of Lorentzian acceleration becomes periodic motion in Euclidean proper time. The Euclidean plane has no distinguished origin of angle, so correlators along the orbit inherit the $2\pi L$ periodicity in imaginary proper time, which by the KMS condition (4.23) is thermality at $1/2\pi L = a/2\pi$.

- *Detector interpretation.* The operational content of (6.11) is that a physical detector carried by the observer thermalizes. The standard model is the *Unruh–DeWitt detector* [23, 38]: a quantum system with two energy levels, a ground state $|0\rangle$ and an excited state $|1\rangle$ separated by a gap ω (an atom with one relevant transition), carried along the worldline $x(\tau)$ and coupled to the field by a term $g\,\mu(\tau)\,\phi(x(\tau))$ in the action, where μ is an operator flipping the detector between its two levels and g is a small coupling. Because the coupling is linear in ϕ, a transition of the detector is accompanied by the emission or absorption of one field quantum, as in the interaction of an atom with the electromagnetic field. The two quantities that characterize the detector are

$$\Gamma_\uparrow \;=\; \frac{g^2\omega}{2\pi}\,\frac{1}{\mathrm{e}^{2\pi\omega/a}-1}, \qquad \Gamma_\downarrow \;=\; \frac{g^2\omega}{2\pi}\Big(1+\frac{1}{\mathrm{e}^{2\pi\omega/a}-1}\Big), \tag{6.13}$$

  the *excitation rate* $\Gamma_\uparrow$, the probability per unit proper time that the detector jumps from $|0\rangle$ to $|1\rangle$ by absorbing energy ω from the field, and the *de-excitation rate* $\Gamma_\downarrow$, the probability per unit proper time that it drops from $|1\rangle$ to $|0\rangle$ by emitting a quantum. Both follow from first-order perturbation theory in g, the rate being the Fourier transform at frequency $\pm\omega$ of the vacuum two-point function of ϕ evaluated along the hyperbola (6.1); the computation is in [4, §3.3]. Read the result as follows.

- For an inertial detector, $a \to 0$, $\Gamma_\uparrow = 0$: there is nothing to absorb in the vacuum. $\Gamma_\downarrow = g^2\omega/2\pi$ is ordinary spontaneous emission.
- The accelerated detector is excited at the rate of a detector immersed in a Planck distribution at temperature $a/2\pi$, and de-excites by spontaneous plus stimulated emission, the 1 and the Bose factor in $\Gamma_\downarrow$.
- The ratio is the detailed balance of a system in a heat bath,

$$\frac{\Gamma_\uparrow}{\Gamma_\downarrow} \;=\; \mathrm{e}^{-2\pi\omega/a} \;=\; \mathrm{e}^{-\omega/T_U} \tag{6.14}$$

(compare (5.15)): stationarity of the level populations, $p_0\Gamma_\uparrow = p_1\Gamma_\downarrow$, gives $p_1/p_0 = \mathrm{e}^{-\omega/T_U}$, a Boltzmann distribution at T_U. The energy for the excitations is supplied by the agent maintaining the acceleration.

- *Tolman's law* [39, 40]. The temperature $1/2\pi L$ depends on the distance from Σ. In a static spacetime $\mathrm{d}s^2 = -g_{00}\,\mathrm{d}t^2 + \dots$, equilibrium with respect to t means correlators periodic in imaginary t with one period $1/T_\infty$, which a static observer with $\mathrm{d}\tau = \sqrt{-g_{00}}\,\mathrm{d}t$ reads as a local temperature

$$T_{\mathrm{loc}}(x)\,\sqrt{-g_{00}(x)} \;=\; T_\infty \;=\; \text{const.} \tag{6.15}$$

In Rindler space $\sqrt{-g_{\eta\eta}} = \rho$ and $T_{\mathrm{loc}}\rho = 1/2\pi$: the invariant statement is thermality at inverse temperature 2π with respect to the boost, and $T_{\mathrm{loc}} = 1/2\pi\rho \to 0$ at infinity. For a black hole the redshift is finite and the temperature at infinity is Hawking's (Section 6.6).

6.5 The thermofield double

The reduced state on either wedge is thermal while the global state is pure. This structure has a name.

- *Purifications.* A purification of a density matrix ρ on $\mathcal{H}$ is a pure state $\Psi \in \mathcal{H}\otimes\mathcal{H}'$ with $\rho = \mathrm{Tr}_{\mathcal{H}'}\,|\Psi\rangle\langle\Psi|$. Purifications always exist, and there is a canonical one.

- *Operators as states.* To a linear map $\mathcal{V} = \sum_{ij} v_{ij}|i\rangle\langle j|$ on $\mathcal{K}$ associate the vector

$$\Psi_{\mathcal{V}} \;=\; \sum_{i,j} v_{ij}\,|i\rangle\otimes|j\rangle' \;\in\; \mathcal{K}\otimes\mathcal{K}', \tag{6.16}$$

with $\mathcal{K}'$ the complex-conjugate space, whose kets $|j\rangle'$ correspond to the bras $\langle j|$; this makes the map basis-independent. Writing out $|\Psi_{\mathcal{V}}\rangle\langle\Psi_{\mathcal{V}}| = \sum_{i,k,j,l} v_{ik}\bar{v}_{jl}\,|i\rangle\langle j|\otimes|k\rangle'\langle l|'$ and tracing over $\mathcal{K}'$ with an orthonormal basis $\{|m\rangle'\}$, the factor $\langle m|'|k\rangle'\langle l|'|m\rangle' = \delta_{km}\delta_{lm}$ sets $k = l = m$, so

$$\mathrm{Tr}_{\mathcal{K}'}\,|\Psi_{\mathcal{V}}\rangle\langle\Psi_{\mathcal{V}}| \;=\; \sum_m\sum_{i,j} v_{im}\,\bar{v}_{jm}\,|i\rangle\langle j| \;=\; \sum_{i,j}\left(\mathcal{V}\mathcal{V}^\dagger\right)_{ij}|i\rangle\langle j| \;=\; \mathcal{V}\mathcal{V}^\dagger, \tag{6.17}$$

since $(\mathcal{V}^\dagger)_{mj} = \bar{v}_{jm}$. Hence if $\mathrm{Tr}\,\mathcal{V}\mathcal{V}^\dagger = 1$, $\Psi_{\mathcal{V}}$ is a purification of $\mathcal{V}\mathcal{V}^\dagger$.

- *The canonical purification.* Given ρ, take $\mathcal{V} = \rho^{1/2}$, the positive square root, so that $\mathcal{V}\mathcal{V}^\dagger = \rho$:

$$\Psi_{\rho^{1/2}} \;=\; \sum_{i,j}\left(\rho^{1/2}\right)_{ij}|i\rangle\otimes|j\rangle' \;\in\; \mathcal{H}\otimes\mathcal{H}', \qquad \mathrm{Tr}_{\mathcal{H}'}\,|\Psi_{\rho^{1/2}}\rangle\langle\Psi_{\rho^{1/2}}| \;=\; \rho, \tag{6.18}$$

the *canonical purification* of ρ.

For the thermal state $\rho_\beta = Z^{-1}\sum_n \mathrm{e}^{-\beta E_n}|n\rangle\langle n|$ the square root is diagonal in the energy basis, and (6.16) gives:

The *thermofield double state* of a system with Hamiltonian H at inverse temperature β is the entangled pure state

$$|\Psi_{\mathrm{TFD}}\rangle \;=\; \frac{1}{\sqrt{Z}}\sum_n \mathrm{e}^{-\beta E_n/2}\,|n\rangle\otimes|n\rangle' \;\;\in\;\; \mathcal{H}\otimes\mathcal{H}', \tag{6.19}$$

in the doubled Hilbert space. Tracing out either factor leaves the thermal density matrix at inverse temperature β on the other: $\mathrm{Tr}_{\mathcal{H}'}\,|\Psi_{\mathrm{TFD}}\rangle\langle\Psi_{\mathrm{TFD}}| = \rho_\beta$.

- *Two copies.* For a system with an antilinear time-reversal symmetry squaring to one, $\mathcal{H}'$ can be identified with $\mathcal{H}$ and one speaks of two copies. All thermal expectation values in one copy are reproduced by pure-state expectation values in Ψ_{TFD}, the entanglement with the second copy replacing classical probability.

- *Oscillator warm-up.* For $H = \omega\,\mathbf{a}^\dagger\mathbf{a}$ and a second copy with operators $\mathbf{a}'$,

$$\begin{aligned} \Psi_{\mathrm{TFD}} \;&=\; \frac{1}{\sqrt{Z}}\sum_{n=0}^{\infty} \mathrm{e}^{-n\beta\omega/2}\,|n\rangle\otimes|n\rangle', \\ \left(\mathbf{a} - \mathrm{e}^{-\beta\omega/2}\,\mathbf{a}'^\dagger\right)\Psi_{\mathrm{TFD}} \;&=\; \left(\mathbf{a}^\dagger - \mathrm{e}^{\beta\omega/2}\,\mathbf{a}'\right)\Psi_{\mathrm{TFD}} \;=\; 0, \end{aligned} \tag{6.20}$$

  To check the first: $\mathbf{a}$ on $|n\rangle\otimes|n\rangle'$ gives $\sqrt{n}\,|n-1\rangle\otimes|n\rangle'$, while $\mathbf{a}'^\dagger$ gives $\sqrt{n+1}\,|n\rangle\otimes|n+1\rangle'$; shifting $n\to n+1$ in the first sum, the two contributions cancel because of the relative weight $\mathrm{e}^{-\beta\omega/2}$ between adjacent levels. Conversely the two conditions are first-order recursions in the double occupation basis and determine the state uniquely: the thermofield double is *the* state annihilated by the $\mathrm{e}^{\pm\beta\omega/2}$-weighted mixtures of the two copies' ladder operators.

6.5.1 The Minkowski vacuum as a thermofield double

- *The half-plane prepares the vacuum.* With $Z = 1$, $\rho_r^{1/2} = \mathrm{e}^{-\pi K_r}$ is computed by the wedge of opening angle π, a half-plane. The path integral on the lower half-plane with boundary values on the positive and negative x axes, i.e. on $\mathcal{S}_r$ and $\mathcal{S}_\ell$, is exactly (6.5) (Figure 6.2(a)). Under the operator–state correspondence (6.16) the operator $\rho_r^{1/2}$ on $\mathcal{H}_r$ is a vector in a doubled space, and the path integral identifies this vector as $\Omega\in\mathcal{H}_\ell\otimes\mathcal{H}_r$.

The Minkowski vacuum, viewed as an entangled state of the two Rindler wedges, is the thermofield double state of the boost Hamiltonian at inverse temperature $\beta = 2\pi$:

$$|\Omega\rangle \;=\; \frac{1}{\sqrt{Z}}\sum_n \mathrm{e}^{-\pi k_n}\,|n\rangle_\ell\otimes|n\rangle_r\,, \tag{6.21}$$

where $|n\rangle_r$ are (formal) eigenstates of K_r with eigenvalues k_n and $|n\rangle_\ell$ their mirrors in the left wedge. Tracing out either wedge yields the boost-thermal state (6.9).

- *CRT.* The thermofield double lives in $\mathcal{H}_r\otimes\mathcal{H}_r'$, the vacuum in $\mathcal{H}_\ell\otimes\mathcal{H}_r$. The bridge is the antiunitary CRT symmetry of every relativistic theory, combining charge conjugation, the reflection $x\to -x$ and time reversal, which maps the right wedge onto the left; being antilinear it identifies $\mathcal{H}_\ell$ with $\mathcal{H}_r'$, exactly as (6.21) requires. The black hole version, the Hartle–Hawking state as the thermofield double of the two exteriors, anticipated by Israel [41], is Section 7.8.

- *Boost modes.* For a free field the same conclusion follows at the operator level [23], and in detail in section 5 of [1]: the Unruh modes, combinations of left- and right-wedge Rindler modes of positive Minkowski frequency, characterized by analyticity in the lower half of the complex null-coordinate plane, show that the vacuum is annihilated by the $e^{\pm\pi\omega}$-weighted mixtures of the two wedges' mode operators, the field-theory version of (6.20). The path-integral argument holds for any interacting theory.

6.6 From Rindler to black holes

- *A horizon is locally Rindler.* For $ds^2 = -f\,dt^2 + dr^2/f + r^2 d\Omega^2$ with a non-extremal horizon at r_h, the near-horizon form (1.11) is

$$ds^2 \;\simeq\; -\kappa^2\rho^2\,dt^2 \;+\; d\rho^2 \;+\; r_h^2\,d\Omega^2, \tag{6.22}$$

  the Rindler metric (6.2) with $\eta = \kappa t$ times a sphere invisible to short-distance physics. Rindler space is the $M \to \infty$ limit of the Schwarzschild horizon.

- *Local temperature.* A static observer at proper distance ρ has $a = f'/2\sqrt{f} \simeq 1/\rho$ by (1.8) and (1.9). If the fields are in the state that looks like the local Minkowski vacuum near the horizon, the Hartle–Hawking state constructed in Chapter 7, she measures

$$T_{\rm loc}(\rho) \;=\; \frac{a}{2\pi} \;\simeq\; \frac{1}{2\pi\rho} \;=\; \frac{\kappa}{2\pi\sqrt{f(r)}}\,, \tag{6.23}$$

  the last form using $\sqrt{f} = \kappa\rho$ near the horizon and remaining exact for the Hartle–Hawking state at all r.

- *Tolman's law gives Hawking's temperature.* By (6.15), $T_{\rm loc}\sqrt{f}$ is constant along the Killing flow and equals the temperature at infinity where $f \to 1$:

$$T_\infty \;=\; T_{\rm loc}\sqrt{f} \;=\; \frac{\kappa}{2\pi} \;=\; \frac{1}{8\pi GM} \;=\; T_H. \tag{6.24}$$

  In Rindler space the redshift grows without bound and the temperature at infinity is zero; for a black hole $\sqrt{f}$ saturates at 1 and a finite thermal flux survives at infinity, at the temperature of Chapter 4. The Hawking effect is the Unruh effect seen through a gravitational redshift.

- *The parallel goes deeper.* The eternal Schwarzschild geometry (Figure 1.1) has two exteriors glued across a bifurcation sphere, as $\mathcal{R}_r$ and $\mathcal{R}_\ell$ across Σ; the Schwarzschild time translation acts on the Kruskal plane as a boost with $\eta = \kappa t$; and the natural vacuum-like state on the full geometry should be the thermofield double (6.19) of the two exteriors at $\beta_H = 2\pi/\kappa$ [41], which Chapter 7 confirms by a Euclidean path integral. What is still missing is the black hole's own thermodynamics, its free energy and entropy, from first principles: the Euclidean approach of Gibbons and Hawking, to which we turn next.

Chapter 7

The Euclidean Approach to Black Hole Thermodynamics

This chapter completes Part I by deriving black hole thermodynamics from the Euclidean gravitational path integral. Continuing Schwarzschild to Euclidean signature produces the "cigar", smooth only if Euclidean time has period $\beta = 8\pi GM$: the Hawking temperature emerges from pure geometry. Evaluating the gravitational action on this saddle gives the free energy and $S = A/4G$, and a second computation exhibits $A/4G$ as the response of the action to a conical deficit at the horizon. We then face the question the formalism raises but does not answer, whether this entropy counts states, and close by constructing the Hartle–Hawking state on half the cigar, the thermofield double of the two exterior regions.

7.1 Thermal partition functions and the Euclidean time circle

- *The trace as a periodic path integral.* For a system with Hamiltonian H, inserting position eigenstates in the canonical partition function,

$$Z(\beta) \;=\; \mathrm{Tr}\, \mathrm{e}^{-\beta H} \;=\; \int \mathrm{d}q\; \langle q|\mathrm{e}^{-\beta H}|q\rangle, \tag{7.1}$$

  and the matrix element is the transition amplitude continued to imaginary time $t = -\mathrm{i}\tau$, a Euclidean path integral from q at $\tau = 0$ to q' at $\tau = \beta$. Setting $q' = q$ and integrating sums over *periodic* trajectories:

$$Z(\beta) \;=\; \int_{q(\tau+\beta)=q(\tau)} Dq\; \mathrm{e}^{-S_E[q]}. \tag{7.2}$$

- *Temperature is the size of the time circle.* For a field theory on space X the thermal partition function is the Euclidean path integral on

$$S^1_\beta \times X, \tag{7.3}$$

  with periodic boundary conditions around the circle for bosons and antiperiodic for fermions. Correlators in this ensemble are periodic in imaginary time with period β, the KMS property of Chapter 6, where the circle was the angular direction of the Euclidean plane and $\beta = 2\pi$ in boost units.

- *Thermodynamics from Z.*

$$\begin{aligned} E \;&=\; -\frac{\partial}{\partial\beta}\log Z, \qquad & S \;&=\; \left(1-\beta\frac{\partial}{\partial\beta}\right)\log Z \;=\; \beta E + \log Z, \\ F \;&=\; -\frac{1}{\beta}\log Z \;=\; E - \frac{S}{\beta}, \end{aligned} \tag{7.4}$$

the second following from $S = -\operatorname{Tr}\rho\log\rho$ on the Gibbs state $\rho = \mathrm{e}^{-\beta H}/Z$, whose logarithm is $-\beta H - \log Z$.

The bold step of Gibbons and Hawking was to apply this to the gravitational field itself: to seek a Euclidean geometry that plays the role of (7.3), and to evaluate the path integral over metrics in a saddle-point approximation around it.

7.2 Euclidean Schwarzschild: the cigar

Continue the Schwarzschild solution to imaginary time, $t = -\mathrm{i}\tau$,

$$\mathrm{d}s_E^2 \;=\; f(r)\,\mathrm{d}\tau^2 \;+\; \frac{\mathrm{d}r^2}{f(r)} \;+\; r^2\,\mathrm{d}\Omega^2, \qquad f(r) = 1 - \frac{2GM}{r}, \qquad \tau \;\sim\; \tau + \beta, \tag{7.5}$$

a positive-definite metric for $r > 2GM$, which at large r approaches the thermal cylinder $S^1_\beta \times \mathbb{R}^3$: the candidate saddle of the thermal ensemble containing a black hole. Smoothness at $r = 2GM$ fixes β.

1. *Zoom in on the horizon.* With $f \simeq 2\kappa(r - r_S)$, $\kappa = f'(r_S)/2 = 1/4GM$, and the proper radial distance

$$\rho \;=\; \int_{r_S}^{r} \frac{\mathrm{d}r'}{\sqrt{f(r')}} \;\simeq\; \sqrt{\frac{2(r-r_S)}{\kappa}} \qquad \Longrightarrow \qquad f \;\simeq\; \kappa^2\rho^2, \tag{7.6}$$

   the near-horizon Euclidean metric is

$$\mathrm{d}s_E^2 \;\simeq\; \rho^2\,\mathrm{d}(\kappa\tau)^2 \;+\; \mathrm{d}\rho^2 \;+\; r_S^2\,\mathrm{d}\Omega^2. \tag{7.7}$$

2. *Polar coordinates.* The first two terms are the flat plane with radius ρ and angle $\kappa\tau$, provided the angle has period 2π; for any other period $\rho = 0$ is a conical singularity, circles at proper radius ρ having circumference $\neq 2\pi\rho$. Since τ has period β, the angle has period $\kappa\beta$.

The Euclidean Schwarzschild geometry (7.5) is a smooth, complete Riemannian manifold if and only if the Euclidean time period is

$$\beta \;=\; \frac{2\pi}{\kappa} \;=\; 8\pi GM, \tag{7.8}$$

i.e. if and only if the temperature equals the Hawking temperature $T_H = \kappa/2\pi = 1/8\pi GM$. A black hole cannot be in equilibrium at any other temperature: regularity of the Euclidean saddle replaces the entire mode-by-mode derivation of Chapter 4.

- *Why the two derivations agree.* In Lorentzian signature nothing in the classical geometry picks out a temperature; in Euclidean signature the temperature is the period assigned to τ and the geometry vetoes all but one choice. The near-horizon region is Rindler space, and the smooth Euclidean continuation of the Minkowski vacuum is the one with angular period 2π (Chapter 6), which is Hartle–Hawking correlations at the local Unruh temperature, redshifting to T_H at infinity.

- *The cigar* (Figure 7.1). The (τ, r) surface caps off smoothly at $r = r_S$ and opens into a cylinder of circumference β; a round sphere of radius r sits over each point. There is no interior and no singularity: the Euclidean section covers only $r \geq r_S$, the horizon has become the regular tip, the Euclidean image of the bifurcation sphere, and $r = 0$ is absent.

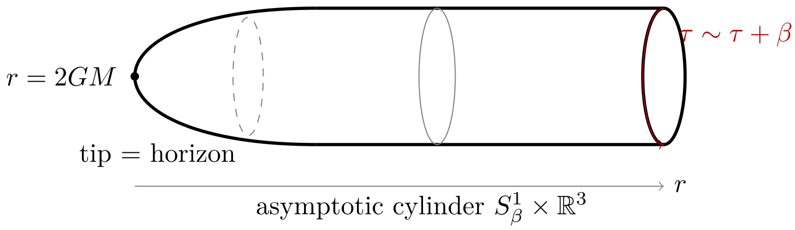


Figure 7.1: The Euclidean Schwarzschild "cigar": the (τ, r) geometry, with the round S^2 of radius r suppressed at every point. The Euclidean time circle has asymptotic circumference β and shrinks smoothly to a point at $r = 2GM$ precisely when $\beta = 8\pi GM$; for any other period the tip has a conical singularity. There is no interior region and no singularity: the tip is the Euclidean image of the bifurcation sphere.

- *Topology.* The cigar is a disk, the time circle shrinking to a point at the tip, so the four-manifold is $\mathbb{R}^2 \times S^2$, not the $S^1 \times \mathbb{R}^3$ of the thermal cylinder. The circle at fixed r is an orbit of ∂_τ of proper length $\beta\sqrt{f(r)}$ and shrinks where ∂_τ vanishes, the Killing horizon of Definition 1.1. In a geometry without a horizon, hot flat space or the exterior of a star, the circle never shrinks. Contractibility of the time circle is the geometric fingerprint of a horizon; this returns in the Hawking–Page transition of Chapter 18.

- *Two saddles.* For given β the path integral with thermal-cylinder boundary conditions has at least two saddles, hot flat space $S^1_\beta \times \mathbb{R}^3$ and the cigar with $M = \beta/8\pi G$; which dominates is settled in Section 7.5.

7.3 The gravitational path integral

Gibbons and Hawking proposed to compute the thermal partition function of gravity as a Euclidean path integral over metrics,

$$Z(\beta) \;=\; \int_{\mathcal{B}} Dg \;\; \mathrm{e}^{-I_E[g]} \;\approx\; \sum_{\text{saddles } g_*} \mathrm{e}^{-I_E[g_*]}, \tag{7.9}$$

with boundary condition $\mathcal{B}$ that the geometry asymptote to $S^1_\beta \times \mathbb{R}^3$, an asymptotically flat space whose time circle has proper circumference β at infinity. The integral is only formally defined (Section 7.7), but for $G \to 0$ at fixed β/G it is dominated by smooth classical solutions with these boundary conditions. The action is the Euclidean Einstein–Hilbert action with its boundary term,

$$I_E \;=\; -\frac{1}{16\pi G}\int_{\mathcal{M}} \mathrm{d}^4x\,\sqrt{g}\,R \;-\; \frac{1}{8\pi G}\oint_{\partial\mathcal{M}} \mathrm{d}^3x\,\sqrt{h}\;\left(K - K_0\right), \tag{7.10}$$

where h is the induced metric on a large boundary surface, K the trace of its extrinsic curvature and K_0 the same for the identical surface embedded in flat space. Both corrections carry all the physics.

- *The Gibbons–Hawking–York term* [42, 43]. We want a variational principle in which only the boundary metric is held fixed, $\delta g_{\mu\nu} = 0$ on $\partial\mathcal{M}$ with the normal derivatives free. The Einstein–Hilbert term alone does not provide one, because R contains second derivatives of the metric. The K term repairs this in three steps.

1. Varying the bulk term gives the Einstein tensor plus a total derivative, which Gauss's theorem turns into a boundary integral,

$$\begin{aligned} \delta\int_{\mathcal{M}} \mathrm{d}^4x\sqrt{g}\,R &= \int_{\mathcal{M}} \mathrm{d}^4x\sqrt{g}\,G_{\mu\nu}\,\delta g^{\mu\nu} + \oint_{\partial\mathcal{M}} \mathrm{d}^3x\sqrt{h}\,n_\mu v^\mu, \\ v^\mu &= \nabla_\nu \delta g^{\mu\nu} - \nabla^\mu(g_{\alpha\beta}\,\delta g^{\alpha\beta}), \end{aligned} \tag{7.11}$$

   and the boundary integral does not vanish for our variations, since $n_\mu v^\mu$ contains normal derivatives of $\delta g_{\mu\nu}$.

2. For the same variations [44, §4.1],

$$\delta\oint_{\partial\mathcal{M}} \mathrm{d}^3x\sqrt{h}\,2K = -\oint_{\partial\mathcal{M}} \mathrm{d}^3x\sqrt{h}\,n_\mu v^\mu, \tag{7.12}$$

   exactly the negative of the unwanted term.

3. Adding (7.11) and (7.12), the combination $\int\sqrt{g}\,R + 2\oint\sqrt{h}\,K$ of (7.10) varies into $\int\sqrt{g}\,G_{\mu\nu}\delta g^{\mu\nu}$ alone, which vanishes on solutions; K_0 depends only on the fixed boundary metric.

In a fixed-temperature ensemble we fix the boundary geometry $S^1_\beta \times S^2_R$, so the term is obligatory.

- *The subtraction.* Even on shell the integral of K diverges linearly as the boundary is pushed to infinity. The divergence is a property of the asymptotic region, common to all geometries with the same boundary data, and is removed by subtracting K_0 of flat space with the same induced boundary metric. This normalizes Z so that hot flat space has $\log Z = 0$ at leading order: the black hole's free energy is measured relative to the thermal vacuum.

In the semiclassical approximation, the thermal partition function of gravity is

$$\log Z(\beta) = -I_E[g_*], \tag{7.13}$$

where I_E is the on-shell Euclidean action (7.10) of the dominant smooth saddle g_* with thermal boundary conditions. All of equilibrium black hole thermodynamics follows from this single formula via the identities (7.4).

7.4 Computing the action

We evaluate I_E on the cigar, following [43].

1. *The bulk term vanishes.* Euclidean Schwarzschild solves the vacuum equations, $R_{\mu\nu} = 0$, so

$$-\frac{1}{16\pi G}\int \mathrm{d}^4x\sqrt{g}\,R = 0: \tag{7.14}$$

   the entire on-shell action resides in the boundary term. The action, and hence free energy and entropy, are quasi-local quantities living at boundaries and horizons, not integrals of a local entropy density; the sharpest version is Section 7.6.

2. *Extrinsic curvature of a large sphere.* The boundary at $r = R$ is $S^1_\beta \times S^2_R$ with induced metric

$$\mathrm{d}s_h^2 = f(R)\,\mathrm{d}\tau^2 + R^2\,\mathrm{d}\Omega^2, \qquad \sqrt{h} = f(R)^{1/2}\,R^2\sin\theta, \tag{7.15}$$

and outward unit normal $n = \sqrt{f}\,\partial_r$. Since $\sqrt{g} = r^2 \sin\theta$, the f's cancelling between $g_{\tau\tau}$ and g_{rr},

$$K \;=\; \nabla_\mu n^\mu \;=\; \frac{1}{r^2}\,\partial_r\Big(r^2\sqrt{f}\Big) \;=\; \frac{2\sqrt{f}}{r} \;+\; \frac{f'}{2\sqrt{f}}\,. \tag{7.16}$$

Integrating over the boundary, with $\mathrm{d}^3x\sqrt{h} = \mathrm{d}\tau\,\mathrm{d}\Omega\,R^2\sqrt{f(R)}$, the τ-integral giving β and the angles 4π,

$$\begin{aligned}\oint \mathrm{d}^3x\,\sqrt{h}\,K \;&=\; \beta\cdot 4\pi R^2\cdot\sqrt{f}\left(\frac{2\sqrt{f}}{R}+\frac{f'}{2\sqrt{f}}\right) \;=\; 4\pi\beta\left(2Rf(R)+\frac{R^2 f'(R)}{2}\right)\\ &=\; 4\pi\beta\,(2R-4GM+GM) \;=\; 4\pi\beta\,(2R-3GM),\end{aligned} \tag{7.17}$$

using $f = 1-2GM/R$ and $f' = 2GM/R^2$. As promised it diverges linearly in R.

3. *The flat-space subtraction.* The comparison surface in flat space, $\mathrm{d}s^2 = \mathrm{d}\tau'^2+\mathrm{d}r^2+r^2\mathrm{d}\Omega^2$, must have the same induced metric: the sphere has radius R, and the time circle, of proper circumference $\beta\sqrt{f(R)}$ in the black hole geometry, must have period

$$\beta' \;=\; \beta\,\sqrt{f(R)}, \tag{7.18}$$

i.e. the local Tolman temperatures at the boundary must agree. With $K_0 = 2/R$ from (7.16) at $f\equiv 1$,

$$\oint \sqrt{h}\,K_0 \;=\; \beta'\cdot 4\pi R^2\cdot\frac{2}{R} \;=\; 8\pi R\,\beta\sqrt{f(R)} \;=\; 4\pi\beta\left(2R-2GM-\frac{G^2M^2}{R}\right)+\mathcal{O}(R^{-2}), \tag{7.19}$$

by $\sqrt{f(R)} = 1-GM/R-G^2M^2/2R^2+\mathcal{O}(R^{-3})$. The divergent $8\pi\beta R$ cancels against (7.17), but the product of the divergent area with the $\mathcal{O}(1/R)$ redshift has left a finite remainder: this interplay is characteristic of gravitational actions and is why the matching (7.18) matters.

4. *Assemble.*

$$\begin{aligned}I_E \;&=\; -\frac{1}{8\pi G}\lim_{R\to\infty}\left[\oint\sqrt{h}\,K-\oint\sqrt{h}\,K_0\right]\\ &=\; -\frac{4\pi\beta}{8\pi G}\lim_{R\to\infty}\Big[(2R-3GM)-(2R-2GM)\Big] \;=\; -\frac{\beta}{2G}\,(-GM) \;=\; \frac{\beta M}{2}\,.\end{aligned} \tag{7.20}$$

With the smoothness condition (7.8), $M = \beta/8\pi G$, as a function of the one variable of the ensemble:

$$\boxed{I_E(\beta) \;=\; \frac{\beta M}{2} \;=\; \frac{\beta^2}{16\pi G}} \tag{7.21}$$

7.5 Black hole thermodynamics from the saddle

Inserting (7.21) into (7.13), $\log Z(\beta) = -\beta^2/16\pi G$, and the identities (7.4) give everything.

- *Energy.*

$$E \;=\; -\frac{\partial}{\partial\beta}\log Z \;=\; \frac{2\beta}{16\pi G} \;=\; \frac{\beta}{8\pi G} \;=\; M. \tag{7.22}$$

The thermodynamic energy equals the ADM mass of the saddle. This is a check, not a tautology: the action knew about M only through the asymptotic falloff, and the factor $\beta M/2$ rather than βM is what makes the derivative come out to M. Dropping the boundary term, or subtracting without matching the circumference, would have broken the first law.

- *Entropy.*

$$S \;=\; \beta E + \log Z \;=\; \frac{\beta^2}{8\pi G} - \frac{\beta^2}{16\pi G} \;=\; \frac{\beta^2}{16\pi G} \;=\; 4\pi G M^2. \tag{7.23}$$

The Euclidean path integral assigns the black hole ensemble the entropy

$$S \;=\; (1-\beta\partial_\beta)\log Z \;=\; 4\pi G M^2 \;=\; \frac{A}{4G}\,, \tag{7.24}$$

with $A = 4\pi r_S^2 = 16\pi G^2 M^2$: the Bekenstein–Hawking entropy, with the coefficient $1/4$ that Chapter 2 could fix only by invoking Hawking's T_H. Here temperature *and* entropy emerge together from the geometry of a single saddle point.

- *Free energy.*

$$F \;=\; -\frac{1}{\beta}\log Z \;=\; \frac{\beta}{16\pi G} \;=\; \frac{M}{2} \;=\; M - T_H\,\frac{A}{4G}\,, \tag{7.25}$$

  the last equality by $T_H \cdot A/4G = (1/8\pi GM)(4\pi GM^2) = M/2$; the first law $\mathrm{d}E = T\mathrm{d}S$ holds identically by construction.

- *Negative specific heat.*

$$C \;=\; \frac{\partial E}{\partial T} \;=\; -\beta^2\frac{\partial E}{\partial \beta} \;=\; -\,\frac{\beta^2}{8\pi G} \;=\; -\,8\pi G M^2 \;<\; 0, \tag{7.26}$$

  in agreement with Chapter 5. In Euclidean language it is a convexity failure: $\partial_\beta^2 \log Z = \langle (E-\langle E\rangle)^2\rangle \geq 0$ in any honest ensemble, whereas here $\partial_\beta^2 \log Z = -1/8\pi G$. The saddle exists but is thermodynamically unstable, the Gaussian integral over metric fluctuations around it has a negative mode, and its free energy $M/2 > 0$ exceeds that of hot flat space: the black hole is a subdominant, metastable contribution, and the canonical ensemble of gravity in flat space does not truly exist. The cure is a box, most elegantly a negative cosmological constant, for which large black holes have positive specific heat (Chapter 18).

7.6 Another computation of the entropy: the conical deficit

The derivation above differentiates a function of β computed from smooth saddles. A second route isolates the entropy geometrically, as a term localized at the horizon; it explains why the entropy is an area and is the ancestor of the derivation of holographic entanglement entropy in Chapter 21 [45].

1. *An off-shell family.* Write the entropy as

$$S \;=\; (1-\beta\partial_\beta)\log Z \;=\; (\beta\partial_\beta - 1)\, I_E(\beta). \tag{7.27}$$

   To differentiate we need I_E for β away from $\beta_H = 8\pi GM$. Keep the geometry (7.5) with M fixed and vary only the period $\tau \sim \tau + \beta$; by (7.7) the tip is then a cone with deficit angle

$$\delta \;=\; 2\pi - \kappa\beta \;=\; 2\pi\left(1 - \frac{\beta}{\beta_H}\right). \tag{7.28}$$

2. *The smooth part is all energy.* Split $I_E(\beta) = I_{\text{tip}}(\beta) + I_{\text{smooth}}(\beta)$. The metric does not depend on τ, so away from the tip every integrand in (7.10) is τ-independent and the τ-integral produces a factor β, as in (7.17): $I_{\text{smooth}} = c\beta$ with $c = M/2$ by (7.20). The operator in (7.27) annihilates it,

$$(\beta\partial_\beta - 1)\, c\beta \;=\; c\beta - c\beta \;=\; 0\,: \tag{7.29}$$

   a β-linear action is all energy and no entropy, and the entire entropy comes from the tip.

3. *Curvature of the tip.* A two-dimensional cone with deficit δ is flat away from the apex but carries a delta function of curvature there: smoothing the apex and applying Gauss–Bonnet,[1]

$$\int_{\text{cone}} \sqrt{g^{(2)}}\, R^{(2)} \;=\; 2\,\delta, \tag{7.30}$$

independent of the smoothing. Near the tip the four-geometry is this cone times the horizon sphere of area A, so $\int \sqrt{g}\, R$ acquires $2\delta A$ and the action the term

$$I_{\text{tip}} \;=\; -\,\frac{1}{16\pi G}\cdot 2\,\delta\, A \;=\; -\,\frac{A}{8\pi G}\, 2\pi\left(1-\frac{\beta}{\beta_H}\right) \;=\; -\,\frac{A}{4G}\left(1-\frac{\beta}{\beta_H}\right). \tag{7.31}$$

4. *Evaluate at the smooth point*, where $I_{\text{tip}} = 0$ but its derivative does not vanish:

$$S \;=\; (\beta\partial_\beta - 1)\, I_{\text{tip}}\Big|_{\beta=\beta_H} \;=\; \beta_H \cdot \frac{A}{4G}\cdot\frac{1}{\beta_H} \;-\; 0 \;=\; \frac{A}{4G}\,. \tag{7.32}$$

- *The entropy is localized at the horizon.* It is the response of the action to a conical opening or closing of the time circle at the fixed-point locus of time translations. Adding (7.31) to $I_{\text{smooth}} = \frac{1}{2}M\beta$ gives $I_E(\beta) = \beta M - A/4G$, the thermodynamic form $\beta E - S$ at fixed M: the bulk contributes the energy, the β-independent constant of the tip term the entropy. Gravitational entropy is an area because the deficit couples to the action through the volume of the fixed-point set of the Euclidean time symmetry, which is the horizon.

- *Two caveats.* The off-shell family is a choice; the answer is insensitive to it because at β_H the variation of the action away from the tip vanishes by the equations of motion, so only the explicit tip term survives, but a fully satisfactory treatment, in which the cone is smoothed and the geometry backreacts, had to wait for [45]. And the conical geometries are not stationary points of the action; they parametrize a direction along which Z is differentiated, which is all (7.27) needs.

7.7 Is the entropy a counting of states?

We obtained $S = A/4G$ twice from $Z(\beta)$, treating (7.9) as an ordinary partition function $Z = \mathrm{Tr}\, \mathrm{e}^{-\beta H} = \sum_n \mathrm{e}^{-\beta E_n}$, for which $\mathrm{e}^{S(E)}$ counts microstates. Does this entitle us to conclude that a Schwarzschild black hole has $\mathrm{e}^{4\pi G M^2}$ states? It is highly suggestive, but by itself it does not establish this [1].

1. *The path integral is formal.* Writing $g_{\mu\nu} = \Omega^2 \hat{g}_{\mu\nu}$, the action (7.10) contains

$$-\,\frac{6}{16\pi G}\int \mathrm{d}^4x\sqrt{\hat{g}}\,(\hat{\partial}\Omega)^2\;:$$

the kinetic term of the conformal factor has the wrong sign, so (7.9) diverges on any real contour. The standard remedy integrates the conformal factor along an imaginary direction, $\Omega = 1 + \mathrm{i}\phi$ [48], which leaves the classical solutions and their actions untouched; a non-perturbative definition is not known.

[1] Gauss–Bonnet for a surface D with the topology of a disk: $\frac{1}{2}\int_D \sqrt{g^{(2)}}\, R^{(2)} + \oint_{\partial D} k_g\, \mathrm{d}s = 2\pi$, with k_g the geodesic curvature of the boundary. Take D to be the cone cut off at distance r from the (smoothed) apex. Outside the smoothing region the metric is flat, $\mathrm{d}s^2 = \mathrm{d}r^2 + r^2\mathrm{d}\phi^2$ with $\phi \sim \phi + 2\pi - \delta$ $\big($(7.7) with $\phi = \kappa\tau\big)$, so the boundary circle has length $(2\pi-\delta)r$ and $k_g = 1/r$, giving $\oint k_g\, \mathrm{d}s = 2\pi - \delta$. Hence $\int_D \sqrt{g^{(2)}}\, R^{(2)} = 2\big[2\pi - (2\pi - \delta)\big] = 2\delta$, independent of r and of how the apex is smoothed [46, 47].

2. *The trace cannot hold as stated.* With $S(E) = 4\pi G E^2$, $\mathrm{Tr}\, \mathrm{e}^{-\beta H} \sim \int \mathrm{d}E\, \mathrm{e}^{4\pi G E^2 - \beta E}$ diverges for every β, the ensemble version of (7.26); the saddle is at best computing the continuation of the microcanonical density of states, $\varrho(M) \sim \mathrm{e}^{A/4G}$, which is the sharp conjecture the computation suggests.

3. *Nothing has been exhibited.* We computed a geometric quantity obeying all thermodynamic identities and are invited to postulate a Hilbert space behind it. The gap has been closed in special corners: Strominger and Vafa [19] counted the microstates of certain supersymmetric black holes, matching $A/4G$ including the coefficient; and AdS/CFT provides a definition of quantum gravity in a box as an ordinary quantum system for which $Z(\beta) = \mathrm{Tr}\, \mathrm{e}^{-\beta H}$ is a state count, so that the black hole entropy is the logarithm of a number of CFT microstates (Chapter 18). A complementary view, black hole entropy as entanglement across the horizon, is Chapter 20.

7.8 The Hartle–Hawking state and the thermofield double

Section 3.5.1 promised a horizon-regular equilibrium state for the eternal black hole, and Chapter 6 ended with the suggestion that it is the thermofield double (6.19) of the two exterior regions. We construct it by transcribing the Euclidean argument of Chapter 6 from the plane to the cigar.

7.8.1 The two-sided Hilbert space

- *The slice.* In the maximally extended geometry (Figure 1.1) the $t = 0$ slice is a Cauchy surface, the union of a right half $\mathcal{S}_r$ in region I and a left half $\mathcal{S}_\ell$ in region III meeting at the bifurcation sphere (Figure 7.2(d)), the counterpart of $\mathcal{S} = \mathcal{S}_\ell \cup \mathcal{S}_r$ in Minkowski space. The Hilbert space is $\mathcal{H}_\ell \otimes \mathcal{H}_r$, with the caveat of Remark 6.1.

- *The Hamiltonians.* Let H_r generate Schwarzschild time translations in region I on $\mathcal{H}_r$, the energy density on $\mathcal{S}_r$ weighted by the redshift. By the near-horizon metric (6.22), where $\eta = \kappa t$,
$$H_r \;\simeq\; \kappa\, K_r, \qquad \text{so that} \qquad \mathrm{e}^{-\beta_H H_r} \;\simeq\; \mathrm{e}^{-2\pi K_r} \quad \text{for} \quad \beta_H = \frac{2\pi}{\kappa}\,, \tag{7.33}$$
with K_r the boost generator (6.8). The Killing flow in region III runs backwards relative to Kruskal time, as the boost does in the left wedge, and H_ℓ denotes the generator of forward evolution there.

7.8.2 The Hartle–Hawking state from the half-cigar

The Euclidean geometry (7.5) is independent of τ, whose translation is generated by H_r. The argument of Section 6.3 goes through in three steps.

1. *Wedges of the cigar.* The Euclidean path integral over the region $0 \le \tau \le \tau_0$, with field configurations fixed on its two edges (Figure 7.2(a)), computes
$$\int_{0 \le \tau \le \tau_0} D\phi\; \mathrm{e}^{-S_E[\phi]} \;=\; \langle \phi' |\, \mathrm{e}^{-\tau_0 H_r}\, | \phi \rangle. \tag{7.34}$$
Near the tip, by (7.7) and (7.33), this is the wedge of Figure 6.3(a) with $\theta = \kappa \tau_0$.

2. *The full period.* For $\tau_0 = \beta_H$ the region is the cigar cut open along $\tau = 0$ (Figure 7.2(b)), and, as in (6.9), the cut cigar is the reduced density matrix of region I:
$$\rho_r \;=\; \frac{1}{Z}\, \mathrm{e}^{-\beta_H H_r}, \qquad Z = \mathrm{Tr}\, \mathrm{e}^{-\beta_H H_r}, \tag{7.35}$$

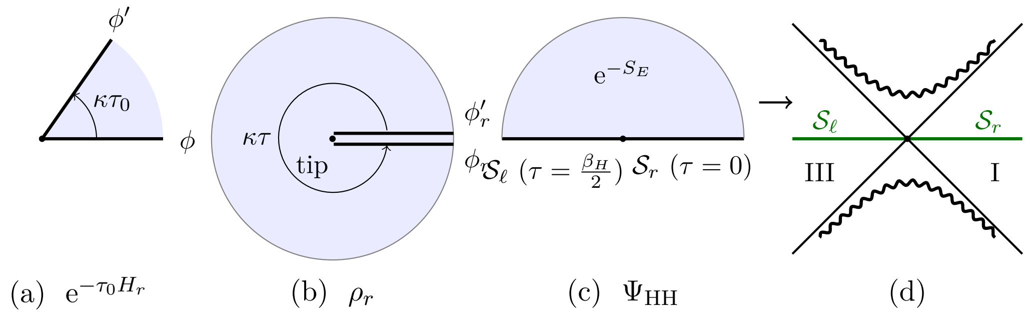


Figure 7.2: The Euclidean cigar seen in the angular $(\rho, \kappa\tau)$ plane, with the tip at the centre and the sphere S^2 suppressed; the whole cigar is topologically this disk, with $\rho \to \infty$ at the rim. (a) The region $0 \le \tau \le \tau_0$, a wedge of opening angle $\kappa\tau_0$ with data ϕ, ϕ' fixed on its two edges: the path integral (7.34), computing $e^{-\tau_0 H_r}$. (b) $\tau_0 = \beta_H$: the cigar cut open along the exterior slice $\tau = 0$, with independent boundary values ϕ_r, ϕ'_r on the two lips of the cut, which is the reduced density matrix ρ_r of (7.35). (c) $\tau_0 = \beta_H/2$: half of the cigar, with data fixed on the two boundary half-lines $\tau = 0$ and $\tau = \beta_H/2$, which is the path integral (7.36). Panels (a)–(c) are Figures 6.3(a), 6.3(b) and 6.2(a) transplanted to the black hole. (d) The two half-lines of (c) are the two halves $\mathcal{S}_r$, $\mathcal{S}_\ell$ of the $t = 0$ slice of the Kruskal geometry, lying in the exterior regions I and III; the half-cigar path integral prepares the state $\Psi_{\rm HH}$ on this slice.

with Z the path integral on the uncut cigar. Only the existence of the Euclidean path integral was used: (7.35) holds for any field content and any interactions.

3. *Half the period.* For $\tau_0 = \beta_H/2$ the region is the half-cigar (Figure 7.2(c)). Its two edges $\tau = 0$ and $\tau = \beta_H/2$ continue to the two halves $\mathcal{S}_r$ and $\mathcal{S}_\ell$ of the Kruskal slice (Figure 7.2(d)): half a turn in Euclidean angle carries region I into region III. By (7.34) and (7.35) with $\tau_0 = \beta_H/2$, the half-cigar path integral is the operator $e^{-\beta_H H_r/2} = \sqrt{Z}\,\rho_r^{1/2}$ on $\mathcal{H}_r$; by the operator–state correspondence (6.16), with $\mathcal{H}'_r$ identified with $\mathcal{H}_\ell$ by CRT as in Section 6.5.1, it is a vector in $\mathcal{H}_\ell \otimes \mathcal{H}_r$,

$$\Psi_{\rm HH}(\phi_\ell, \phi_r) \;=\; \frac{1}{\sqrt{Z}} \int_{\text{half-cigar}} D\phi \;\, e^{-S_E[\phi]} \;\;\in\;\; \mathcal{H}_\ell \otimes \mathcal{H}_r, \tag{7.36}$$

the *Hartle–Hawking state* of the eternal black hole [24]. Expanding $\rho_r^{1/2}$ in eigenstates of H_r, as in (6.18)–(6.19), gives its explicit form.

The Hartle–Hawking state of the eternal Schwarzschild black hole is the thermofield double (6.19) of its two exterior regions at inverse temperature $\beta_H = 8\pi GM$:

$$|\Psi_{\rm HH}\rangle \;=\; \frac{1}{\sqrt{Z}} \sum_n e^{-\beta_H E_n/2}\, |n\rangle_\ell \otimes |n\rangle_r\,, \tag{7.37}$$

where $|n\rangle_r$ are the eigenstates of H_r with eigenvalues E_n and $|n\rangle_\ell$ their CRT images in region III. Tracing out either exterior leaves the thermal state (7.35) on the other.

Equation (7.37) is (6.21) with $K_r \to H_r$ and $2\pi \to \beta_H$, as (7.33) dictates. That the eternal black hole should be understood in this way, as two copies of the exterior in the state (7.37), was proposed by Israel [41] shortly after Hawking's discovery and independently of the Euclidean formalism.

7.8.3 Properties of the Hartle–Hawking state

Each of the following is read off from (7.36) or (7.37).

- *Each exterior is in equilibrium at T_H.* By (7.35) every observable in region I has a thermal expectation value at inverse temperature β_H with respect to H_r. For a static observer this is the local temperature (6.23), and Tolman's law (6.15) gives the temperature at infinity (6.24), now exactly at all r: at infinity the state is a thermal bath at T_H, with as much radiation falling in as coming out. This is the black hole in equilibrium in the thermal box of Section 5.2, not an evaporating one.

- *Regularity at the horizon.* Near the tip the metric (7.7) is flat and the half-cigar is the half-plane, so there $\Psi_{\rm HH}$ reduces to the Minkowski vacuum (6.21) of the local Rindler wedges. Its correlators are regular on both horizons and an observer falling through sees only vacuum fluctuations. In mode language, $\Psi_{\rm HH}$ is positive frequency with respect to the Kruskal coordinates at the horizon rather than Schwarzschild time, which is what makes it Hadamard, in contrast with the Boulware vacuum of Section 3.5.1, and it is how Hartle and Hawking defined the state [24].

- *Two-sided correlations.* Inserting $\mathcal{O}_r$ on $\mathcal{S}_r$ and $\mathcal{O}_\ell$ on $\mathcal{S}_\ell$ in (7.36) and gluing to the conjugate half-cigar leaves the uncut cigar with two insertions at Euclidean times 0 and $\beta_H/2$, a thermal trace:

$$\langle\Psi_{\rm HH}|\,\mathcal{O}_\ell\,\mathcal{O}_r\,|\Psi_{\rm HH}\rangle \;=\; \frac{1}{Z}\,\mathrm{Tr}\left[\mathrm{e}^{-\beta_H H_r/2}\,\mathcal{O}\,\mathrm{e}^{-\beta_H H_r/2}\,\mathcal{O}\right], \tag{7.38}$$

  with $\mathcal{O}$ the operator in region I of which $\mathcal{O}_\ell$ is the CRT image. It does not vanish, although regions I and III are spacelike separated. The same gluing shows that Lorentzian correlators are continuations of functions periodic in τ with period β_H, the KMS condition (4.23) at β_H, which is how Hartle and Hawking identified the temperature [24].

- *A pure state with mixed restrictions.* The global state (7.37) is pure, while its restriction (7.35) to either exterior is mixed. That is what it means for the two exteriors to be entangled: (7.37) is a superposition of products $|n\rangle_\ell|n\rangle_r$ rather than a single product state, so the fields in region I are correlated with those in region III, as (7.38) shows, without any signal passing between them. Chapter 19 defines the measure of this entanglement, the von Neumann entropy of (7.35); it is the entropy of the fields outside the horizon at T_H, ultraviolet divergent for the reason of Remark 6.1, and how it combines with $A/4G$ into a cutoff-independent total is Chapter 20.

- *Outlook.* The thermofield double becomes fully operational once the two copies in (7.37) are honest quantum systems. AdS/CFT provides this: Maldacena's eternal AdS black hole [49] is dual to two copies of a conformal field theory in the state (7.37) (Section 18.7), which turns every statement of this section, including the state counting missing in Section 7.7, into a statement about an ordinary quantum system. Building that machinery is the task of Part II.

Part II

The AdS/CFT Correspondence

Chapter 8

Conformal Field Theories I: The Conformal Group

With this chapter Part II begins; the destination is the AdS/CFT correspondence, and we start with the field-theory side. We define conformal transformations, solve the conformal Killing equation in $d \geq 3$, and show that the resulting algebra is $\mathfrak{so}(2, d)$, the group that reappears in Chapter 10 as the isometries of AdS_{d+1}. We then describe how conformal symmetry acts on quantum fields: scaling dimensions, primaries and descendants, and the organization of the Hilbert space into lowest-weight representations graded by the dilatation operator. Throughout, d is the spacetime dimension of the field theory, the signature is mostly-plus, and $d \geq 3$ unless stated otherwise.

8.1 Why conformal field theories?

- *Within field theory.* CFTs are the fixed points of the renormalization group, the landmarks from which the space of quantum field theories is charted, and the theories about which the most can be said without solving any dynamics: symmetry fixes two- and three-point functions completely (Chapter 9).

- *For quantum gravity.* The isometry group of AdS_{d+1} is $\mathrm{SO}(2, d)$, which is the conformal group of d-dimensional Minkowski space. Whatever describes quantum gravity in AdS_{d+1} must carry an action of $\mathrm{SO}(2, d)$, and a conformal field theory in d dimensions is the natural candidate. This chapter and the next two supply the kinematics; the dynamical evidence occupies Chapters 12–17.

8.2 Conformal transformations and the conformal Killing equation

- *Scale invariance.* A theory without dimensionful parameters is classically invariant under the dilatation
$$x^\mu \;\to\; x'^\mu = \lambda\, x^\mu, \qquad \phi(x) \;\to\; \phi'(x') = \lambda^{-\Delta}\, \phi(x), \tag{8.1}$$
which rescales the field with a weight Δ, its *scaling dimension.* For a free scalar, $S = -\frac{1}{2}\int \mathrm{d}^d x\, (\partial\phi)^2$ picks up $\lambda^{d-2-2\Delta}$ and is invariant for
$$\Delta_\phi \;=\; \frac{d-2}{2}, \tag{8.2}$$

the canonical dimension; a mass term scales as $\lambda^{d-2\Delta}$ and breaks the invariance, since masses set a scale. Quantum mechanically the running of couplings breaks it too; Section 8.5 describes when it survives.

- *Definition.* A *conformal transformation* of Minkowski space is a change of coordinates $x \to x'(x)$ that rescales the line element by a positive position-dependent factor,

$$\mathrm{d}x'^2 \;=\; \Omega^2(x)\,\mathrm{d}x^2. \tag{8.3}$$

It preserves angles and the causal structure. Poincaré transformations are the case $\Omega = 1$, dilatations the case of constant Ω.

- *The conformal Killing equation.* Infinitesimally, $x'^\mu = x^\mu + v^\mu(x)$ and $\Omega = 1 + \frac{1}{2}\omega(x)$, so (8.3) becomes $\partial_\mu v_\nu + \partial_\nu v_\mu = \omega\,\eta_{\mu\nu}$, and the trace gives $\omega = \frac{2}{d}\,\partial\cdot v$:

$$\boxed{\partial_\mu v_\nu + \partial_\nu v_\mu - \frac{2}{d}\,(\partial\cdot v)\,\eta_{\mu\nu} \;=\; 0.} \tag{8.4}$$

Killing vectors satisfy the stronger $\partial_\mu v_\nu + \partial_\nu v_\mu = 0$; the conformal equation allows a pure-trace right-hand side.

8.3 The conformal group in $d \geq 3$

We solve (8.4), written as $\partial_\mu v_\nu + \partial_\nu v_\mu = \omega\,\eta_{\mu\nu}$ with $\omega = \frac{2}{d}\partial\cdot v$, by a chain of differentiations.

1. Apply ∂^ν:

$$\Box v_\mu + \partial_\mu(\partial\cdot v) = \partial_\mu\omega \quad\Longrightarrow\quad \Box v_\mu = \Big(\frac{2}{d} - 1\Big)\,\partial_\mu(\partial\cdot v). \tag{8.5}$$

2. Apply $\Box$ to the equation and use (8.5):

$$(2-d)\,\partial_\mu\partial_\nu(\partial\cdot v) = \eta_{\mu\nu}\,\Box(\partial\cdot v). \tag{8.6}$$

Its trace gives $(d-1)\Box(\partial\cdot v) = 0$, hence for $d \geq 2$ $\Box(\partial\cdot v) = 0$ and $(2-d)\,\partial_\mu\partial_\nu(\partial\cdot v) = 0$. For $d \geq 3$ this says $\partial\cdot v$ is at most linear,

$$\partial\cdot v \;=\; d\,(\lambda + 2\,b\cdot x)\,, \qquad \omega \;=\; 2\lambda + 4\,b\cdot x, \tag{8.7}$$

with constants λ, b_μ. For $d = 2$ the condition is empty (Remark 8.1).

3. Take ∂_ρ of the equation and add the cyclic permutations with signs $(+,+,-)$; four of six terms cancel and

$$2\,\partial_\mu\partial_\nu v_\rho \;=\; \eta_{\nu\rho}\,\partial_\mu\omega + \eta_{\rho\mu}\,\partial_\nu\omega - \eta_{\mu\nu}\,\partial_\rho\omega. \tag{8.8}$$

Since $\partial_\mu\omega = 4b_\mu$ is constant, v is a polynomial of degree two; substituting a general quadratic ansatz into (8.4) and matching gives the general solution.

The general conformal Killing vector of $d \geq 3$ Minkowski space is

$$v^\mu(x) \;=\; \underbrace{a^\mu}_{\text{translations } P_\mu} + \underbrace{\omega^\mu{}_\nu\, x^\nu}_{\text{Lorentz } J_{\mu\nu}} + \underbrace{\lambda\, x^\mu}_{\text{dilatation } D} + \underbrace{b^\mu x^2 - 2x^\mu\,(b\cdot x)}_{\text{special conformal } K_\mu}, \qquad \omega_{\mu\nu} = -\omega_{\nu\mu}.$$

- *Counting.* $d + \frac{d(d-1)}{2} + 1 + d = \frac{(d+1)(d+2)}{2}$ parameters, the dimension of $\mathrm{SO}(2,d)$; Section 8.4 makes this exact.

- *Finite transformations.*

$$\begin{aligned} \text{translations:} \quad & x'^\mu = x^\mu + a^\mu, && \Omega = 1, \\ \text{Lorentz:} \quad & x'^\mu = \Lambda^\mu{}_\nu\, x^\nu, && \Omega = 1, \\ \text{dilatations:} \quad & x'^\mu = \lambda\, x^\mu, && \Omega = \lambda, \\ \text{special conformal:} \quad & x'^\mu = \frac{x^\mu + b^\mu x^2}{1 + 2\, b\cdot x + b^2 x^2}, && \Omega = \frac{1}{1 + 2\, b\cdot x + b^2 x^2}. \end{aligned} \tag{8.9}$$

- *The inversion* $I : x^\mu \mapsto x^\mu/x^2$ is conformal, with Jacobian

$$\frac{\partial x'^\mu}{\partial x^\nu} = \frac{1}{x^2}\left(\delta^\mu_\nu - \frac{2\, x^\mu x_\nu}{x^2}\right) \equiv \frac{1}{x^2}\, I^\mu{}_\nu(x), \qquad \mathrm{d}x'^2 = \frac{\mathrm{d}x^2}{(x^2)^2}, \qquad \Omega_I(x) = \frac{1}{x^2}, \tag{8.10}$$

where $I^\mu{}_\nu$ is a position-dependent reflection, $I^\mu{}_\rho I^\rho{}_\nu = \delta^\mu_\nu$ with determinant -1. The inversion is not connected to the identity; adjoining it enlarges $\mathrm{SO}(2, d)$ to $\mathrm{O}(2, d)$. It is the essential computational tool of Chapter 9 because of the next two facts.

Lemma 8.1 (Inversion and distances)**.** *Under* $x' = x/x^2$, $y' = y/y^2$,

$$(x' - y')^2 = \frac{(x - y)^2}{x^2\, y^2}. \tag{8.11}$$

Proof. $\left(\frac{x}{x^2} - \frac{y}{y^2}\right)^2 = \frac{1}{x^2} - \frac{2x\cdot y}{x^2 y^2} + \frac{1}{y^2} = \frac{(x-y)^2}{x^2 y^2}$. □

Proposition 8.1 ($K = I\, P\, I$)**.** *The finite special conformal transformation with parameter* b *is an inversion, a translation by* b*, and a second inversion:* $SCT_b = I \circ T_b \circ I$.

Proof. The first two maps send x^μ to $x^\mu/x^2 + b^\mu$, whose norm is $(1 + 2b\cdot x + b^2 x^2)/x^2$, so the final inversion gives $x'^\mu = (x^\mu + b^\mu x^2)/(1 + 2b\cdot x + b^2 x^2)$, which is (8.9). □

Special conformal transformations are translations of the point at infinity, which is why an SCT can map a finite point to infinity: the cure is the conformal compactification of Minkowski space, met again in Chapter 10 as the boundary of AdS.

Remark 8.1 (The special role of $d = 2$)**.** In $d = 2$ the condition on $\partial\cdot v$ is vacuous. In Euclidean signature with $z = x^1 + \mathrm{i}x^2$, (8.4) becomes the Cauchy–Riemann equations and every holomorphic map is conformal: the algebra is infinite dimensional, two copies of the Virasoro algebra in the quantum theory [6], while the globally defined maps still form $\mathrm{SO}(2, 2)$. This course works in $d \geq 3$.

8.4 The conformal algebra and $\mathfrak{so}(2, d)$

- *Generators* as differential operators on functions of x:

$$\begin{aligned} P_\mu &= -\mathrm{i}\, \partial_\mu, \\ J_{\mu\nu} &= \mathrm{i}\left(x_\nu \partial_\mu - x_\mu \partial_\nu\right), \\ D &= -\mathrm{i}\, x^\mu \partial_\mu, \\ K_\mu &= -\mathrm{i}\left(x^2 \partial_\mu - 2x_\mu\, x^\nu \partial_\nu\right), \end{aligned} \tag{8.12}$$

with conventions such that a finite transformation acts as $\mathrm{e}^{-\mathrm{i}\epsilon\cdot G}$ and all four are hermitian in a unitary theory.

- *Dilatation as a grading.* $x\cdot\partial$ counts the scaling degree, $+1$ per x and -1 per ∂; P_μ has degree -1 and K_μ degree $+1$, so

$$[D, P_\mu] \;=\; +\mathrm{i}\,P_\mu, \qquad [D, K_\mu] \;=\; -\mathrm{i}\,K_\mu. \tag{8.13}$$

- *The bracket* $[K, P]$. With $K_\mu = -\mathrm{i}A_\mu$, $A_\mu = x^2\partial_\mu - 2x_\mu\, x\cdot\partial$,

$$\begin{aligned} [\, x^2\partial_\mu,\, \partial_\nu\,] = -2x_\nu\partial_\mu, \qquad [\, -2x_\mu\, x^\rho\partial_\rho,\, \partial_\nu\,] = 2\,\eta_{\mu\nu}\, x\cdot\partial + 2\, x_\mu\partial_\nu, \\ [K_\mu, P_\nu] = (-\mathrm{i})^2\,[A_\mu, \partial_\nu] = 2\,(x_\nu\partial_\mu - x_\mu\partial_\nu) - 2\eta_{\mu\nu}\, x\cdot\partial = -2\mathrm{i}\, J_{\mu\nu} - 2\mathrm{i}\,\eta_{\mu\nu} D. \end{aligned} \tag{8.14}$$

$$\boxed{\begin{aligned} [J_{\mu\nu}, J_{\rho\sigma}] &= \mathrm{i}\,(\eta_{\mu\rho}J_{\nu\sigma} - \eta_{\nu\rho}J_{\mu\sigma} - \eta_{\mu\sigma}J_{\nu\rho} + \eta_{\nu\sigma}J_{\mu\rho})\,, \\ [J_{\mu\nu}, P_\rho] &= \mathrm{i}\,(\eta_{\mu\rho}P_\nu - \eta_{\nu\rho}P_\mu)\,, \qquad [J_{\mu\nu}, K_\rho] = \mathrm{i}\,(\eta_{\mu\rho}K_\nu - \eta_{\nu\rho}K_\mu)\,, \\ [D, P_\mu] &= \mathrm{i}P_\mu, \qquad [D, K_\mu] = -\mathrm{i}K_\mu, \qquad [D, J_{\mu\nu}] = 0, \\ [K_\mu, P_\nu] &= -2\mathrm{i}J_{\mu\nu} - 2\mathrm{i}\,\eta_{\mu\nu}D, \qquad [P_\mu, P_\nu] = [K_\mu, K_\nu] = 0. \qquad (8.15) \end{aligned}}$$

D is a Lorentz scalar, P and K are Lorentz vectors and ladder operators for D, and K with P close onto a Lorentz transformation plus a dilatation: the backbone of the representation theory of Section 8.7.

8.4.1 The isomorphism with $\mathfrak{so}(2, d)$

With indices $M, N = 0, \ldots, d+1$, the metric with two time directions $\eta_{MN} = \mathrm{diag}(\eta_{\mu\nu}, +1, -1)$, and

$$J_{\mu\nu} = J_{\mu\nu}, \qquad J_{\mu\, d} = \frac{K_\mu - P_\mu}{2}, \qquad J_{\mu\, d+1} = -\frac{K_\mu + P_\mu}{2}, \qquad J_{d\, d+1} = D, \tag{8.16}$$

Theorem 8.1. *the conformal algebra* (8.15) *is*

$$[J_{MN}, J_{RS}] = \mathrm{i}\,(\eta_{MR}J_{NS} - \eta_{NR}J_{MS} - \eta_{MS}J_{NR} + \eta_{NS}J_{MR})\,, \tag{8.17}$$

the Lie algebra $\mathfrak{so}(2, d)$ *of pseudo-rotations of* $\mathbb{R}^{2,d}$*: the conformal group of d-dimensional Minkowski space is a cover of* $\mathrm{SO}(2, d)$*.*

Proof (representative checks). The claim is a finite list of commutators; we verify the three structurally distinct ones and leave the remaining (entirely analogous) cases to the reader's private amusement.

(i) $[J_{\mu d}, J_{\nu d}]$. The right-hand side of (8.17) gives $\mathrm{i}\,\eta_{dd}J_{\mu\nu} = \mathrm{i}J_{\mu\nu}$. Directly:

$$\begin{aligned} \Big[\frac{K_\mu - P_\mu}{2}, \frac{K_\nu - P_\nu}{2}\Big] &= -\frac{1}{4}\,([K_\mu, P_\nu] + [P_\mu, K_\nu]) = -\frac{1}{4}\,([K_\mu, P_\nu] - [K_\nu, P_\mu]) \\ &= -\frac{1}{4}\,(-2\mathrm{i}J_{\mu\nu} - 2\mathrm{i}\eta_{\mu\nu}D + 2\mathrm{i}J_{\nu\mu} + 2\mathrm{i}\eta_{\nu\mu}D) = \mathrm{i}\, J_{\mu\nu}.\checkmark \end{aligned} \tag{8.18}$$

(ii) $[J_{\mu\, d+1}, J_{\nu\, d+1}]$. Now the right-hand side is $\mathrm{i}\,\eta_{d+1\, d+1}J_{\mu\nu} = -\mathrm{i}J_{\mu\nu}$, and indeed the same computation with both relative signs flipped gives $+\frac{1}{4}([K_\mu, P_\nu] - [K_\nu, P_\mu]) = -\mathrm{i}J_{\mu\nu}$. The opposite signs in (i) and (ii) are exactly what distinguishes a spacelike from a timelike extra direction: the plane $(\mu, d+1)$ containing two timelike directions would generate compact rotations if both were timelike, and this sign bookkeeping is what fixes the signature to be $(2, d)$.

(iii) $[J_{d\,d+1}, J_{\mu\,d+1}]$. The algebra (8.17) predicts $\mathrm{i}\,\eta_{d+1\,d+1}\,J_{d\mu} = -\mathrm{i}J_{d\mu} = \mathrm{i}J_{\mu d}$. Directly:

$$\Big[D, -\frac{K_\mu + P_\mu}{2}\Big] = -\frac{1}{2}\,(-\mathrm{i}K_\mu + \mathrm{i}P_\mu) = \mathrm{i}\,\frac{K_\mu - P_\mu}{2} = \mathrm{i}J_{\mu d}.\checkmark \tag{8.19}$$

(iv) Mixed check: $[J_{\mu d}, J_{\nu\,d+1}]$ should equal $\mathrm{i}\,\eta_{\mu\nu}J_{d\,d+1} = \mathrm{i}\,\eta_{\mu\nu}D$:

$$\Big[\frac{K_\mu - P_\mu}{2}, -\frac{K_\nu + P_\nu}{2}\Big] = -\frac{1}{4}\,([K_\mu, P_\nu] + [K_\nu, P_\mu]) = -\frac{1}{4}\,(-4\mathrm{i}\,\eta_{\mu\nu}D) = \mathrm{i}\,\eta_{\mu\nu}D.\checkmark \tag{8.20}$$

$\square$

- *The same group as AdS.* $\mathrm{SO}(2,d)$ reappears in Chapter 10 as the isometry group of AdS_{d+1}, acting linearly on the embedding coordinates X^M of the hyperboloid in $\mathbb{R}^{2,d}$; the index M here is that embedding index. Boundary conformal group and bulk isometry group are one and the same, the kinematical backbone of holography.

- *The conformal Hamiltonian.*
$$H_{\text{conf}} \;=\; \frac{P^0 + K^0}{2} \tag{8.21}$$
generates the compact $\mathrm{SO}(2)$ of rotations in the plane of the two timelike directions; the maximal compact subgroup is $\mathrm{SO}(2) \times \mathrm{SO}(d)$. Chapter 9 shows it is the Hamiltonian on the cylinder $\mathbb{R} \times S^{d-1}$, and Chapter 10 that it generates global time in AdS. Quantum theories represent the universal cover, which is why dimensions are real numbers rather than integers.

8.5 Conformal quantum field theories

- *Running breaks scale invariance.* Renormalization introduces a scale μ and the couplings run, $\mu\,\mathrm{d}g/\mathrm{d}\mu = \beta(g)$. In Yang–Mills theory a scale Λ_{QCD} appears by dimensional transmutation and the trace of the stress tensor is proportional to the beta function, $T^\mu{}_\mu \sim \beta(g)g^{-3}F^a_{\mu\nu}F^{a\,\mu\nu}$: the renormalization group equation is the anomalous Ward identity for dilatations. Dimensions are corrected by anomalous dimensions, $\Delta = \Delta_0 + \gamma(g)$ with $\gamma = \frac{1}{2}\mu\,\mathrm{d}\ln Z_\phi/\mathrm{d}\mu$.

- *Fixed points* (Figure 8.1). At a zero $\beta(g_*) = 0$ the running stops, the stress tensor is traceless as an operator equation, and the theory is a CFT with dimensions $\Delta_0 + \gamma(g_*)$, generically irrational. The infrared endpoint of a flow from a massive theory can be an interacting CFT: the Wilson–Fisher point of ϕ^4 theory below four dimensions controls the critical three-dimensional Ising model, with measurable exponents. The ultraviolet and infrared limits of essentially any QFT are conformal, possibly trivially.

- *Finite theories.* If $\beta(g) = 0$ for all g there is a line of fixed points. The example that dominates the second half of the course is $\mathcal{N} = 4$ super Yang–Mills, an $\mathrm{SU}(N)$ gauge theory with, per gauge generator, one gauge field, four Weyl fermions and six real scalars, all in the adjoint (constructed in Chapter 16; no supersymmetry is needed here). The one-loop beta function
$$\beta(g) = -\frac{g^3}{16\pi^2}\left[\frac{11}{3}C(\text{adj}) - \frac{2}{3}\sum_{\text{Weyl f.}} C(r_f) - \frac{1}{6}\sum_{\text{real sc.}} C(r_s)\right] \tag{8.22}$$
vanishes because all Casimirs are equal and $\frac{11}{3} - \frac{2}{3}\cdot 4 - \frac{1}{6}\cdot 6 = 0$; the cancellation persists to all orders and the anomalous dimensions of the elementary fields vanish. With $\tau = \frac{4\pi\mathrm{i}}{g^2} + \frac{\theta}{2\pi}$, the theory is a CFT for every complex τ.

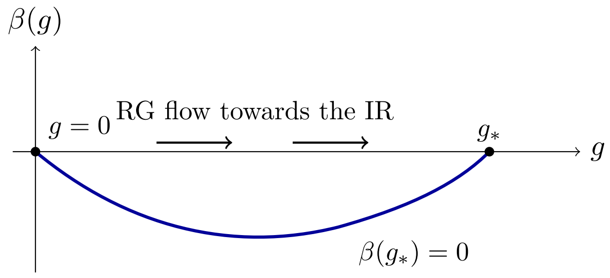


Figure 8.1: A beta function with an infrared fixed point: $\beta < 0$ means the coupling grows towards the infrared until it reaches the zero g_*, where the flow stops and the theory becomes conformal, with operator dimensions $\Delta = \Delta_0 + \gamma(g_*)$.

- *No S-matrix.* P^2 is not a Casimir: by (8.13),

$$[D, P^\mu P_\mu] = 2\mathrm{i}\, P^2, \tag{8.23}$$

so a dilatation rescales masses. Masses are zero or continuous, there are no asymptotic particle states, and the observables are correlation functions of local operators, labelled by scaling dimension and Lorentz quantum numbers rather than mass and momentum.

8.6 Action on fields: primaries and descendants

1. *Scaling dimension.* A local operator has definite dimension Δ if

$$[D, \mathcal{O}(0)] \;=\; \mathrm{i}\,\Delta\,\mathcal{O}(0). \tag{8.24}$$

2. *Translating the dilatation.* With $\mathcal{O}(x) = \mathrm{e}^{\mathrm{i}P\cdot x}\mathcal{O}(0)\mathrm{e}^{-\mathrm{i}P\cdot x}$, so that $[P_\mu, \mathcal{O}(x)] = -\mathrm{i}\partial_\mu\mathcal{O}(x)$, the Baker–Campbell–Hausdorff formula

$$\mathrm{e}^{-\mathrm{i}P\cdot x}G\,\mathrm{e}^{\mathrm{i}P\cdot x} = G + [\,-\mathrm{i}x\cdot P,\; G] + \frac{1}{2}[\,-\mathrm{i}x\cdot P,\, [\,-\mathrm{i}x\cdot P, G]] + \cdots \tag{8.25}$$

terminates after one step for $G = D$, since $[-\mathrm{i}x^\nu P_\nu, D] = -x\cdot P$ and $[P, P] = 0$: $\mathrm{e}^{-\mathrm{i}P\cdot x}D\,\mathrm{e}^{\mathrm{i}P\cdot x} = D - x\cdot P$. Hence

$$\begin{aligned} [D, \mathcal{O}(x)] &= \mathrm{e}^{\mathrm{i}Px}\,[D - x\cdot P,\; \mathcal{O}(0)]\,\mathrm{e}^{-\mathrm{i}Px} = \mathrm{i}\,(\Delta + x^\mu\partial_\mu)\,\mathcal{O}(x), \\ \mathrm{e}^{\mathrm{i}\alpha D}\,\mathcal{O}(x)\,\mathrm{e}^{-\mathrm{i}\alpha D} &= \lambda^{-\Delta}\,\mathcal{O}(x/\lambda), \end{aligned} \tag{8.26}$$

with $\lambda = \mathrm{e}^\alpha$, the quantum version of (8.1).

3. *P raises, K lowers.* By the Jacobi identity and (8.13),

$$[D, [P_\mu, \mathcal{O}(0)]] = [[D, P_\mu], \mathcal{O}(0)] + [P_\mu, [D, \mathcal{O}(0)]] = \mathrm{i}(\Delta + 1)\,[P_\mu, \mathcal{O}(0)], \tag{8.27}$$

and $[K_\mu, \mathcal{O}(0)]$ has dimension $\Delta - 1$. In a unitary theory dimensions are bounded below (Chapter 9), so every ladder terminates.

Definition 8.1. A local operator $\mathcal{O}(0)$ of definite dimension and Lorentz representation is *primary* if

$$[K_\mu, \mathcal{O}(0)] \;=\; 0. \tag{8.28}$$

Its *descendants* are the derivatives $\partial_{\mu_1}\cdots\partial_{\mu_n}\mathcal{O}$, of dimension $\Delta + n$; a primary with its descendants is a *conformal family.*

- *The spectrum.* The operator content of a CFT is its list of primaries, the pairs $(\Delta_i, \mathrm{spin}_i)$. In gauge theories the physical operators are the gauge-invariant composites.

- *Finite transformation of a primary.* Under $x \to x'$ with $\mathrm{d}x'^2 = \Omega^2 \mathrm{d}x^2$,

$$\boxed{\mathcal{O}'(x') \;=\; \Omega(x)^{-\Delta}\, \mathcal{D}[R(x)]\, \mathcal{O}(x), \qquad R^{\mu}{}_{\nu}(x) = \Omega(x)^{-1}\,\frac{\partial x'^{\mu}}{\partial x^{\nu}} \in \mathrm{SO}(1, d-1), \quad (8.29)}$$

  with $\mathcal{D}[R]$ the matrix of the local Lorentz rotation in the representation of $\mathcal{O}$, trivial for a scalar. Descendants transform inhomogeneously, which is why primaries are the building blocks.

- *K at a general point.* Applying (8.25) to $G = K_\mu$ the series terminates at second order,

$$\mathrm{e}^{-\mathrm{i}Px}\, K_\mu\, \mathrm{e}^{\mathrm{i}Px} = K_\mu + 2x^\nu J_{\mu\nu} + 2x_\mu D + x^2 P_\mu - 2x_\mu\, x\cdot P, \tag{8.30}$$

  the first-order term coming from (8.14) and the second-order one from $[J, P]$ and $[D, P]$. On a scalar primary only D and P survive:

$$[K_\mu, \mathcal{O}(x)] = \mathrm{i}\left(2\Delta\, x_\mu + 2x_\mu\, x^\nu \partial_\nu - x^2 \partial_\mu\right) \mathcal{O}(x). \tag{8.31}$$

  A primary is annihilated by K_μ only at the origin; elsewhere the right-hand side is fixed by Δ.

8.7 Lowest-weight representations

The Hilbert space is organized by diagonalizing D together with the Lorentz generators and using P_μ and K_μ as raising and lowering operators, as $a^\dagger$ and a grade the harmonic oscillator.

- *States from operators.* To every local operator associate $|\mathcal{O}\rangle = \mathcal{O}(0)|0\rangle$, with $|0\rangle$ the conformally invariant vacuum (that this is a bijection is the state–operator correspondence of Chapter 9). For a primary,

$$D|\mathcal{O}\rangle = \mathrm{i}\Delta|\mathcal{O}\rangle, \qquad K_\mu|\mathcal{O}\rangle = 0, \qquad P_{\mu_1}\cdots P_{\mu_n}|\mathcal{O}\rangle \neq 0 \;\text{ in general.} \tag{8.32}$$

  The factor i is an artefact of Lorentzian conventions; in the Euclidean radial quantization of Chapter 9 D has real eigenvalue Δ and $P_\mu^\dagger = K_\mu$.

- *The ladder.* D grades the multiplet, with level-n states $P_{\mu_1}\cdots P_{\mu_n}|\mathcal{O}\rangle$ of dimension $\Delta + n$; P_μ raises and generates the descendants; K_μ lowers, and the primary is the state of lowest dimension. Since dimensions are bounded below in a unitary theory, repeated lowering from any state reaches a primary: every state lies in a lowest-weight multiplet, and the Hilbert space is a direct sum of conformal multiplets, one per primary.

- *Shortening.* Unitarity implies sharp bounds $\Delta \geq f(\mathrm{spin})$, derived in Chapter 9; at saturation some descendants vanish, as $\Box\phi = 0$ for the free scalar, shortening the multiplet and protecting the dimension against quantum corrections.

All of this structure was extracted without writing a single interaction; the next chapter pushes it further and determines two- and three-point functions exactly.

Chapter 9

Conformal Field Theories II: Constraints from Conformal Invariance

Conformal symmetry determines two- and three-point functions of primaries exactly and constrains everything else through the operator product expansion; this chapter cashes that in. We derive the conformal Ward identities, the fixed forms of two- and three-point functions, and the correlators of conserved currents and the stress tensor. Radial quantization and the state–operator correspondence turn the OPE into a convergent expansion, and unitarity yields the bounds $\Delta \geq \frac{d-2}{2}$ for scalars and $\Delta \geq \ell + d - 2$ for spin $\ell \geq 1$, saturated by free fields and conserved currents. We work in Euclidean signature, where correlators are single-valued; Lorentzian correlators follow by continuation.

9.1 Conformal Ward identities

- *The master constraint.* For scalar primaries, $\mathcal{O}'(x') = \Omega(x)^{-\Delta}\mathcal{O}(x)$ with $\mathcal{O}' = U\mathcal{O}U^\dagger$ and $U|0\rangle = |0\rangle$; inserting $U^\dagger U$ between all operators,

$$\boxed{\langle \mathcal{O}_1(x_1')\,\mathcal{O}_2(x_2')\cdots\mathcal{O}_n(x_n')\rangle \;=\; \Omega(x_1)^{-\Delta_1}\,\Omega(x_2)^{-\Delta_2}\cdots\Omega(x_n)^{-\Delta_n}\,\langle \mathcal{O}_1(x_1)\cdots\mathcal{O}_n(x_n)\rangle, \quad (9.1)}$$

  one condition for every conformal transformation, from which Sections 9.3 and 9.4 follow.

- *Infinitesimal form.* For dilatations and special conformal transformations,

$$\begin{aligned} \sum_{i=1}^{n}\left(\Delta_i + x_i^\mu \frac{\partial}{\partial x_i^\mu}\right)\langle \mathcal{O}_1(x_1)\cdots\mathcal{O}_n(x_n)\rangle \;&=\; 0, \\ \sum_{i=1}^{n}\left(x_i^2\,\frac{\partial}{\partial x_i^\mu} - 2x_{i\mu}\,x_i^\nu\frac{\partial}{\partial x_i^\nu} - 2\Delta_i\,x_{i\mu}\right)\langle \mathcal{O}_1(x_1)\cdots\mathcal{O}_n(x_n)\rangle \;&=\; 0, \end{aligned} \quad (9.2)$$

  with the translation and rotation identities. For one operator, translation invariance makes $\langle\mathcal{O}\rangle$ constant and the first identity gives $\Delta\langle\mathcal{O}\rangle = 0$: one-point functions of all primaries except the identity vanish. Nothing condenses in a CFT.

- *Local Ward identities.* The symmetry is generated locally by the currents $j_\mu[v] = T_{\mu\nu}v^\nu$ of Section 9.2, and the path-integral argument (transform only inside a region containing some insertions) gives

$$\partial^\mu \langle T_{\mu\nu}(x)\, \mathcal{O}_1(x_1)\cdots\mathcal{O}_n(x_n)\rangle = -\sum_{i=1}^{n}\delta^d(x-x_i)\,\frac{\partial}{\partial x_i^\nu}\langle\mathcal{O}_1(x_1)\cdots\mathcal{O}_n(x_n)\rangle, \tag{9.3}$$

up to convention-dependent contact terms, and for the trace

$$\langle T^\mu{}_\mu(x)\,\mathcal{O}_1(x_1)\cdots\mathcal{O}_n(x_n)\rangle = -\sum_{i=1}^{n}\delta^d(x-x_i)\,\Delta_i\,\langle\mathcal{O}_1(x_1)\cdots\mathcal{O}_n(x_n)\rangle : \tag{9.4}$$

$T^\mu{}_\mu$ vanishes away from insertions and its contact terms measure the dimensions. The stress tensor is the generating operator of the whole symmetry, and (9.3) fixes its normalization absolutely, which is why C_T in Section 9.5 is a convention-independent number.

9.2 The stress tensor: symmetric, traceless, improved

- *Canonical tensor and improvements.* The Noether current of translations,

$$T^{\text{can}}_{\mu\nu} = \frac{\partial\mathcal{L}}{\partial(\partial^\mu\phi_a)}\,\partial_\nu\phi_a - \eta_{\mu\nu}\,\mathcal{L}, \tag{9.5}$$

  is conserved on shell but for spinning fields neither symmetric nor traceless. Both are repaired by improvements $T_{\mu\nu} = T^{\text{can}}_{\mu\nu} + \partial^\rho B_{\rho\mu\nu}$ with $B_{\rho\mu\nu} = -B_{\mu\rho\nu}$, automatically conserved and leaving the charges $P_\nu = \int T_{0\nu}$ unchanged. The Belinfante procedure uses the spin part of the Lorentz currents to make $T_{\mu\nu}$ symmetric in any Lorentz-invariant theory.

- *Tracelessness is conformal invariance.* To every conformal Killing vector v associate $j_\mu[v] = T_{\mu\nu}v^\nu$. By conservation, symmetry of T and (8.4),

$$\partial^\mu j_\mu[v] = T_{\mu\nu}\,\partial^\mu v^\nu = \frac{1}{2}\,T^{\mu\nu}\,(\partial_\mu v_\nu + \partial_\nu v_\mu) = \frac{1}{d}\,(\partial\cdot v)\,T^\mu{}_\mu. \tag{9.6}$$

  For Killing vectors $\partial\cdot v = 0$ and the current is conserved automatically; for dilatations and special conformal transformations conservation requires exactly $T^\mu{}_\mu = 0$.

In a Poincaré-invariant theory with a symmetric, conserved, *traceless* stress tensor, the currents $j_\mu[v] = T_{\mu\nu}v^\nu$ are conserved for every conformal Killing vector v: tracelessness is the local expression of scale invariance, and it delivers full conformal invariance.

- *Scale versus conformal.* Scale invariance alone guarantees only $T^\mu{}_\mu = \partial^\mu V_\mu$ for a local *virial current* V_μ, and the trace can be improved away when V_μ is itself a divergence, $V_\mu = \partial^\nu\sigma_{\nu\mu}$. In $d = 2$ unitary scale-invariant theories with discrete spectrum are conformal (Zamolodchikov–Polchinski), and in $d = 4$ no unitary interacting counterexample is known [7]; we use the two terms interchangeably for the theories of interest.

- *Maxwell theory is the free counterexample.* For $S = -\frac{1}{4}\int F^2$, scale invariance gives $\Delta_A = \frac{d-2}{2}$, $\Delta_F = \frac{d}{2}$, and the symmetric gauge-invariant tensor has

$$\begin{aligned} T_{\mu\nu} &= F_{\mu\rho}F_\nu{}^\rho - \frac{1}{4}\,\eta_{\mu\nu}F_{\rho\sigma}F^{\rho\sigma}, \\ T^\mu{}_\mu &= \left(1-\frac{d}{4}\right)F_{\rho\sigma}F^{\rho\sigma} \;=\; \partial_\mu V^\mu, \qquad V^\mu \;=\; \frac{d-4}{2}\,A_\rho F^{\rho\mu}, \end{aligned} \tag{9.7}$$

  using $\partial_\mu(F^{\mu\nu}A_\nu) = \frac{1}{2}F^{\mu\nu}F_{\mu\nu}$ on shell. The trace vanishes only in $d = 4$, where $F_{\mu\nu}$ is a primary of dimension 2 saturating its unitarity bound; in $d \neq 4$ the virial current is not a double divergence (it is not even gauge invariant), so the theory is scale but not conformally invariant.

Example 9.1 (Improved stress tensor of the free scalar)**.** The canonical tensor of the free massless scalar is symmetric but not traceless,

$$T^{\rm can}_{\mu\nu} = \partial_\mu\phi\,\partial_\nu\phi - \frac{1}{2}\,\eta_{\mu\nu}\,(\partial\phi)^2, \qquad \eta^{\mu\nu}T^{\rm can}_{\mu\nu} = \Big(1-\frac{d}{2}\Big)(\partial\phi)^2. \tag{9.8}$$

The improvement $T_{\mu\nu} = T^{\rm can}_{\mu\nu} + \xi(\eta_{\mu\nu}\Box - \partial_\mu\partial_\nu)\phi^2$ is conserved for any constant ξ, since $\partial^\mu(\eta_{\mu\nu}\Box - \partial_\mu\partial_\nu) = 0$, and its trace is

$$\eta^{\mu\nu}T_{\mu\nu} = \Big(1-\frac{d}{2}\Big)(\partial\phi)^2 + 2\xi(d-1)\Big[(\partial\phi)^2 + \phi\,\Box\phi\Big] \overset{\Box\phi=0}{=} 0 \qquad \text{for} \qquad \xi \;=\; \frac{d-2}{4(d-1)}, \tag{9.9}$$

the conformal coupling, the same number that appears in the curvature coupling $\frac{1}{2}\xi R\phi^2$ making the scalar Weyl invariant on a curved background. The free scalar is thus a CFT with $\Delta_\phi = \frac{d-2}{2}$. From now on $T_{\mu\nu}$ denotes the improved tensor: conserved, symmetric, traceless.

9.3 Two-point functions

Impose (9.1) on $G(x_1,x_2) = \langle\mathcal{O}_1(x_1)\mathcal{O}_2(x_2)\rangle$, one symmetry at a time.

1. *Translations and rotations*: $G = f(r)$ with $r = |x_1 - x_2|$.

2. *Dilatations*: $f(\lambda r) = \lambda^{-\Delta_1-\Delta_2} f(r)$ for all $\lambda > 0$, so

$$f(r) \;=\; \frac{C_{12}}{r^{\Delta_1+\Delta_2}}. \tag{9.10}$$

3. *Inversion*, which suffices since $K = IPI$. With $\Omega_I = 1/x^2$ and Lemma 8.1, $|x'_1 - x'_2| = |x_{12}|/|x_1||x_2|$, the two sides of (9.1) are

$$C_{12}\,\frac{(|x_1||x_2|)^{\Delta_1+\Delta_2}}{r^{\Delta_1+\Delta_2}} \qquad \text{and} \qquad \frac{C_{12}}{r^{\Delta_1+\Delta_2}}\,|x_1|^{2\Delta_1}\,|x_2|^{2\Delta_2}, \tag{9.11}$$

and matching the powers of $|x_1|$ and $|x_2|$ separately requires $\Delta_1+\Delta_2 = 2\Delta_1 = 2\Delta_2$. For different dimensions the correlator vanishes.

Two-point functions of primary operators are diagonal in the scaling dimension, and fixed up to normalization:

$$\langle\mathcal{O}_i(x)\,\mathcal{O}_j(y)\rangle \;=\; \frac{C_i\,\delta_{ij}}{|x-y|^{2\Delta_i}}. \tag{9.12}$$

Rescaling each operator sets $C_i = 1$, the standard normalization, after which no rescaling freedom remains.

- Orthogonality between operators of equal dimension is arranged by diagonalizing C_{ij}, which is positive definite in a unitary theory (Section 9.7); primaries and descendants, and descendants at different levels, are orthogonal, their correlators being derivatives of (9.12).

- For spinning primaries the inversion matrix carries the indices; for a vector of dimension Δ,

$$\langle\mathcal{O}_\mu(x)\,\mathcal{O}_\nu(0)\rangle \;=\; C\,\frac{I_{\mu\nu}(x)}{x^{2\Delta}}, \qquad I_{\mu\nu}(x) = \delta_{\mu\nu} - \frac{2\,x_\mu x_\nu}{x^2}. \tag{9.13}$$

9.4 Three-point functions

1. *No invariants.* Any three points can be mapped to 0, a unit vector and infinity, so no continuous conformal invariant survives and the correlator is fixed up to constants. Translations, rotations and dilatations give

$$\langle \mathcal{O}_1(x_1)\mathcal{O}_2(x_2)\mathcal{O}_3(x_3)\rangle \;=\; \frac{\lambda_{123}}{r_{12}^a\, r_{23}^b\, r_{31}^c}, \qquad a+b+c \;=\; \Delta_1+\Delta_2+\Delta_3. \tag{9.14}$$

2. *Inversion.* By Lemma 8.1 each distance transforms, and the left-hand side of (9.1) acquires $|x_1|^{a+c}|x_2|^{a+b}|x_3|^{b+c}$ while the right-hand side supplies $|x_1|^{2\Delta_1}|x_2|^{2\Delta_2}|x_3|^{2\Delta_3}$. Matching the three powers,

$$a = \Delta_1+\Delta_2-\Delta_3, \qquad b = \Delta_2+\Delta_3-\Delta_1, \qquad c = \Delta_3+\Delta_1-\Delta_2. \tag{9.15}$$

Three-point functions of scalar primaries are fixed by conformal symmetry up to one constant per triple:

$$\langle \mathcal{O}_1(x_1)\,\mathcal{O}_2(x_2)\,\mathcal{O}_3(x_3)\rangle = \frac{\lambda_{123}}{r_{12}^{\Delta_1+\Delta_2-\Delta_3}\; r_{23}^{\Delta_2+\Delta_3-\Delta_1}\; r_{31}^{\Delta_3+\Delta_1-\Delta_2}}. \tag{9.16}$$

- *CFT data.* With two-point functions normalized to unity the λ_{123} are the *structure constants* or OPE coefficients; for spinning operators a finite number of tensor structures appear, each with its own constant. The pairs $(\Delta_i, \mathrm{spin}_i)$ with the set $\{\lambda_{ijk}\}$ are the CFT data, and they determine every correlator (Section 9.6).

- *Four points.* After using the whole group, two invariants remain, the cross-ratios

$$u \;=\; \frac{r_{12}^2\, r_{34}^2}{r_{13}^2\, r_{24}^2}, \qquad v \;=\; \frac{r_{14}^2\, r_{23}^2}{r_{13}^2\, r_{24}^2}, \qquad \langle \mathcal{O}(x_1)\,\mathcal{O}(x_2)\,\mathcal{O}(x_3)\,\mathcal{O}(x_4)\rangle \;=\; \frac{g(u,v)}{r_{12}^{2\Delta}\, r_{34}^{2\Delta}}, \tag{9.17}$$

  invariant under inversions by Lemma 8.1. The function g is genuine dynamics; the OPE computes it as a sum over CFT data in three channels, and their equality, crossing symmetry, is the engine of the conformal bootstrap [7]. For us the four-point function is the first observable where bulk dynamics such as graviton exchange enters.

9.5 Currents, the stress tensor, and central charges

- *Conserved currents.* A current is a vector primary, so by (9.13)

$$\langle J_\mu(x)\, J_\nu(0)\rangle = C_J\, \frac{I_{\mu\nu}(x)}{x^{2(d-1)}}. \tag{9.18}$$

  Conservation forces the dimension: for a vector primary of dimension Δ, using $\partial_\mu x^{-2\Delta} = -2\Delta x_\mu x^{-2\Delta-2}$,

$$\partial^\mu\left(\frac{\delta_{\mu\nu}}{x^{2\Delta}} - \frac{2\,x_\mu x_\nu}{x^{2\Delta+2}}\right) = \Big[-2\Delta-2(d+1)+2(2\Delta+2)\Big]\frac{x_\nu}{x^{2\Delta+2}} = 2\,(\Delta-d+1)\,\frac{x_\nu}{x^{2\Delta+2}}, \tag{9.19}$$

  which vanishes away from the origin iff $\Delta = d-1$. Conservation, saturation of the unitarity bound of Section 9.8, and a protected canonical dimension are one and the same. C_J is physical once the current is normalized by its Ward identity.

- *The stress tensor.* A symmetric traceless tensor primary of dimension exactly d, with

$$\langle T_{\mu\nu}(x)\, T_{\rho\sigma}(0)\rangle = C_T\, \frac{\mathcal{I}_{\mu\nu,\rho\sigma}(x)}{x^{2d}}, \qquad \mathcal{I}_{\mu\nu,\rho\sigma} = \frac{1}{2}\,(I_{\mu\rho}I_{\nu\sigma} + I_{\mu\sigma}I_{\nu\rho}) - \frac{1}{d}\,\delta_{\mu\nu}\,\delta_{\rho\sigma}, \tag{9.20}$$

  the unique symmetric traceless inversion-covariant structure. Since the normalization of T is fixed by (9.3), C_T is an intrinsic number, additive over decoupled sectors, counting local degrees of freedom: proportional to the Virasoro c in $d = 2$ and to the coefficient c of the Weyl-squared term in the trace anomaly in $d = 4$. $\langle TTT\rangle$ and $\langle TJJ\rangle$ are likewise fixed up to a few constants related to a, c and C_J. These are the quantities with a clean holographic computation, and Chapter 14 matches the gravity prediction to the exact a and c of $\mathcal{N} = 4$ super Yang–Mills.

9.6 The operator product expansion

For two scalar primaries,

$$\mathcal{O}_i(x)\, \mathcal{O}_j(0) \;=\; \sum_k \frac{\lambda_{ijk}}{|x|^{\Delta_i+\Delta_j-\Delta_k}} \Big(\mathcal{O}_k(0) + \underbrace{\#\, x^\mu \partial_\mu \mathcal{O}_k(0) + \cdots}_{\text{descendants}}\Big), \tag{9.21}$$

the sum running over primaries of any spin, the leading power fixed by dimensional analysis.

- *One number per family.* The constant multiplying $\mathcal{O}_k$ is the structure constant of (9.16): inserting (9.21) into $\langle \mathcal{O}_i(x)\mathcal{O}_j(0)\mathcal{O}_k(y)\rangle$ and using orthogonality of two-point functions projects out the $\mathcal{O}_k$ term. The descendant coefficients are fixed by conformal symmetry level by level (Section 9.7).

- *Everything follows.* Applying the OPE to nearby pairs reduces any n-point function to $(n-1)$-point functions and, iterating, to two-point functions: the CFT data determine all correlators. In the four-point function the OPE in the $12 \to 34$ channel organizes $g(u, v)$ into conformal blocks, one per exchanged primary, and equality with the $14 \to 23$ channel is the crossing equation.

- *Convergence.* Unlike Wilson's OPE in a generic QFT, the conformal OPE converges at finite separation, as radial quantization shows next.

9.7 Radial quantization and the state–operator correspondence

Canonical quantization needs a foliation. In Euclidean signature a CFT is naturally foliated by concentric spheres S^{d-1} around the origin, evolved radially by the dilatation: *radial quantization.*

- *The cylinder map.* With $r = \mathrm{e}^\tau$,

$$\mathrm{d}s^2_{\mathbb{R}^d} = \mathrm{d}r^2 + r^2\, \mathrm{d}\Omega^2_{d-1} = \mathrm{e}^{2\tau}\left(\mathrm{d}\tau^2 + \mathrm{d}\Omega^2_{d-1}\right), \qquad \mathcal{O}_{\text{cyl}}(\tau, n) \;=\; r^\Delta\, \mathcal{O}_{\text{plane}}(rn), \tag{9.22}$$

  so flat space minus the origin is conformal to the cylinder $\mathbb{R}_\tau \times S^{d-1}$, and up to the anomaly of Chapter 14 the theory on the plane is the theory on the cylinder (Figure 9.1). Dilatations become time translations, so the dilatation is the cylinder Hamiltonian, $H_{\text{cyl}} = D$; the origin is $\tau = -\infty$; the inversion is $\tau \to -\tau$; rotations of the sphere give $\mathrm{SO}(d)$ quantum numbers.

- *Energies are dimensions.* Mapping (9.12) with $|x_1 - x_2|^2 = r_1^2 + r_2^2 - 2r_1r_2\cos\theta$,

$$\langle \mathcal{O}_{\text{cyl}}(\tau_1, n_1)\, \mathcal{O}_{\text{cyl}}(\tau_2, n_2)\rangle = \frac{(r_1r_2)^\Delta}{|x_1 - x_2|^{2\Delta}} = \Big[\,2\cosh(\tau_1 - \tau_2) - 2\cos\theta\,\Big]^{-\Delta}, \tag{9.23}$$

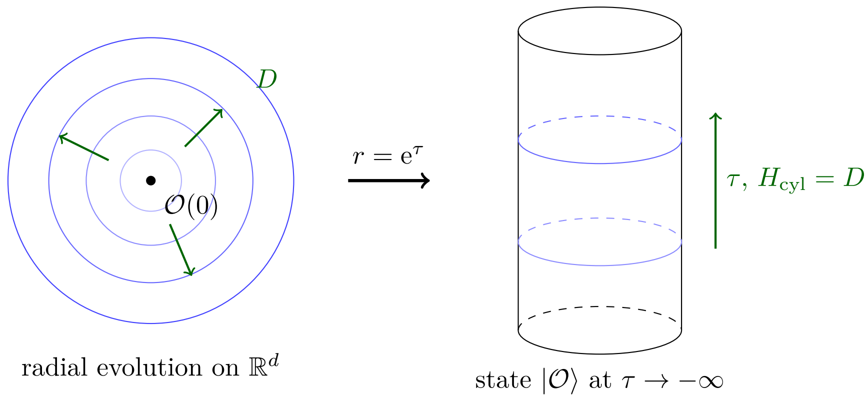


Figure 9.1: Radial quantization. Left: Euclidean $\mathbb{R}^d$ foliated by spheres around the origin; an operator inserted at the origin creates the state on each sphere, and dilatations move outward from sphere to sphere. Right: after the Weyl rescaling (9.22), the same theory lives on the cylinder $\mathbb{R} \times S^{d-1}$; the dilatation becomes the Hamiltonian and the operator insertion becomes an initial state in the infinite past.

which depends on $\tau_1 - \tau_2$ only and decays as $\mathrm{e}^{-\Delta(\tau_1 - \tau_2)}$: the lowest state exchanged has energy Δ. In Lorentzian language H_{cyl} is the conformal energy (8.21), which reappears in Chapter 10 as global time translation in AdS.

9.7.1 The state–operator correspondence

- *Operator to state.* The path integral over the interior of the unit sphere with $\mathcal{O}$ at the origin is a wavefunctional on the sphere, the state $|\mathcal{O}\rangle = \mathcal{O}(0)|0\rangle$ of the radial Hilbert space $\mathcal{H}_{S^{d-1}}$. On the cylinder the insertion sits at $\tau = -\infty$ and infinite Euclidean evolution projects onto the eigenstate of D with energy Δ and the spin of $\mathcal{O}$.

- *State to operator.* Any eigenstate $|\psi\rangle$ of D on the unit sphere serves as boundary condition for the interior path integral; radial evolution $\mathrm{e}^{-\tau D}$ squeezes it into an arbitrarily small ball, and a ball with prescribed boundary data and nothing inside is, to every operator outside, a local insertion $\mathcal{O}_\psi(0)$ with $\mathcal{O}_\psi(0)|0\rangle = |\psi\rangle$.

State–operator correspondence. In a conformal field theory there is a one-to-one correspondence between local operators and states of the theory on S^{d-1}:

$$\mathcal{O} \;\longleftrightarrow\; |\mathcal{O}\rangle = \mathcal{O}(0)|0\rangle, \qquad \Delta_{\mathcal{O}} \;=\; E_{\text{cyl}}, \qquad \text{spin}_{\mathcal{O}} = \mathrm{SO}(d) \text{ quantum numbers.} \tag{9.24}$$

Primaries correspond to lowest-weight states ($K_\mu|\mathcal{O}\rangle = 0$) and descendants to their P_μ-excitations.

This is special to CFT: in a massive theory the radius of the sphere matters. Two consequences follow.

- *Convergence of the OPE. ⋆ Self-study.* With the other insertions outside a sphere separating $\{0, x\}$ from them, the interior path integral is a state on that sphere, expandable in the eigenbasis of D, which is organized into conformal multiplets:

$$\mathcal{O}_1(x)\,\mathcal{O}_2(0)\,|0\rangle = \sum_k \sum_{n\geq 0} c_n^{\mu_1\cdots\mu_n}(x)\,P_{\mu_1}\cdots P_{\mu_n}|\mathcal{O}_k\rangle. \tag{9.25}$$

Acting with D in two ways, on the left through $[D, \mathcal{O}_i(x)] = (\Delta_i + x\cdot\partial)\mathcal{O}_i(x)$ and on the right through $D\,P^n|\mathcal{O}_k\rangle = (\Delta_k + n)P^n|\mathcal{O}_k\rangle$, and equating coefficients of the independent basis states,

$$x\cdot\partial_x\, c_n^{\mu_1\cdots\mu_n}(x) = \big(\Delta_k + n - \Delta_1 - \Delta_2\big)\, c_n^{\mu_1\cdots\mu_n}(x), \tag{9.26}$$

so with the tensor factors $x^{\mu_1}\cdots x^{\mu_n}$ carrying degree n, every level scales with the same power $|x|^{\Delta_k-\Delta_1-\Delta_2}$:

$$\mathcal{O}_1(x)\,\mathcal{O}_2(0)\,|0\rangle = \sum_k \lambda_{12k}\,|x|^{\Delta_k-\Delta_1-\Delta_2}\Big(|\mathcal{O}_k\rangle + a_1\,x^\mu P_\mu|\mathcal{O}_k\rangle \\ + a_2\,x^\mu x^\nu P_\mu P_\nu|\mathcal{O}_k\rangle + b_2\,x^2P^2|\mathcal{O}_k\rangle + \cdots\Big). \tag{9.27}$$

Acting with K_μ in two ways fixes the constants: on the left by (8.31), in the Euclidean form $[K_\mu, \mathcal{O}_1(x)] = (2\Delta_1 x_\mu + 2x_\mu x\cdot\partial - x^2\partial_\mu)\mathcal{O}_1(x)$ and $[K_\mu, \mathcal{O}_2(0)]|0\rangle = 0$, on the right by lowering,

$$K_\mu|\mathcal{O}_k\rangle = 0, \qquad K_\mu P_\nu|\mathcal{O}_k\rangle = [K_\mu, P_\nu]|\mathcal{O}_k\rangle = 2\Delta_k\,\delta_{\mu\nu}|\mathcal{O}_k\rangle. \tag{9.28}$$

Matching the coefficient of $|\mathcal{O}_k\rangle$ at order $x^{\kappa+1}$, with $\kappa = \Delta_k - \Delta_1 - \Delta_2$, the left gives $(2\Delta_1 + \kappa)x_\mu$ and the right $2\Delta_k a_1 x_\mu$:

$$a_1 = \frac{\Delta_k + \Delta_1 - \Delta_2}{2\Delta_k}, \tag{9.29}$$

and at each next order K_μ maps the unknown level-$(n{+}1)$ coefficients onto the known level-n ones. One constant λ_{12k} per primary survives, and translating states back to operators this is the OPE (9.21). As the expansion of a normalizable state in an orthogonal basis it converges whenever a separating sphere exists: the conformal OPE has a finite radius of convergence, like a Taylor series.

- *Conjugation.* Hermitian conjugation on the cylinder involves $\tau \to -\tau$, on the plane the inversion, because Euclidean evolution is not unitary and reflection positivity replaces unitarity.[1] In Euclidean conventions, keeping P_μ and K_μ as in (8.12) and defining $M_{\mu\nu} \equiv \mathrm{i}J_{\mu\nu} = x_\mu\partial_\nu - x_\nu\partial_\mu$ and $D \equiv -\mathrm{i}D_{\mathrm{Lor}}$ (with a spin part $\Sigma_{\mu\nu}$ added to $M_{\mu\nu}$ on operators with indices), all structure constants are real,

$$\begin{aligned} [D, P_\mu] &= P_\mu, & [D, K_\mu] &= -K_\mu, \\ [K_\mu, P_\nu] &= 2\,\delta_{\mu\nu}D - 2M_{\mu\nu}, & [M_{\mu\nu}, P_\rho] &= \delta_{\nu\rho}P_\mu - \delta_{\mu\rho}P_\nu, \end{aligned} \tag{9.30}$$

and since $IP_\mu I = K_\mu$,

$$P_\mu^\dagger \;=\; K_\mu, \qquad D^\dagger = D, \qquad M_{\mu\nu}^\dagger = -M_{\mu\nu}\;: \tag{9.31}$$

D has real eigenvalue Δ on $|\mathcal{O}\rangle$, and P_μ, K_μ are an honest raising/lowering pair on a positive-definite Hilbert space.

9.8 Unitarity bounds

In a unitary CFT every state has non-negative norm, and applied to descendants this gives sharp lower bounds on dimensions. Take a primary $|\Delta\rangle$ with $\langle\Delta|\Delta\rangle = 1$, $D|\Delta\rangle = \Delta|\Delta\rangle$, $K_\mu|\Delta\rangle = 0$, and use (9.30) and (9.31).

[1] Reflection positivity, the Euclidean avatar of unitarity: with $\Theta\,\mathcal{O}(\tau,\vec{x})\,\Theta^{-1} = \mathcal{O}^\dagger(-\tau,\vec{x})$, one requires $\langle(\Theta F)F\rangle \geq 0$ for every F built from insertions at $\tau < 0$. The correlator computes $\langle\psi_F|\psi_F\rangle$ for the state F prepares on $\tau = 0$, so this is the condition under which the path integral defines a positive-definite inner product, from which a unitary Lorentzian theory is reconstructed [50].

1. *Level one, scalar.* For a scalar the spin Casimir $\sum_{\mu<\nu} \|M_{\mu\nu}|\Delta\rangle\|^2$ vanishes, so $M_{\mu\nu}|\Delta\rangle = 0$ for all $\mu\nu$ and
$$\langle\Delta|K_\nu\, P_\mu|\Delta\rangle = \langle\Delta|(2\delta_{\mu\nu} D - 2M_{\nu\mu})|\Delta\rangle = 2\,\Delta\;\delta_{\mu\nu}\, : \tag{9.32}$$
$\Delta \geq 0$, with equality only for the identity.

2. *Level two, the scalar bound.* Moving one K through the two P's of $P^2|\Delta\rangle \equiv P_\nu P_\nu|\Delta\rangle$, with $DP_\nu|\Delta\rangle = (\Delta+1)P_\nu|\Delta\rangle$ and $[M_{\mu\nu}, P_\nu] = (d-1)P_\mu$ from the last commutator in (9.30),
$$K_\mu\, P^2|\Delta\rangle = 2(\Delta+1)\, P_\mu|\Delta\rangle - 2\,[M_{\mu\nu}, P_\nu]|\Delta\rangle + 2\Delta\, P_\mu|\Delta\rangle = 4\left(\Delta - \frac{d-2}{2}\right) P_\mu|\Delta\rangle, \tag{9.33}$$
and then by (9.32)
$$\left\|P^2|\Delta\rangle\right\|^2 = \langle\Delta|K_\mu K_\mu\, P^2|\Delta\rangle = 8\,d\,\Delta\left(\Delta - \frac{d-2}{2}\right) \qquad \Longrightarrow \qquad \Delta \;\geq\; \frac{d-2}{2}. \tag{9.34}$$
The other level-two channels and higher levels give nothing stronger. At saturation $P^2|\Delta\rangle$ is null and must vanish, which by the state–operator correspondence is $\Box\mathcal{O} = 0$: a scalar primary saturating the bound is a free field.

9.8.1 Higher spin: currents and the stress tensor ⋆

For a primary in the symmetric-traceless spin-ℓ representation the binding constraint appears at level one.

1. With a, b labelling an orthonormal basis of the spin space and $(\Sigma_{\mu\nu})_{ab} = \langle a|M_{\mu\nu}|b\rangle$ the representation matrices, the level-one Gram matrix is
$$\begin{aligned} G \equiv \langle\Delta; a|\, K_\nu P_\mu\, |\Delta; b\rangle = 2\Delta\,\delta_{\mu\nu}\,\delta_{ab} - 2\langle a|M_{\nu\mu}|b\rangle \;&=\; 2\Delta\,\mathbf{1}\otimes\mathbf{1} \;-\; 2B, \\ B \equiv \frac{1}{2}\sum_{\alpha\beta} m_{\alpha\beta}\otimes\Sigma_{\alpha\beta}, \end{aligned} \tag{9.35}$$
with $(m_{\mu\nu})_{\rho\sigma} = \delta_{\mu\rho}\delta_{\nu\sigma} - \delta_{\nu\rho}\delta_{\mu\sigma}$ the vector generators: a spin–orbit problem on $V\otimes R_\ell$, the SO(d) analogue of $2\vec{L}\cdot\vec{S}$.

2. The Casimir trick: with $J_{\alpha\beta} = m_{\alpha\beta}\otimes\mathbf{1} + \mathbf{1}\otimes\Sigma_{\alpha\beta}$ and $C_2 \equiv -\frac{1}{2}\sum_{\alpha\beta} J_{\alpha\beta}J_{\alpha\beta}$, squaring gives
$$B \;=\; \tfrac{1}{2}\Big[C_2(V) + C_2(R_\ell) - C_2(V\otimes R_\ell)\Big], \tag{9.36}$$
constant on each irreducible component of $V\otimes R_\ell = R_{\ell+1}\oplus R_{\ell-1}\oplus R_{[\ell,1]}$. With $C_2(R_s) = s(s+d-2)$, $C_2(V) = d-1$ and $C_2(R_{[\ell,1]}) = \ell(\ell+d-2)+d-3$,
$$b_{\ell+1} = -\ell, \qquad b_{[\ell,1]} = 1, \qquad b_{\ell-1} = \ell+d-2. \tag{9.37}$$

3. Positivity of G requires $\Delta \geq b$ in every channel; the binding one is $b_{\ell-1}$, the spin-$(\ell-1)$ channel obtained by contracting the index of P_μ with one index of the primary [7].

$$\boxed{\Delta \;\geq\; \ell+d-2 \quad (\ell\geq 1), \qquad \Delta \;\geq\; \frac{d-2}{2} \quad (\ell=0,\; \mathcal{O}\neq\mathbf{1}).} \tag{9.38}$$

At saturation the null descendant in the spin-$(\ell-1)$ channel is the operator equation $\partial^\mu\mathcal{O}_{\mu\mu_2\cdots\mu_\ell} = 0$: the operator is a conserved current, and conversely a conserved spin-ℓ current sits exactly at $\Delta = \ell+d-2$.

9.8.2 Consequences

- *Null descendants.* At the bound a descendant has zero norm: $P^2|\Delta\rangle$ for a scalar at $\Delta = \frac{d-2}{2}$, by (9.34), and $P^\mu|\Delta;\mu\mu_2\cdots\mu_\ell\rangle$ for spin ℓ at $\Delta = \ell + d - 2$. A zero-norm vector in a positive-definite Hilbert space is orthogonal to every state, hence zero, and by the state–operator correspondence its vanishing is an operator equation: $\Box\mathcal{O} = 0$ or $\partial^\mu\mathcal{O}_{\mu\cdots} = 0$. A multiplet with such a missing descendant is *short*.

- *Short multiplets are protected.* Dimensions vary continuously with the couplings, but the number of states at a level is an integer. If a short multiplet left the bound, its null descendant would regain a norm and reappear as a state, a discontinuous jump. This can happen only by *recombination*, a second multiplet with the spin and dimension of the missing descendant joining the first into a long multiplet, which is then free to move. Without such a partner the dimension is pinned to the bound at any coupling: this is the mechanism behind non-renormalization theorems.

- *Conserved currents.* J_μ has $\Delta = d-1$ and $T_{\mu\nu}$ has $\Delta = d$ at any coupling; in $d = 4$, $\Delta_J = 3$ and $\Delta_T = 4$. A current leaves the bound only by recombining with a scalar of dimension d that becomes its divergence, $\partial^\mu J_\mu = \mathcal{O}$, which is what a broken symmetry does. In $\mathcal{N} = 4$ SYM the same mechanism, for the superconformal algebra, protects the chiral (BPS) operators whose dimensions are fixed by their R-charges, the operators followed to strong coupling in Chapter 17.

- *Free fields.* A scalar primary saturates $\Delta = \frac{d-2}{2}$ only if $\Box\mathcal{O} = 0$. For an elementary scalar, with classical dimension $\Delta_0 = \frac{d-2}{2}$ from (8.2), the anomalous dimension $\gamma = \Delta - \Delta_0$ of Section 8.5 is therefore $\gamma \geq 0$, and $\gamma > 0$ once the field interacts. No such sign constraint applies to composites such as ϕ^2, of classical dimension $d-2$.

- *Boundedness.* Dimensions in a unitary CFT are real and bounded below, with the identity the unique operator at $\Delta = 0$, which justifies the lowest-weight structure of Chapter 8.

A CFT is specified by a discrete list of primaries obeying the unitarity bounds and by the structure constants; symmetry then determines two- and three-point functions and, through the convergent OPE, all correlators. Missing is a mechanism computing the data $(\Delta_i, \lambda_{ijk})$ for a given theory. For a special class of CFTs the AdS/CFT correspondence provides it, reading the data off a gravity computation in anti-de Sitter space, the subject of the next chapter.

Chapter 10

Anti-de Sitter Space

We turn to the gravitational side: anti-de Sitter space. Constructed as a hyperboloid in $\mathbb{R}^{2,d}$, AdS_{d+1} has isometry group $\mathrm{SO}(2,d)$, the conformal group of d-dimensional Minkowski space, manifest from the start. We develop global and Poincaré coordinates and the Euclidean continuation, and study propagation: light reaches the boundary in finite time while massive particles never do, so AdS is a gravitational box whose walls are the conformal boundary. We derive the mass–dimension relation $\Delta(\Delta - d) = m^2L^2$ and the Breitenlohner–Freedman bound, and close with a table of symmetries and scales.

10.1 The hyperboloid and its isometries

- *Einstein spaces.* Gravity in $D = d + 1$ dimensions with a cosmological constant,
$$S \;=\; \frac{1}{16\pi G}\int \mathrm{d}^{d+1}x\sqrt{|g|}\;(R - 2\Lambda)\,, \qquad R_{\mu\nu} - \frac{1}{2}\,R\,g_{\mu\nu} + \Lambda\,g_{\mu\nu} \;=\; 0, \tag{10.1}$$
has, by tracing, $R = \frac{2(d+1)}{d-1}\Lambda$ and hence $R_{\mu\nu} = \frac{2\Lambda}{d-1}g_{\mu\nu}$: every solution is an Einstein space.

- *Maximal symmetry.* If the full Riemann tensor is built from the metric, $R_{\mu\nu\rho\sigma} = \pm\ell^{-2}(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho})$, the space admits the maximal number $\frac{(d+1)(d+2)}{2}$ of Killing vectors and solves (10.1) automatically. There are four cases: the sphere S^{d+1} and hyperbolic space H^{d+1} in Euclidean signature, de Sitter dS_{d+1} and anti-de Sitter AdS_{d+1} in Lorentzian signature, all quadrics in a flat $(d+2)$-dimensional space. For AdS_{d+1} of radius L,
$$R_{\mu\nu\rho\sigma} = -\frac{1}{L^2}\left(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}\right), \qquad R_{\mu\nu} = -\frac{d}{L^2}\,g_{\mu\nu}, \qquad \Lambda = -\frac{d(d-1)}{2L^2}. \tag{10.2}$$

- *Definition.* In $\mathbb{R}^{2,d}$ with coordinates $X^M = (X^0, X^1, \ldots, X^d, X^{d+1})$ and two time directions,
$$\mathrm{d}s^2 \;=\; -(\mathrm{d}X^0)^2 + \sum_{i=1}^{d}(\mathrm{d}X^i)^2 - (\mathrm{d}X^{d+1})^2, \qquad -(X^0)^2 + \sum_{i=1}^{d}(X^i)^2 - (X^{d+1})^2 \;=\; -L^2, \tag{10.3}$$
AdS_{d+1} is the quadric with the induced metric.

> The isometry group of AdS_{d+1} is $\mathrm{SO}(2,d)$, with $\frac{(d+1)(d+2)}{2}$ generators J_{MN} acting linearly on the embedding coordinates (the full group of the quadric is $\mathrm{O}(2,d)$, the extra reflections being the counterpart of the inversion of Section 8.3). This is *the same group* as the conformal group of d-dimensional Minkowski space constructed in Chapter 8—including the index structure: the generators J_{MN} of Chapter 8, assembled from $P_\mu, K_\mu, D, J_{\mu\nu}$, are the AdS isometries.

This group-theoretic identity is the kinematical foundation of AdS/CFT: a quantum gravity theory in AdS_{d+1} has exactly the symmetries of a CFT_d. The rest of the chapter makes the match concrete, chart by chart.

10.2 Global coordinates and the cylinder

1. *The global chart.* Solve the quadric by
$$X^0 = L\cosh\rho\,\cos\tau, \qquad X^{d+1} = L\cosh\rho\,\sin\tau, \qquad X^i = L\sinh\rho\, n^i, \tag{10.4}$$
with $\rho \geq 0$, $\tau \in [0, 2\pi)$ and n^i a unit vector on S^{d-1}; the constraint holds because $-\cosh^2\rho + \sinh^2\rho = -1$.

2. *Induced metric.* With the cross terms $\sum_i n^i \mathrm{d}n^i$ vanishing since $n^2 = 1$,
$$\begin{aligned}(\mathrm{d}X^0)^2 + (\mathrm{d}X^{d+1})^2 &= L^2\left(\sinh^2\rho\,\mathrm{d}\rho^2 + \cosh^2\rho\,\mathrm{d}\tau^2\right),\\ \sum_i (\mathrm{d}X^i)^2 &= L^2\left(\cosh^2\rho\,\mathrm{d}\rho^2 + \sinh^2\rho\,\mathrm{d}\Omega_{d-1}^2\right),\end{aligned} \tag{10.5}$$
and subtracting according to (10.3):

$$\boxed{\mathrm{d}s^2 \;=\; L^2\left(-\cosh^2\!\rho\,\mathrm{d}\tau^2 + \mathrm{d}\rho^2 + \sinh^2\!\rho\,\mathrm{d}\Omega_{d-1}^2\right) \qquad \text{(global coordinates)}. \quad (10.6)}$$

- *Static form.* With $r = L\sinh\rho$, $t = L\tau$,
$$\mathrm{d}s^2 = -\Big(1 + \frac{r^2}{L^2}\Big)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{1 + \dfrac{r^2}{L^2}} + r^2\,\mathrm{d}\Omega_{d-1}^2, \tag{10.7}$$
a static spherically symmetric spacetime with $f(r) = 1 + r^2/L^2$: no horizon, and a potential growing without bound. This is the form in which AdS reappears in Chapter 18 with a black hole added.

- *Universal cover.* On the hyperboloid τ is an angle and closed timelike curves pass through every point. One passes to the universal cover, $\tau \in \mathbb{R}$, which is what AdS_{d+1} means from now on. The matching statement on the CFT side is that quantum theories represent the universal cover of $\mathrm{SO}(2, d)$, with continuous conformal energies: the unwrapped $\mathrm{SO}(2)$ is the rotation of the (X^0, X^{d+1}) plane, i.e. global time translation.

- *Conformal boundary.* With $\tan\theta = \sinh\rho$, $\theta \in [0, \frac{\pi}{2})$,
$$\mathrm{d}s^2 \;=\; \frac{L^2}{\cos^2\theta}\left(-\mathrm{d}\tau^2 + \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\Omega_{d-1}^2\right): \tag{10.8}$$
up to the conformal factor, one half of the Einstein static universe $\mathbb{R} \times S^d$, a solid cylinder whose rim $\theta = \frac{\pi}{2}$ is at infinite proper but finite conformal distance (Figure 10.1).

> AdS_{d+1} has a conformal boundary
> $$\partial\,\mathrm{AdS}_{d+1} \;=\; \mathbb{R}_\tau \times S^{d-1},$$
> a timelike hypersurface at infinite proper distance but finite conformal distance. Only the *conformal class* of the boundary metric is well defined, since rescaling the arbitrary function used to strip the divergence rescales the boundary metric. $\mathbb{R}\times S^{d-1}$ is precisely the cylinder on which we formulated radial quantization of a CFT_d in Chapter 9.

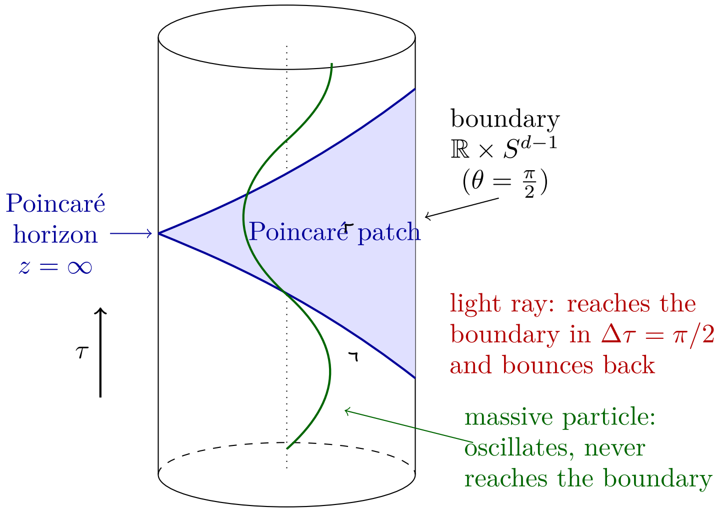


Figure 10.1: Global anti-de Sitter space as a solid cylinder (drawn after the conformal rescaling (10.8)). The boundary of the cylinder is the conformal boundary $\mathbb{R}\times S^{d-1}$. Light rays (red; the arrows indicate the direction of propagation) travel at 45° and reach the boundary in finite coordinate time; massive particles (green) oscillate about the centre like in a harmonic trap. The shaded wedge is the region covered by Poincaré coordinates, bounded by the Poincaré horizon $z=\infty$; its boundary portion is conformal to d-dimensional Minkowski space.

The boundary carries a conformal class rather than a metric, which is exactly what a conformal field theory couples to.

10.3 Poincaré coordinates

1. *A Minkowski slicing.* With x^μ Minkowski coordinates, $x^2=\eta_{\mu\nu}x^\mu x^\nu$ and $z>0$, set

$$X^\mu=\frac{L\,x^\mu}{z},\qquad X^d=\frac{z^2+x^2-L^2}{2z},\qquad X^{d+1}=\frac{z^2+x^2+L^2}{2z}. \tag{10.9}$$

The constraint holds since $\eta_{\mu\nu}X^\mu X^\nu=L^2x^2/z^2$ and $(X^d)^2-(X^{d+1})^2=(X^d-X^{d+1})(X^d+X^{d+1})=-(L^2/z)(z^2+x^2)/z$.

2. *Induced metric.* With the light-cone pair $u=X^{d+1}-X^d=L^2/z$, $w=X^{d+1}+X^d=(z^2+x^2)/z$, the ambient metric is $\eta_{\mu\nu}\mathrm{d}X^\mu\mathrm{d}X^\nu-\mathrm{d}u\,\mathrm{d}w$, and

$$\begin{aligned}\eta_{\mu\nu}\mathrm{d}X^\mu\mathrm{d}X^\nu&=L^2\left(\frac{\mathrm{d}x^2}{z^2}-\frac{2\,x\!\cdot\!\mathrm{d}x\;\mathrm{d}z}{z^3}+\frac{x^2\,\mathrm{d}z^2}{z^4}\right),\\ -\mathrm{d}u\,\mathrm{d}w&=L^2\left(\frac{\mathrm{d}z^2}{z^2}+\frac{2\,x\!\cdot\!\mathrm{d}x\;\mathrm{d}z}{z^3}-\frac{x^2\,\mathrm{d}z^2}{z^4}\right),\end{aligned} \tag{10.10}$$

from $\mathrm{d}u=-L^2\mathrm{d}z/z^2$ and $\mathrm{d}w=\mathrm{d}z+2x\!\cdot\!\mathrm{d}x/z-x^2\mathrm{d}z/z^2$. The cross terms cancel in the sum:

$$\boxed{\mathrm{d}s^2\;=\;L^2\,\frac{\mathrm{d}z^2+\eta_{\mu\nu}\,\mathrm{d}x^\mu\mathrm{d}x^\nu}{z^2}\qquad\text{(Poincaré coordinates, } z>0).} \tag{10.11}$$

Equivalent forms trade z for $u=L^2/z$ (in the convention of [2], $\mathrm{d}s^2=L^2\mathrm{d}u^2/u^2+(u^2/L^2)\mathrm{d}x^2$) or for $z=L\mathrm{e}^{-\tilde r/L}$ (the domain-wall form $\mathrm{d}\tilde r^2+\mathrm{e}^{2\tilde r/L}\mathrm{d}x^2$).

- *Warped Minkowski slices.* At fixed z the geometry is Minkowski space with warp factor L^2/z^2: an observer at z sees boundary lengths rescaled by L/z. Small z magnifies, large z shrinks; this is the geometric seed of the UV/IR relation of Section 12.5.

- *Boundary.* Stripping L^2/z^2 at $z = 0$ leaves the flat metric: the conformal boundary of the patch is d-dimensional Minkowski space, conformal to one diamond of the cylinder $\mathbb{R} \times S^{d-1}$, whose full form is its conformal compactification.

- *Poincaré horizon.* As $z \to \infty$, $g_{tt} = -L^2/z^2 \to 0$: $z = \infty$ is a Killing horizon of Poincaré time, with nothing singular there, since the curvature (10.2) is constant. The patch covers the wedge $u > 0$ of the hyperboloid, about half of it (Figure 10.1); global coordinates continue across.

- *Isometries made manifest.* Global coordinates display $\mathrm{SO}(2) \times \mathrm{SO}(d)$: translations of τ, the conformal Hamiltonian, and rotations of S^{d-1}. Poincaré coordinates display $\mathrm{ISO}(1, d-1) \times \mathrm{SO}(1, 1)$: the Poincaré group on each slice, P_μ and $J_{\mu\nu}$ of the boundary, and the rescaling
$$(z, x^\mu) \;\longmapsto\; (\lambda z,\; \lambda x^\mu), \tag{10.12}$$
which preserves (10.11) and acts on the boundary as the dilatation D. The remaining d isometries act on the boundary as K_μ and exist by the embedding construction. Every conformal generator of the boundary extends to an isometry of the bulk, and (10.12) shows that the new direction z transforms like an inverse energy scale.

10.4 Euclidean AdS: hyperbolic space

- *Continuation.* $\tau \to -\mathrm{i}\tau_E$, i.e. $X^{d+1} \to -\mathrm{i}X^{d+1}$, turns the quadric into $-(X^0)^2 + \sum_{i=1}^{d+1}(X^i)^2 = -L^2$ with $X^0 \geq L$: one sheet of a two-sheeted hyperboloid in $\mathbb{R}^{1,d+1}$, hyperbolic space H^{d+1}, with isometry group $\mathrm{SO}(1, d+1)$, the conformal group of Euclidean $\mathbb{R}^d$. The Poincaré chart now covers the whole space,
$$\mathrm{d}s^2 \;=\; L^2\,\frac{\mathrm{d}z^2 + \delta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu}{z^2}, \qquad z > 0, \tag{10.13}$$
the higher-dimensional upper half-space.

- *What changes.* The Poincaré horizon disappears: $z \to \infty$ at finite x is a single point at infinite distance. The boundary compactifies: $\partial H^{d+1} = \mathbb{R}^d \cup \{\infty\} = S^d$, consistent with continuing $\mathbb{R} \times S^{d-1}$ and compactifying. The ball model makes the sphere explicit: the map $y = \big(2x,\, |x|^2 + z^2 - 1\big)/\big(|x|^2 + (z+1)^2\big)$ sends the half space to the unit ball with $1 - |y|^2 = 4z/(|x|^2 + (z+1)^2)$, and (10.13) becomes
$$\mathrm{d}s^2 \;=\; \frac{4L^2\,\delta_{ab}\,\mathrm{d}y^a \mathrm{d}y^b}{(1 - |y|^2)^2}, \qquad |y| < 1, \tag{10.14}$$
with the boundary sphere $|y| = 1$ at infinite distance.

- *The hyperbolic plane.* For $d = 1$, with $w = t_E + \mathrm{i}z$, (10.13) is $L^2 \mathrm{d}w\,\mathrm{d}\bar{w}/(\mathrm{Im}\,w)^2$, whose isometries are the real Möbius maps $\mathrm{PSL}(2, \mathbb{R}) \simeq \mathrm{SO}(1, 2)$ and whose geodesics are half-circles meeting the boundary at right angles; the Cayley map $y = (w - \mathrm{i})/(w + \mathrm{i})$ gives the disk form of (10.14).

10.5 Geodesics: AdS as a box

We set $L = 1$ where harmless.

- *Light reaches the boundary in finite time.* For a radial null ray in (10.6), $\mathrm{d}\tau = \mathrm{d}\rho/\cosh\rho$, so from the centre to $\rho = \infty$
$$\Delta\tau \;=\; \int_0^\infty \frac{\mathrm{d}\rho}{\cosh\rho} \;=\; \Big[\arctan(\sinh\rho)\Big]_0^\infty \;=\; \frac{\pi}{2}, \tag{10.15}$$
i.e. $\Delta t = \pi L/2$ in (10.7); in the frame (10.8) light travels at 45° to the rim at $\theta = \frac{\pi}{2}$. The proper distance is infinite, yet light gets there and back in a time of order L.

- *AdS is not globally hyperbolic.* Data on a complete slice do not determine the future, since information enters from the timelike boundary. Dynamics requires boundary conditions at $\theta = \frac{\pi}{2}$, reflecting ones being natural, and the boundary stays in causal contact with the interior forever: AdS is a covariant, maximally symmetric box. In Chapter 18 this makes black holes thermodynamically stable; in AdS/CFT the need for boundary data is the classical shadow of a theory defined on the boundary.

- *Massive particles never get out.* Consider a radial timelike geodesic of (10.6), $\rho(\lambda)$, $\tau(\lambda)$, with λ the proper time in units of L and dots $\mathrm{d}/\mathrm{d}\lambda$. Two facts fix the motion: ∂_τ is a Killing vector, so the momentum conjugate to τ in the geodesic Lagrangian $\frac{1}{2}g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu$, namely $-g_{\tau\tau}\dot{\tau}/L^2 = \cosh^2\!\rho\,\dot{\tau} \equiv E$, is conserved; and the tangent vector is unit timelike, $g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu/L^2 = -\cosh^2\!\rho\,\dot{\tau}^2 + \dot{\rho}^2 = -1$. Eliminating $\dot{\tau}$ gives
$$\dot{\rho}^2 \;=\; \frac{E^2 - \cosh^2\rho}{\cosh^2\rho}, \qquad \sinh\rho(\lambda) \;=\; \sqrt{E^2-1}\,\sin\lambda : \tag{10.16}$$
a turning point at $\cosh\rho_{\max} = E$, and every bound orbit a harmonic oscillation of proper period $2\pi L$ independent of the amplitude. AdS confines massive particles like an isotropic harmonic trap of frequency $1/L$, and all geodesics from a point refocus at the antipode after $\Delta\tau = \pi$.

- *Boundary-anchored spacelike geodesics* (Figure 10.2). On a constant-t slice of the Poincaré patch, a copy of H^d with metric $L^2(\mathrm{d}z^2 + \mathrm{d}\vec{x}^2)/z^2$, consider the geodesic anchored at $x^1 = \pm\ell/2$ on the boundary, a curve $z(x)$ in the (x^1, z) plane.

  1. Its length functional is
  $$\text{Length} \;=\; L\int \mathrm{d}x\,\frac{\sqrt{1+z'^2}}{z}, \tag{10.17}$$
  and since the integrand has no explicit x-dependence its "Hamiltonian" is conserved:[1]
  $$z'\,\frac{\partial}{\partial z'}\left(\frac{\sqrt{1+z'^2}}{z}\right) - \frac{\sqrt{1+z'^2}}{z} \;=\; -\,\frac{1}{z\sqrt{1+z'^2}} \;=\; -\,\frac{1}{z_*}, \tag{10.18}$$
  with z_* the turning point, where $z' = 0$.
  2. Thus $z\sqrt{1+z'^2} = z_*$, i.e. $z'^2 = (z_*^2 - z^2)/z^2$, whose solution through $x = \pm\ell/2$ at $z = 0$ is the semicircle
  $$z^2 + x^2 \;=\; z_*^2, \qquad z_* \;=\; \frac{\ell}{2}, \tag{10.19}$$
  meeting the boundary orthogonally as in the hyperbolic plane. A boundary region of size ℓ probes depth $z \sim \ell$: the UV/IR relation made quantitative.

[1] In the language of classical mechanics, with x playing the role of time, the integrand $\sqrt{1+z'^2}/z$ is a Lagrangian and the combination below its Hamiltonian, or Routhian; see §41 of Landau–Lifshitz [51], whose eq. (41.6) is precisely the equation below with $\dot{\xi} = z'$.

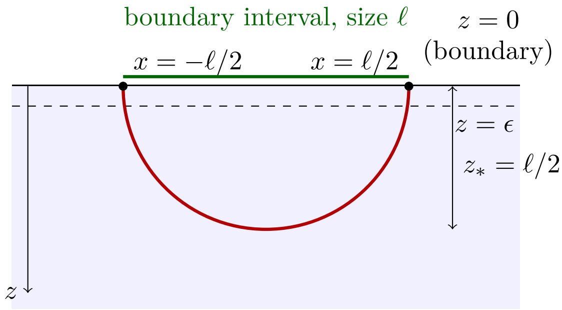


Figure 10.2: A spacelike geodesic in the Poincaré patch anchored at two boundary points: a semicircle of radius $\ell/2$ reaching down to $z_* = \ell/2$. Its length, cut off at $z = \epsilon$, is $2L\ln(\ell/\epsilon)$—the seed of the Ryu–Takayanagi formula of Chapter 21.

3. The length diverges at $z = 0$ because of the $1/z$ in the metric. Parametrize $x = \frac{\ell}{2}\cos\eta$, $z = \frac{\ell}{2}\sin\eta$, so that $\mathrm{d}x^2 + \mathrm{d}z^2 = \frac{\ell^2}{4}\mathrm{d}\eta^2$, and cut the curve off at $z = \epsilon$, i.e. at $\sin\eta_\epsilon = 2\epsilon/\ell$:

$$\text{Length} \;=\; L\int_{\eta_\epsilon}^{\pi-\eta_\epsilon} \frac{\mathrm{d}\eta}{\sin\eta} \;=\; L\left[\ln\tan\tfrac{\eta}{2}\right]_{\eta_\epsilon}^{\pi-\eta_\epsilon} \;=\; 2L\,\ln\cot\frac{\eta_\epsilon}{2}. \tag{10.20}$$

4. For $\epsilon \ll \ell$, $\cot(\eta_\epsilon/2) \simeq \ell/\epsilon$, so

$$\text{Length}(\ell) \;=\; 2L\,\ln\frac{\ell}{\epsilon} \;+\; O(\epsilon^2/\ell^2). \tag{10.21}$$

In Chapter 21, Length$/4G$ becomes the entanglement entropy of an interval, and the logarithm reproduces $\frac{c}{3}\ln(\ell/\epsilon)$ of two-dimensional CFT, with ϵ the field-theory cutoff.

10.6 Scalar fields in AdS and the mass–dimension relation

For a free scalar of mass m in the Poincaré patch, with $\sqrt{-g} = (L/z)^{d+1}$ and $g^{zz} = z^2/L^2$, $(\Box_g - m^2)\phi = 0$ reads

$$\frac{z^{d+1}}{L^2}\,\partial_z\!\left(z^{1-d}\,\partial_z\phi\right) + \frac{z^2}{L^2}\,\eta^{\mu\nu}\partial_\mu\partial_\nu\phi \;=\; m^2\phi. \tag{10.22}$$

1. *Indicial exponents.* Near $z = 0$ try $\phi \sim z^\Delta$; the x-derivative term carries an extra z^2 and is subleading, and $z^{d+1}\partial_z(z^{1-d}\Delta z^{\Delta-1}) = \Delta(\Delta-d)z^\Delta$:

$$\boxed{\Delta\,(\Delta - d) \;=\; m^2L^2 \qquad\Longrightarrow\qquad \Delta_\pm = \frac{d}{2}\pm\sqrt{\frac{d^2}{4}+m^2L^2}\,, \qquad \Delta_+ + \Delta_- = d.} \tag{10.23}$$

- *Near-boundary expansion.* Every solution behaves as

$$\phi(z,x) \;\simeq\; \alpha(x)\,z^{\Delta_-}(1+O(z^2)) \;+\; \beta(x)\,z^{\Delta_+}(1+O(z^2)). \tag{10.24}$$

Anticipating Chapter 12, a bulk scalar of mass m is dual to a primary of dimension Δ_+: the notation is not an accident.

- *Normalizability.* The Klein–Gordon norm involves $\int_0 \mathrm{d}z \sqrt{-g}\, |g^{tt}||\phi|^2 \sim \int_0 \mathrm{d}z\, z^{1-d}|\phi|^2$, which for $\phi \sim z^\Delta$ converges at $z = 0$ iff

$$\Delta \; > \; \frac{d-2}{2}. \tag{10.25}$$

Since $\Delta_+ \geq d/2$ the z^{Δ_+} mode is always normalizable, a dynamical excitation in the box; the z^{Δ_-} mode is generically non-normalizable, of infinite energy, and turning it on changes the asymptotics: it is a boundary condition. The condition (10.25) is the CFT unitarity bound of Chapter 9. In the dictionary of Chapter 12 the non-normalizable mode $\alpha(x)$ is the source of $\mathcal{O}$ and the normalizable mode $\beta(x)$ the expectation value $\langle\mathcal{O}\rangle$.

10.7 The Breitenlohner–Freedman bound

The exponents (10.23) become complex for $m^2L^2 < -d^2/4$. The precise statement [52]: AdS is stable against scalar perturbations iff $m^2L^2 \geq -d^2/4$, so a finite window of negative mass squared is harmless. Derivation by reduction to a Schrödinger problem:

1. For $\phi = \mathrm{e}^{-\mathrm{i}\omega t} f(z)$ at zero boundary momentum (nonzero $\vec{k}$ only adds a positive term), (10.22) is $z^2 f'' + (1-d) z f' + \omega^2 z^2 f = m^2 L^2 f$.

2. The substitution $f = z^{(d-1)/2}\psi$ removes the first derivative, $z^2 f'' + (1-d)zf' = z^{(d-1)/2}[z^2\psi'' - \frac{d^2-1}{4}\psi]$, the coefficient arising as $\frac{(d-1)(d-3)}{4} - \frac{(d-1)^2}{2}$, and

$$-\,\psi'' \;+\; \frac{\nu^2 - \frac{1}{4}}{z^2}\,\psi \;=\; \omega^2\psi, \qquad \nu^2 = m^2L^2 + \frac{d^2}{4}. \tag{10.26}$$

3. Stability is the attractive inverse-square potential c/z^2 on the half-line [53]:[2] for $c \geq -\frac{1}{4}$ the operator is bounded below and $\omega^2 \geq 0$; for $c < -\frac{1}{4}$ the particle falls to the centre, there are normalizable modes with $\omega^2 < 0$ for every $|\omega|$, growing like $\mathrm{e}^{|\omega| t}$, and the background is unstable. With $c = \nu^2 - \frac{1}{4}$ the condition is $\nu^2 \geq 0$:

Breitenlohner–Freedman bound. A scalar field on AdS_{d+1} is stable if and only if

$$m^2L^2 \;\geq\; -\,\frac{d^2}{4}, \tag{10.27}$$

i.e. precisely when the exponents $\Delta_\pm$ in (10.23) are real. Moderately tachyonic fields do not destabilize AdS: the curvature-induced gradient energy of any would-be decay mode, confined by the AdS box, overwhelms the negative potential energy.

- *Mechanism.* In flat space a tachyon condenses in arbitrarily soft modes at no gradient cost; in AdS every excitation carries gradient energy of order $1/L^2$, and only when $-m^2$ exceeds this cost does the instability win. Many Kaluza–Klein scalars of the compactifications of Chapter 16 have $m^2 < 0$ in the window and are dual to healthy operators with $\frac{d-2}{2} < \Delta < d$.

[2]With the boundary condition at $z = 0$ that selects a single one of the two powers $z^{1/2+\nu}$ or $z^{1/2-\nu}$, i.e. Δ_+ or Δ_-; these are the scale-invariant boundary conditions, and the bound below is the statement for them. For $0 < \nu < 1$ both powers are normalizable, and a boundary condition mixing them, which introduces a scale, admits exactly one $\omega^2 < 0$ mode for one sign of the mixing [53, §IV]; in AdS this is the analysis of [54], and in the dictionary of Chapter 12 a double-trace deformation of the wrong sign.

- *Two quantizations.* In the window

$$-\frac{d^2}{4} \le m^2L^2 \le -\frac{d^2}{4}+1 \qquad \Longleftrightarrow \qquad 0 \le \nu \le 1, \tag{10.28}$$

  both $z^{\Delta_\pm}$ are normalizable by (10.25), and one must choose which mode is boundary data and which is dynamical: two consistent quantizations of one field [52], dual to CFTs in which the operator has dimension Δ_+ or Δ_-, the window being exactly where both respect the unitarity bound (Chapter 12).

10.8 Symmetries and scales: a first dictionary

The kinematical skeleton of the correspondence, assembled before the correspondence is stated; the last two rows, the UV/IR relation, are developed in Section 12.5.

CFT in d dimensions	**Gravity on AdS_{d+1}**
conformal group $\mathrm{SO}(2,d)$	isometry group $\mathrm{SO}(2,d)$
radial quantization on $\mathbb{R}\times S^{d-1}$; conformal Hamiltonian $H_{\text{conf}} = \frac{1}{2}(P^0+K^0)$	global coordinates; translations of global time τ; boundary cylinder $\mathbb{R}\times S^{d-1}$
quantization on Minkowski space $\mathbb{R}^{1,d-1}$; P_μ, $J_{\mu\nu}$	Poincaré patch; isometries of the Minkowski slices
dilatation D	$(z,x)\to(\lambda z,\lambda x)$
scaling dimension Δ of a scalar primary	scalar field mass: $m^2L^2=\Delta(\Delta-d)$
unitarity bound $\Delta \ge \frac{d-2}{2}$	normalizability of the z^Δ mode; BF bound $m^2L^2 \ge -\frac{d^2}{4}$
UV regulator (short distances)	radial cutoff near the boundary, $z \ge \epsilon$
energy scale μ of a process	radial position: $\mu \sim 1/z$ (UV $\leftrightarrow z\to 0$, IR $\leftrightarrow z \to \infty$)

What is still missing is dynamics: which gravitational theory, which CFT, and why. The next chapter supplies the field-theory half of the answer, the large-N limit, and Chapter 12 states the correspondence.

Chapter 11

The Large-N Limit of Gauge Theories

Why should a quantum field theory have anything to do with strings or gravity in higher dimensions? The missing dynamical idea is 't Hooft's large-N limit [55]. Taking $N \to \infty$ at fixed $\lambda = g^2_{\mathrm{YM}} N$, double-line notation turns every Feynman diagram into a two-dimensional surface, and a connected diagram scales as N^{2-2g}: the free energy is a genus expansion dominated by planar diagrams, single-trace correlators factorize, and $1/N$ plays the role of $\hbar$. The genus expansion is formally the loop expansion of a string theory with $g_s \sim 1/N$, a string living, as we argue, in more than four dimensions, the extra dimension being an energy scale: the anti-de Sitter geometry of Chapter 10.

11.1 Why a large number of colours?

QCD, the SU(3) gauge theory of the strong interactions, has no small parameter at low energies. 't Hooft's proposal [55] was to invent one: replace the gauge group by SU(N) or U(N) and expand in $1/N$.

- *Systematic.* A well-defined leading term with computable corrections, valid at any coupling.
- *Physical.* At leading order mesons and glueballs are free and stable, with interactions suppressed by powers of $1/N$.
- *Simplifying.* Only the planar diagrams survive.
- *Stringy.* The $1/N$ expansion has the structure of a string perturbation series, the bridge over which AdS/CFT is built.
- *Our running example.* We mostly care about theories whose coupling does not run, above all $\mathcal{N} = 4$ super Yang–Mills (Chapter 16), for which N and $\lambda = g^2_{\mathrm{YM}} N$ are independent dimensionless parameters. Nothing below requires conformal invariance: the counting is combinatorial and applies to any gauge theory with adjoint matter.

11.2 The 't Hooft limit

Consider U(N) Yang–Mills with adjoint matter,

$$\mathcal{L} \;=\; -\frac{1}{2g^2_{\mathrm{YM}}}\,\mathrm{Tr}\,F_{\mu\nu}F^{\mu\nu} \;+\; \mathcal{L}_{\mathrm{matter}}, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + \mathrm{i}\,[A_\mu, A_\nu], \tag{11.1}$$

with A_μ and the matter fields $N\times N$ hermitian matrices, the trace in the fundamental, and the coupling as an overall prefactor also for the matter terms, so that every propagator carries $g_{\rm YM}^2$ and every vertex $1/g_{\rm YM}^2$.

- *The naive limit fails.* The one-loop gluon self-energy has two vertices, three propagators and one free colour index summed over N values, so it is proportional to $g_{\rm YM}^2 N$, which diverges as $N\to\infty$ at fixed $g_{\rm YM}$; so do infinitely many other diagrams.
- *The cure.* All factors of N come with factors of $g_{\rm YM}^2$, so take the two limits together:

The 't Hooft limit.

$$N \;\to\; \infty, \qquad g_{\rm YM} \;\to\; 0, \qquad \lambda \;\equiv\; g_{\rm YM}^2\, N \;\text{ fixed.} \tag{11.2}$$

The fixed combination λ is the *'t Hooft coupling.* In this limit the self-energy is finite and equal to λ; the same is true of the entire perturbative expansion, order by order in $1/N$.

- *Feynman rules.* Rescaling so that all coupling dependence sits in front of the action,

$$\begin{gathered}\mathcal{L} \;=\; \frac{N}{\lambda}\,\mathrm{Tr}\Big(-\tfrac12 F_{\mu\nu}F^{\mu\nu} - (D_\mu\phi)^2 - \cdots\Big),\\ \text{propagator} \propto \frac{\lambda}{N}, \qquad \text{vertex} \propto \frac{N}{\lambda}, \qquad \text{index loop} \propto N,\end{gathered} \tag{11.3}$$

so the only task is to count closed colour-index loops.

11.3 Double-line notation

- *Index lines.* An adjoint field is an $N\times N$ matrix $\phi^i{}_{\bar\jmath}$ with one fundamental and one anti-fundamental index, the adjoint being $\mathbf{N}\otimes\bar{\mathbf{N}}$ for $\mathrm{U}(N)$. Its propagator is diagonal in the matrix entries,

$$\langle \phi^i{}_{\bar\jmath}(x)\;\phi^k{}_{\bar l}(y)\rangle \;=\; \frac{\lambda}{N}\;\delta^i{}_{\bar l}\,\delta^k{}_{\bar\jmath}\;D(x-y), \tag{11.4}$$

and is drawn as a double line, two oppositely oriented index lines. A single-trace vertex such as $\mathrm{Tr}\,\phi^3 = \phi^i{}_{\bar\jmath}\phi^j{}_{\bar k}\phi^k{}_{\bar\imath}$ joins the index lines cyclically (Figure 11.1).
- *Index loops.* In a vacuum diagram every index line closes into a loop carrying a free colour index summed over N values, the factor N per loop in (11.3).
- *The self-energy again* (Figure 11.2). Four propagator segments, two vertices and one closed index loop give

$$\left(\frac{\lambda}{N}\right)^4\left(\frac{N}{\lambda}\right)^2 N \;=\; \lambda\times\frac{\lambda}{N}, \tag{11.5}$$

a correction of relative order λ to the tree propagator, finite in the 't Hooft limit.

11.4 Counting powers of N: the genus expansion

1. *The weight of a diagram.* A connected vacuum diagram with V vertices, E propagators and F closed index loops has weight

$$\left(\frac{N}{\lambda}\right)^V\left(\frac{\lambda}{N}\right)^E N^F \;=\; \lambda^{\,E-V}\,N^{\,V-E+F}. \tag{11.6}$$

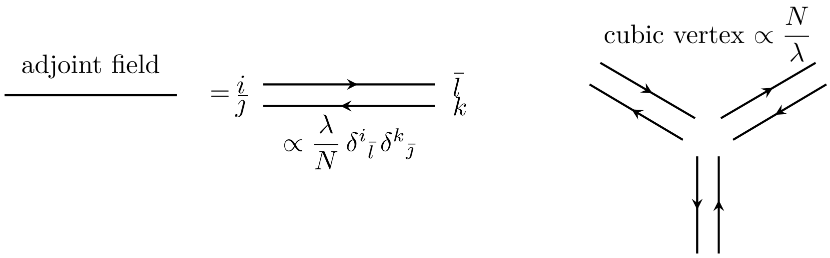


Figure 11.1: Double-line notation. An adjoint propagator is a pair of oppositely oriented index lines; a single-trace vertex ties the index lines together cyclically. Colour flows along the oriented single lines.

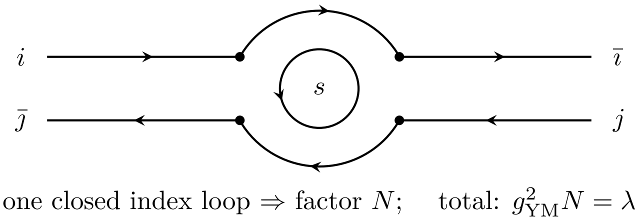


Figure 11.2: The one-loop gluon self-energy in double-line notation. The internal index line closes into a loop carrying the free colour index s, summed over N values. The correction is of order $g^2_{\mathrm{YM}}N = \lambda$, finite in the 't Hooft limit.

2. *Diagrams are surfaces.* Glue a disk onto each index loop: the loops become faces, the double-line propagators edges, the vertices points, and the result is a closed oriented surface tiled by the diagram. Euler's theorem gives

$$\chi \;=\; V - E + F \;=\; 2 - 2g, \tag{11.7}$$

with g the genus, the number of handles of the minimal surface on which the diagram can be drawn without crossings.

A connected vacuum diagram of a U(N) gauge theory with adjoint matter, drawn on a surface of minimal genus g, contributes

$$\lambda^{E-V}\, N^{2-2g}. \tag{11.8}$$

In the 't Hooft limit the diagrams are graded by their topology: the leading diagrams, of order N^2, are the *planar* diagrams—those that can be drawn on a plane (equivalently a sphere) without crossings. Each additional handle costs a factor $1/N^2$.

- *Topology, not loop order.* At every order in λ there are planar diagrams, so the coefficients of the $1/N$ expansion are full series in λ; within a genus, λ counts momentum loops, a connected diagram with $\ell = E - V + 1$ loops weighing $\lambda^{\ell-1}$.

- *Examples* (Figure 11.3). Two cubic vertices joined by three propagators: with the lines untwisted, $V = 2$, $E = 3$, $F = 3$, so $\chi = 2$ and the weight is λN^2; with one propagator twisted, the same spacetime integrand but all index lines merging into a single loop, $F = 1$, $\chi = 0$: a torus, weight λN^0. Crossings force handles and each handle costs N^{-2}.

- *The free energy.* Summing connected vacuum diagrams,

planar: $V = 2,\ E = 3,\ F = 3$
$\chi = 2$ (sphere): λN^2

non-planar: $V = 2,\ E = 3,\ F = 1$
$\chi = 0$ (torus): λN^0

Figure 11.3: Two vacuum diagrams with identical spacetime structure (two cubic vertices, three propagators) but different colour routing. Left: the planar "theta" diagram has three index loops and weight λN^2. Right: twisting the middle propagator merges all index lines into a single loop; the diagram can only be drawn without crossings on a torus, and its weight is λN^0.

Genus expansion of the free energy.

$$F \;=\; \sum_{g=0}^{\infty} N^{2-2g}\, f_g(\lambda) \;=\; N^2 f_0(\lambda) + f_1(\lambda) + \frac{1}{N^2} f_2(\lambda) + \cdots\,, \tag{11.9}$$

where $f_g(\lambda)$ resums *all* diagrams of genus g, to all orders in λ. As $N \to \infty$ only the planar sum $f_0(\lambda)$ survives.

The N^2 counts degrees of freedom: the action $\frac{N}{\lambda}\operatorname{Tr} F^2$ is of order N^2 on a generic configuration, since a trace is of order N. The same scaling reappears in Chapter 18, where the black hole dual to the deconfined phase has $F \sim N^2$ and the confined phase $F \sim O(1)$.

- *Two caveats.* Planarity reorganizes perturbation theory but does not resum it: $f_0(\lambda)$ contains infinitely many diagrams and is not known exactly in any four-dimensional gauge theory (integrability comes close for $\mathcal{N} = 4$ SYM). And for SU(N) the propagator acquires the term $-\frac{1}{N}\delta^i{}_{\bar{j}}\delta^k{}_{\bar{l}}$ removing the trace part, which is $1/N$-suppressed: SU(N) and U(N) share the planar limit.

- *Fundamental matter.* A quark q_i carries a single index and its loop is an index line without partner, a boundary of the surface; with b boundaries $\chi = 2 - 2g - b$, so each quark loop costs $1/N$ and the expansion proceeds in $1/N$ rather than $1/N^2$. Closed surfaces are worldsheets of closed strings and surfaces with boundaries of open strings ending on D-branes (Chapter 15); $\mathcal{N} = 4$ SYM has only adjoint matter, so closed strings suffice.

11.5 Single-trace operators and large-N factorization

- *Single-trace operators.* The gauge-invariant observables are traces of products of adjoint fields at a point,

$$\mathcal{O}(x) \;=\; \frac{1}{N}\operatorname{Tr}\left(\phi_1\phi_2\cdots\phi_k\right)(x), \tag{11.10}$$

normalized to be of order one on a generic configuration; multi-trace operators are their products. A source term $N\int J\mathcal{O}$ enters the action like the 't Hooft Lagrangian, N times a trace, so $F(J) = \sum_g N^{2-2g} f_g(\lambda; J)$ keeps the genus expansion and

$$\langle \mathcal{O}_1(x_1) \cdots \mathcal{O}_n(x_n) \rangle_{\mathrm{c}} \;=\; \frac{1}{N^n} \frac{\delta^n F}{\delta J_1(x_1) \cdots \delta J_n(x_n)} \;=\; O(N^{2-n}), \tag{11.11}$$

the derivatives of F being of order N^2 and the $1/N^n$ coming from the normalization.

Connected correlation functions of unit-normalized single-trace operators scale as

$$\langle \mathcal{O} \rangle \sim N, \qquad \langle \mathcal{O}_1 \mathcal{O}_2 \rangle_{\mathrm{c}} \sim 1, \qquad \langle \mathcal{O}_1 \mathcal{O}_2 \mathcal{O}_3 \rangle_{\mathrm{c}} \sim \frac{1}{N}, \qquad \langle \mathcal{O}_1 \cdots \mathcal{O}_n \rangle_{\mathrm{c}} \sim N^{2-n}. \tag{11.12}$$

Every additional operator insertion costs a factor $1/N$: single-trace operators become weakly interacting at large N, with effective coupling $1/N$.

With operators normalized to unit two-point function, as in Chapter 9, the pattern is the same from the two-point level up: three-point couplings are $O(1/N)$, n-point $O(N^{2-n})$.

- *Factorization.* For non-vanishing one-point functions,

$$\langle \mathcal{O}_1 \, \mathcal{O}_2 \rangle \;=\; \langle \mathcal{O}_1 \rangle \langle \mathcal{O}_2 \rangle \;+\; \langle \mathcal{O}_1 \mathcal{O}_2 \rangle_{\mathrm{c}} \;=\; \langle \mathcal{O}_1 \rangle \langle \mathcal{O}_2 \rangle \left(1 + O\left(\frac{1}{N^2}\right)\right), \tag{11.13}$$

  the disconnected term being $O(N^2)$ against the connected $O(1)$: gauge-invariant observables do not fluctuate, their relative variance being $O(1/N^2)$.

- *A classical limit.* Absence of fluctuations defines a classical limit: with action $S/\hbar$ expectation values factorize up to $O(\hbar)$, and here the action is N^2 times an N-independent functional.

The large-N limit is a classical limit, with

$$\frac{1}{N^2} \;\longleftrightarrow\; \hbar. \tag{11.14}$$

It is *not* the classical limit of the gauge theory itself (f_0 contains arbitrarily many powers of λ, i.e. arbitrarily many gauge theory loops); it is the classical limit of some *other* description, whose loop expansion parameter is $1/N^2$ and whose classical solution encodes the full planar sum.

This hypothetical description, a single configuration whose classical observables reproduce all planar correlators, is the *master field.* AdS/CFT realizes it: for $\mathcal{N} = 4$ SYM at large λ the master field is a classical gravitational configuration, anti-de Sitter space, and the $1/N^2$ fluctuations are those of gravity.

11.6 The string interpretation

- *String perturbation theory is a sum over topologies.* A string sweeps out a two-dimensional worldsheet, and closed-string amplitudes are sums over closed oriented surfaces weighted by the string coupling (Chapter 15),

$$\mathcal{A} \;=\; \sum_{g=0}^{\infty} g_s^{2g-2} \mathcal{A}_g \;=\; \frac{1}{g_s^2} \mathcal{A}_{\text{sphere}} + \mathcal{A}_{\text{torus}} + g_s^2 \, \mathcal{A}_{2-\text{handles}} + \cdots, \tag{11.15}$$

  each handle a string loop costing g_s^2. Compared with (11.9) the two series are formally identical:

$$g_s \sim \frac{1}{N}, \tag{11.16}$$

(more precisely $g_s = c(\lambda)/N$, with the λ-dependence fixed by the specific realization), and with $f_g(\lambda)$ playing the role of the genus-g string amplitude. *The $1/N$ expansion of a gauge theory is the loop expansion of a string theory*: planar diagrams are tree-level closed-string amplitudes (spheres), and each $1/N^2$ correction is a string loop (a handle) [55].

- *Tilings.* The faces of a double-line diagram tile the surface it is drawn on, so summing planar diagrams sums over tilings of the sphere, a discretization of the sum over sphere geometries defining a tree-level string amplitude. At small λ few vertices dominate and the tiling is coarse; at large λ diagrams with enormous numbers of vertices dominate and the sum is plausibly governed by a smooth worldsheet:

$$\begin{aligned} &\lambda \ll 1: \text{ gauge theory perturbative, string picture rough;} \\ &\lambda \gg 1: \text{ gauge theory strongly coupled, string picture smooth.} \end{aligned} \tag{11.17}$$

  This is the seed of the strong/weak nature of AdS/CFT, of its power and of the difficulty of testing it.

- *Which string, and where?* The natural guess, a string in the four dimensions of the gauge theory, fails twice: a string in flat D dimensions is consistent only for $D = 26$ (bosonic) or $D = 10$ (superstring), the worldsheet conformal symmetry being anomalous otherwise; and the closed-string spectrum always contains a massless spin-two particle, a graviton, which a hadronic string cannot have and a gravitational interpretation welcomes.

- *The extra dimension is the scale.* Polyakov's lesson from the first failure: in a non-critical dimension the conformal mode of the worldsheet metric, the Liouville field, becomes a dynamical field on the worldsheet, effectively an extra target coordinate. A four-dimensional string moves in five dimensions, and since the string describes the gauge theory at all scales, the fifth coordinate is the renormalization group scale.

- *Symmetry fixes the geometry.* For a conformal gauge theory, take x^μ and one coordinate z with the meaning of inverse energy. d-dimensional Poincaré invariance forces $\mathrm{d}s^2 = \Omega^2(z)(\mathrm{d}z^2 + \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu)$, and scale invariance, under which a length transforms as $z \to \lambda z$, forces $\Omega = L/z$:

$$\mathrm{d}s^2 = L^2 \frac{\mathrm{d}z^2 + \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu}{z^2}, \tag{11.18}$$

  AdS_5 in the Poincaré coordinates (10.11). If the planar limit of a four-dimensional conformal gauge theory is a classical string background, that background contains anti-de Sitter space.

Summary of the argument, and a preview. The $1/N$ expansion says a large-N gauge theory *is* a string theory with $g_s \sim 1/N$; consistency of string theory says the string lives in higher dimensions, with the extra dimension acting as energy scale; the symmetries of a CFT_4 then fix the five-dimensional geometry to be AdS_5. What large-N counting alone cannot supply is the *identity* of the string theory, the value of L, and a derivation. For $\mathcal{N} = 4$ super Yang–Mills all three will be provided in Chapter 16 by an honest string-theory construction (D3-branes), yielding Type IIB string theory on $AdS_5 \times S^5$ with $g_s \sim g_{\mathrm{YM}}^2$ and $L^4/\alpha'^2 = \lambda$. The next chapter states this correspondence in general terms and begins to develop its consequences.

Chapter 12

The AdS/CFT Correspondence

This is the central chapter of the course. To the evidence accumulated so far, the shared symmetry group $\mathrm{SO}(2,d)$ of Chapters 8 and 10 and the stringy $1/N$ expansion of Chapter 11, we add the holographic principle from Part I and state the conjecture: a quantum gravity (string) theory on $\mathrm{AdS}_{d+1} \times$ (compact) *equals* a conformal field theory in d dimensions. We distinguish its strong and weak forms, quote the parameter map of the canonical example (derived in Chapter 16), and develop the working dictionary: bulk fields $\leftrightarrow$ boundary operators, boundary values as sources, the GKPW formula, masses $\leftrightarrow$ dimensions, normalizable modes $\leftrightarrow$ states, radial position $\leftrightarrow$ energy scale. We close by asking which field theories can have a semiclassical gravity dual, and what the duality does and does not promise.

12.1 The holographic principle

- *Two puzzles.* The match between Chapters 8 and 10 is structural, down to the generators J_{MN} and the conformal class of the boundary metric, and Chapter 11 added that a conformal large-N gauge theory is a string theory whose target space contains AdS_5. Two puzzles remain. Many inequivalent theories share a symmetry group. And how could a theory *with* gravity in $d+1$ dimensions carry the same information as a theory *without* gravity in d dimensions, when a local field theory has independent degrees of freedom at every point and $(d+1)$-dimensional space has more points? The second puzzle is answered by Part I.

- *Black hole entropy bounds the entropy of everything.* Run the following thought experiment, due in this form to 't Hooft [56] and Susskind [57]. Take a spherical region of area A containing a gravitating system of entropy S and energy E. If the energy inside a radius R exceeds the Schwarzschild bound $E \lesssim R/2G$, the region is already inside a black hole; so add energy, gently, until the system collapses into a black hole that just fills the region. The collapse is an ordinary process, so by the Generalized Second Law of Chapter 2 the entropy cannot decrease,

$$S \;\leq\; S_{\text{final}} \;=\; S_{\text{BH}} \;=\; \frac{A}{4G}, \tag{12.1}$$

and the energy added only *increased* the left-hand side, so the original system obeys the bound a fortiori.

Holographic entropy bound. The entropy of any system contained in a region bounded by a surface of area A satisfies

$$S \le \frac{A}{4G}: \tag{12.2}$$

the maximal entropy, the logarithm of the number of independent states, of a gravitating region scales with the *area* of its boundary, not with its volume.

- *Why this is shocking.* In a local QFT with cutoff ϵ a region of volume V has of order V/ϵ^{d-1} degrees of freedom and a maximal entropy scaling with the volume, which for large regions exceeds the area in cutoff units. Almost all of these states are not available: long before reaching them the system collapses into a black hole, whose states are counted by area. In a counting sense the degrees of freedom of quantum gravity in a region are those of an ordinary quantum system on its *boundary*, one per Planck-area cell: the *holographic principle* [56, 57].

- *Why anti-de Sitter space.* The principle does not say *which* boundary theory describes a given bulk, nor how the dictionary works. AdS is where it becomes precise, for two reasons met in Chapter 10: light reaches the conformal boundary in finite time and boundary data are needed to evolve the bulk, so the boundary is part of the dynamical specification of the theory; and the boundary carries exactly the structure, a conformal class of metrics with symmetry $\mathrm{SO}(2, d)$, on which a CFT_d lives. The boldest reading of the symmetry match is then:

The AdS/CFT correspondence [58, 59, 60]. A theory of quantum gravity (in all known realizations, a string theory) on a spacetime asymptotic to $\mathrm{AdS}_{d+1} \times \mathcal{M}$, with $\mathcal{M}$ a compact space, is *exactly equivalent*, as a quantum theory, state by state and observable by observable, to a conformal field theory in d dimensions, which can be thought of as living on the conformal boundary of AdS_{d+1}. The CFT is not an approximation, a limit, or a boundary condition of the gravity theory: it is a complete, independent, and non-perturbative *definition* of it.

- *The compact factor, and states.* $\mathcal{M}$ (a five-sphere in the canonical example) contributes, upon Kaluza–Klein reduction (Chapter 17), a tower of fields on AdS_{d+1}, and its isometries appear in the CFT as global internal symmetries. Being an equivalence of theories, the correspondence does not require the bulk to be exactly AdS: any configuration with the right asymptotics, a black hole in AdS for instance (Chapter 18), is a state of the *same* CFT.

- *The canonical example*, proposed by Maldacena [58] and derived in Chapter 16 from a stack of N D3-branes, is

$$\mathcal{N} = 4 \text{ supersymmetric } \mathrm{SU}(N) \text{ Yang–Mills theory in } d = 4 \iff \text{Type IIB string theory on } \mathrm{AdS}_5 \times S^5,$$

with the string coupling and the AdS radius fixed by the gauge theory parameters,

$$g_s \sim g_{\mathrm{YM}}^2, \qquad \frac{L^4}{\alpha'^2} = \lambda = g_{\mathrm{YM}}^2 N, \tag{12.3}$$

where $\alpha' = \ell_s^2$ is the square of the string length (Chapter 15) and factors of 4π are suppressed until Chapter 16. Both relations are what Chapter 11 anticipated: $g_s \sim 1/N$ at fixed λ (here $g_s \sim \lambda/N$), and a smooth geometry on the string scale, $L \gg \ell_s$, precisely at strong 't Hooft coupling. For most of this chapter we keep any d and any consistent bulk theory, since the dictionary uses only the structure of AdS asymptotics.

12.2 Strong and weak forms of the conjecture

The correspondence as stated is the *strong form*, an exact equivalence for all N and λ. Since string theory has no independent non-perturbative definition, the strong form is best regarded as a definition whose self-consistency is tested by its semiclassical limits, and in practice one works with weaker forms obtained by expanding around corners of parameter space, translated with (12.3):

form	**gauge theory**	**bulk theory**
strong	any N, any λ	full quantum Type IIB string theory
intermediate	$N \to \infty$, any λ	classical (tree-level, $g_s \to 0$) string theory on $\mathrm{AdS}_5 \times S^5$; worldsheet α' corrections retained
weak	$N \to \infty$, then $\lambda \to \infty$	classical Type IIB *supergravity* on $\mathrm{AdS}_5 \times S^5$

The weak form is the one we compute with. Classical gravity describes the bulk when two independent conditions hold:

- *Small string-loop corrections*: $g_s \ll 1$, i.e. at fixed λ, $N \gg 1$; quantum gravity corrections are controlled by $g_s^2 \sim 1/N^2$, the genus expansion of Chapter 11.

- *Small worldsheet corrections*: the geometry must be weakly curved on the string scale, $L \gg \ell_s$, i.e. $\lambda = L^4/\alpha'^2 \gg 1$; stringy corrections to supergravity are controlled by powers of $\alpha'/L^2 = \lambda^{-1/2}$. (Comparing with the Planck length instead, $\ell_P^8 \sim G_{10} \sim g_s^2\alpha'^4$, so $L \gg \ell_P$ requires $N \gg 1$ directly.)

> Classical gravity in the bulk is reliable precisely when the gauge theory is in its most inaccessible regime, $N \to \infty$ *and* $\lambda \gg 1$; when the gauge theory is weakly coupled the dual geometry is curved on the string scale. AdS/CFT is a *strong/weak-coupling duality*: hard on one side means easy on the other. This is why it is powerful (strongly coupled gauge theory from classical gravity, quantum gravity defined by an ordinary QFT) and why it is hard to test naively (no observable is easily computable on both sides at once; Chapter 14 discusses the protected quantities that evade this).

12.3 The field–operator map and the GKPW formula

- *What is to be matched.* On the CFT side the observables are the correlation functions of gauge-invariant operators, which by the state–operator map of Chapter 9 carry complete information about the theory. On the gravity side the natural objects are the bulk fields: the metric g_{MN}, gauge fields A_M, scalars ϕ, and whatever else the bulk theory contains. The dictionary rests on one fact established in Chapter 10: a bulk field is not determined by its dynamics but needs *boundary data*, one function of the d boundary coordinates per field, the coefficient $\alpha(x)$ of the non-normalizable mode. A function of d coordinates is exactly what a CFT can supply.

- *Sources.* In quantum field theory an operator $\mathcal{O}(x)$ is probed by coupling it to an external source $h(x)$,

$$\mathcal{L}_{\mathrm{CFT}} \;\longrightarrow\; \mathcal{L}_{\mathrm{CFT}} + \int \mathrm{d}^d x\, h(x)\, \mathcal{O}(x), \tag{12.4}$$

which defines the generating functional of connected correlators,

$$\mathrm{e}^{W[h]} = \left\langle \mathrm{e}^{\int \mathrm{d}^d x\, h\, \mathcal{O}} \right\rangle_{\mathrm{CFT}}, \qquad \langle \mathcal{O}(x_1)\cdots\mathcal{O}(x_n)\rangle_{\mathrm{c}} = \left.\frac{\delta^n W}{\delta h(x_1)\cdots\delta h(x_n)}\right|_{h=0}. \tag{12.5}$$

We work in Euclidean signature (Lorentzian remarks at the end of Chapter 13). If $\mathcal{O}$ has dimension Δ and spin s, the source h has dimension $d-\Delta$ and the conjugate spin indices.

Every bulk field ϕ corresponds to a single-trace primary operator $\mathcal{O}$ of the boundary CFT with the same $\mathrm{SO}(2,d)$ (and internal) quantum numbers, and the boundary value of ϕ is identified with the source h of $\mathcal{O}$. Schematically,

$$\phi(z,x) \xrightarrow{z\to 0} h(x) \qquad \Longleftrightarrow \qquad \mathcal{L}_{\mathrm{CFT}} + \int \mathrm{d}^d x\, h\, \mathcal{O}. \tag{12.6}$$

(Bulk solutions do not literally approach a constant at the boundary: the precise statement is $\phi \sim z^{d-\Delta} h(x)$ as $z \to 0$, Section 12.4.)

- *Which operator pairs with which field?* In explicit realizations the map is fixed by the microscopic construction (Chapters 16–17), but three universal entries follow from the linearized couplings of any QFT to background fields,

$$\int \mathrm{d}^d x \sqrt{g}\, \Big(\tfrac{1}{2}\, \delta g_{\mu\nu}\, T^{\mu\nu} + A_\mu\, J^\mu + \phi\, \tfrac{1}{N}\, \mathrm{Tr}\, F_{\mu\nu} F^{\mu\nu} + \cdots \Big). \tag{12.7}$$

  The *graviton*, present in every gravitational bulk theory, couples to the *stress tensor*, present in every local CFT. A bulk *gauge field* A_M couples to a *conserved current*: $W[A_\mu]$ depends on the boundary gauge field only up to gauge transformations, which is current conservation, $\partial_\mu J^\mu = 0$. This is the general rule that *gauge symmetries in the bulk are global symmetries of the boundary theory*; the isometries of $\mathcal{M}$ yield bulk gauge fields whose currents generate the CFT's internal symmetry (the R-symmetry, in the canonical example). A massless bulk *scalar* (the dilaton, in string realizations) couples to $\mathrm{Tr}\, F^2$, of dimension exactly $d = 4$ in a conformal gauge theory.

- *The GKPW formula* (Gubser, Klebanov, Polyakov [59] and Witten [60]). The CFT generating functional depends on an arbitrary d-dimensional configuration $h(x)$, and the bulk theory has a natural functional of the same data: the partition function with the boundary condition that the field dual to $\mathcal{O}$ approaches h at the conformal boundary,

$$\left\langle \mathrm{e}^{\int \mathrm{d}^d x\, h\, \mathcal{O}} \right\rangle_{\mathrm{CFT}} = Z_{\mathrm{grav}}\big[\, \phi \xrightarrow{z\to 0} h \,\big] \equiv \int_{\phi\,\to\, h \text{ at } \partial\mathrm{AdS}} \mathcal{D}\phi\ \mathrm{e}^{-S_{\mathrm{bulk}}[\phi]}. \tag{12.8}$$

- *Reading the formula.* The left-hand side is a CFT object, a functional of the arbitrary source; the right-hand side is the quantum-gravity path integral over all bulk configurations with prescribed asymptotics (written for one scalar; in general all bulk fields, gravity included, each with its own boundary data, summed over topologies). Since $W[h]$ for all sources determines the CFT completely, (12.8) *is* the equivalence of the two theories, stated as an equality of generating functionals.

- *Semiclassical limit.* In the weak form the bulk path integral is dominated by its saddle point, the classical solution $\hat{\phi}(z,x)$ with the prescribed boundary behaviour,

$$Z_{\mathrm{grav}}[\phi \to h] \simeq \mathrm{e}^{-S_{\mathrm{bulk}}[\hat{\phi}]}, \qquad \text{i.e.} \qquad W[h] \simeq -S_{\mathrm{bulk}}[\hat{\phi}\,]. \tag{12.9}$$

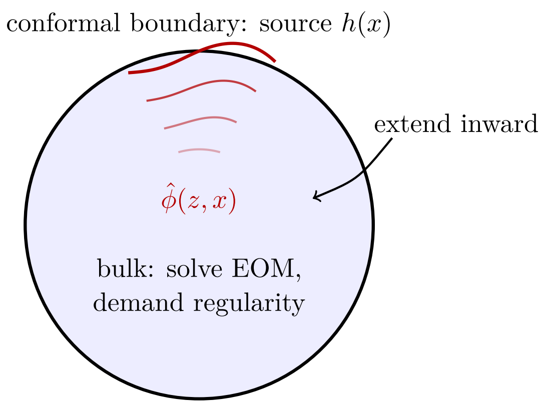


Figure 12.1: The GKPW prescription in Euclidean AdS (drawn as the ball model of Chapter 10). The source $h(x)$, an arbitrary function on the boundary sphere, is extended into the bulk as the unique regular solution $\hat{\phi}$ of the field equations with asymptotics $\hat{\phi} \sim z^{d-\Delta}h(x)$; the on-shell action of $\hat{\phi}$ generates the connected correlators of the dual operator $\mathcal{O}$.

In the semiclassical regime, the generating functional of connected CFT correlators is minus the *on-shell* bulk action, regarded as a functional of the boundary data. Correlation functions follow by functional differentiation with respect to the sources. An *off-shell* problem in d dimensions is solved by an *on-shell* problem in $d+1$ dimensions, the hallmark of all holographic computations.

- *A Dirichlet problem* (Figure 12.1). As in electrostatics, the boundary data are extended into the bulk by solving the field equations, and the "energy" of the resulting configuration is the generating object. The equations being second order, uniqueness needs a second condition in the interior: in Euclidean signature this is smoothness at the centre of the ball of Chapter 10; in Lorentzian signature one must choose conditions at the Poincaré horizon, which is where states and $i\epsilon$ prescriptions enter (Chapter 13).

- *Boundary representative.* In global coordinates (12.8) computes correlators on the cylinder, in Poincaré coordinates on flat space; a bulk coordinate change acts on the boundary as a conformal transformation, so the conformal representative is matched by a bulk diffeomorphism: bulk gauge redundancy, boundary physics.

12.4 Masses and dimensions; sources and states

Chapter 10 found for a free scalar the two-parameter near-boundary asymptotics

$$\phi(z,x) \simeq \alpha(x)\, z^{d-\Delta}\,(1+O(z^2)) + \beta(x)\, z^{\Delta}\,(1+O(z^2)), \qquad \Delta = \frac{d}{2} + \sqrt{\frac{d^2}{4} + m^2L^2}\,, \tag{12.10}$$

where Δ is now the larger root Δ_+ of $\Delta(\Delta-d) = m^2L^2$ and $\Delta_- = d-\Delta$. The dictionary interprets every ingredient.

- *The exponent is a scaling dimension.* Under the bulk dilatation $(z,x) \to (\lambda z, \lambda x)$ the coefficient α transforms as a field of dimension $d-\Delta$ and β as one of dimension Δ. Since the source of an operator of dimension Δ has dimension $d-\Delta$ by (12.4), we identify $\alpha(x) = h(x)$ and Δ with the conformal dimension of $\mathcal{O}$; the boundary condition in (12.8) reads, precisely,

$$\phi(z,x) \xrightarrow{z\to 0} z^{d-\Delta}\, h(x). \tag{12.11}$$

This is why "$\phi \to h$" cannot be taken literally: the solution vanishes at the boundary for $\Delta < d$, diverges for $\Delta > d$, and approaches a constant only for $\Delta = d$. The metric, which blows up as z^{-2}, has as "boundary value" the boundary conformal metric, source of the dimension-d operator $T_{\mu\nu}$.

- *The mode α is a source.* It is non-normalizable (Chapter 10): turning it on changes the asymptotics, i.e. the theory, just as a source $h\,\mathcal{O}$ changes the CFT action. A constant source is a change of couplings.

- *The mode β is a state.* It is normalizable and labels a state of the theory in the given background; its boundary interpretation is the *expectation value*, as Chapter 13 derives:

$$\langle \mathcal{O}(x) \rangle \;=\; (2\Delta - d)\,\beta(x) \qquad \text{(up to normalization conventions).} \tag{12.12}$$

bulk field ϕ, mass m	**boundary operator $\mathcal{O}$, dimension Δ**
$m^2L^2 = \Delta(\Delta - d)$ (scalar)	conformal dimension Δ
non-normalizable mode $\alpha(x)\,z^{d-\Delta}$	source / coupling $h(x)$ for $\mathcal{O}$
normalizable mode $\beta(x)\,z^{\Delta}$	state; $\langle\mathcal{O}\rangle \propto (2\Delta - d)\,\beta$
regular solution with $\alpha \neq 0$	vacuum correlators with source $h = \alpha$
solution with $\beta \neq 0$, $\alpha = 0$	excited state / VEV of $\mathcal{O}$ (spontaneous)

- *Fields of higher spin.* The same near-boundary analysis gives for each spin its own mass–dimension relation, always the quadratic Casimir of $\mathrm{SO}(2, d)$ in two languages. For $d = 4$ [2]:

bulk field on AdS_5	**mass–dimension relation**	**dual operator**
scalar ϕ	$m^2L^2 = \Delta(\Delta - 4)$	scalar $\mathcal{O}_\Delta$
vector A_M	$m^2L^2 = (\Delta - 1)(\Delta - 3)$	vector operator
massless vector	$\Delta = 3$	conserved current J_μ
antisymmetric B_{MN}	$m^2L^2 = (\Delta - 2)^2$	2-form operator
spin-$\frac{1}{2}$ ψ	$\lvert m\rvert L = \Delta - 2$	fermionic operator
graviton g_{MN} (massless)	$\Delta = 4$	stress tensor $T_{\mu\nu}$

A *massless* vector has $\Delta = 3 = d - 1$, the dimension at which a vector operator saturates the unitarity bound of Chapter 9 and is conserved: bulk gauge invariance and boundary current conservation are the same statement. Likewise the massless graviton maps to $\Delta = 4 = d$, the conserved stress tensor; a CFT whose stress tensor had an anomalous dimension would need a massive graviton. "Mass" in AdS is convention-laden (curvature couplings can be moved in and out of m^2); the invariant statement is the $\mathrm{SO}(2, d)$ representation, i.e. Δ.

- *Relevant, marginal, irrelevant; tachyons welcome.* A source $h\,\mathcal{O}$ is a coupling of dimension $d - \Delta$, so the deformation is relevant for $\Delta < d$, marginal for $\Delta = d$, irrelevant for $\Delta > d$:

$$\underbrace{-\frac{d^2}{4} \le m^2L^2 < 0,}_{\text{relevant } (\frac{d}{2}\le\Delta<d)} \qquad \underbrace{m^2 = 0}_{\text{marginal } (\Delta=d)}\;, \qquad \underbrace{m^2 > 0}_{\text{irrelevant } (\Delta>d)}\;. \tag{12.13}$$

Relevant operators, the most interesting deformations of a CFT, are dual to fields tachyonic in the flat-space sense, harmless down to the Breitenlohner–Freedman bound

$m^2L^2 \geq -d^2/4$ of Chapter 10, exactly where Δ would become complex. In the window $-d^2/4 \leq m^2L^2 \leq -d^2/4+1$ both modes in (12.10) are normalizable and one may *choose* which is the source, the two quantizations of Chapter 10, defining two CFTs in which $\mathcal{O}$ has dimension Δ_+ or Δ_-, the latter covering the unitary range $\frac{d-2}{2} \leq \Delta < \frac{d}{2}$. On $\mathrm{AdS}_5 \times S^5$ the lightest Kaluza–Klein scalar has $m^2L^2 = -4$, saturating the bound, dual to the $\Delta = 2$ operator $\mathrm{Tr}\,\phi^{(i}\phi^{j)}$ (Chapter 17).

12.5 The UV/IR connection

The identification of the radial coordinate with energy scale, previewed in Chapter 10, is central to how the correspondence is used.

- *Dilatations move the radius.* Under the boundary dilatation $x \to \lambda x$ the bulk isometry is $(z, x) \to (\lambda z, \lambda x)$: a CFT process at energy μ maps to one at μ/λ, and its bulk representative moves from depth z to λz,

$$z \sim \frac{1}{\mu}, \qquad z \to 0 \leftrightarrow \mathrm{UV}, \qquad z \to \infty \leftrightarrow \mathrm{IR}. \tag{12.14}$$

- *Warping does the same.* An object of proper size ℓ_{proper} at depth z occupies boundary size $\ell_{\mathrm{coord}} = (z/L)\,\ell_{\mathrm{proper}}$, since a constant-$z$ slice has metric $(L/z)^2 \mathrm{d}x^2$: the same bulk object, placed deeper, is a fatter, lower-energy boundary excitation. Holographic computations of a correlator at separation $|x|$ are dominated by the region $z \sim |x|$, as the geodesic of Chapter 10 ($z_* = \ell/2$) exemplified and the propagators of Chapter 13 confirm.

- *The radial cutoff is a UV regulator.* Quantities in AdS diverge near the boundary (infinite volume, divergent on-shell actions); quantities in CFT diverge in the UV. Cutting off the bulk at $z = \epsilon$ regularizes the CFT with a short-distance cutoff ϵ, and removing it with counterterms, *holographic renormalization* (Chapter 13), is the bulk image of QFT renormalization.

- *Counting degrees of freedom.* This closes the circle with the holographic principle. $\mathcal{N} = 4$ SYM on a spatial volume V_3 with cutoff ϵ has $N_{\mathrm{dof}} \sim N^2 V_3/\epsilon^3$ degrees of freedom (N^2 matrix components per cutoff cell). On the gravity side the holographic bound applied to the region $z \geq \epsilon$ assigns its boundary surface an entropy capacity $A/4G_5$ with

$$A = \left(\frac{L}{\epsilon}\right)^3 V_3 \qquad \Longrightarrow \qquad \frac{A}{4G_5} \sim \frac{L^3}{G_5}\frac{V_3}{\epsilon^3} \sim N^2\,\frac{V_3}{\epsilon^3}, \tag{12.15}$$

using $L^3/G_5 \sim N^2$, which follows from (12.3): $G_{10} \sim g_s^2 \alpha'^4$, reduction on the S^5 gives $G_5 \sim G_{10}/L^5$, so $L^3/G_5 \sim L^8/(g_s^2\alpha'^4) \sim \lambda^2/g_s^2 \sim N^2$. The area of the holographic screen in Planck units counts exactly the degrees of freedom of the cutoff gauge theory, N-dependence included: AdS/CFT does not merely accommodate the holographic principle, it saturates it.

12.6 Which theories have a gravity dual, and what the duality promises

Given *any* effective gravitational theory with an AdS vacuum, the GKPW prescription generates "correlators" with the formal properties of a CFT (conformal covariance and consistency with the operator product expansion are verified in Chapter 13). Conversely, every CFT plausibly has some bulk description, but generically a *string* theory with strings as large as the AdS radius, no better than the CFT itself. The useful question is which CFTs have a dual well approximated by semiclassical Einstein gravity coupled to a few light fields; the structure of the correspondence dictates the requirements [2].

- *Large N.* The bulk loop expansion parameter is $G/L^{d-1} \sim 1/N^2$ by (12.15), so classical gravity requires $N \gg 1$: many degrees of freedom per point and factorized correlators, as in Chapter 11.

- *Strong coupling, i.e. a gap in the spectrum.* Einstein gravity plus a few fields has few light excitations, all of spin ≤ 2. By the mass–dimension relation the dual CFT must have a *sparse* set of single-trace operators of low dimension and spin ≤ 2, with the duals of massive string states pushed to $\Delta \sim mL \sim L/\ell_s \sim \lambda^{1/4}$. A free CFT fails: it has infinite towers of conserved currents of every spin ($\mathrm{Tr}\,\phi\,\partial^s\phi + \cdots$, $\Delta = s+2$), requiring infinitely many massless higher-spin bulk fields, a stringy rather than gravitational regime. Strong 't Hooft coupling lifts these operators: the gap *is* $\lambda \gg 1$ seen from the boundary.

- *Fine print.* Einstein gravity imposes sharper constraints: in $d = 4$ the two central charges must coincide at leading order, $a = c$ (Chapter 14 computes both from gravity); free theories, with $a \neq c$, fail again.

Semiclassical gravity duals belong to CFTs that are (i) large-N (many degrees of freedom, factorized correlators) and (ii) strongly coupled, with a sparse spectrum of low-dimension single-trace operators of spin ≤ 2 and a parametrically large gap to everything else. Theories failing these criteria have duals that are irreducibly stringy or strongly quantum. $\mathcal{N} = 4$ SYM at large N and large λ passes all tests; free or weakly coupled theories fail them all.

- *What the correspondence delivers.* A non-perturbative *definition* of quantum gravity in asymptotically AdS spacetimes, encoded in a quantum field theory with manifest unitarity, the tool Part III brings back to the puzzles of Part I; and a computational window on *strongly coupled* field theory, where classical gravity yields correlators, anomalies, transport coefficients and entanglement entropies at $\lambda \to \infty$ (Chapters 13–14).

- *What it does not deliver.* Quantum gravity in flat or cosmological spacetimes, the boundary being a crutch we do not know how to remove; QCD itself, of which $\mathcal{N} = 4$ SYM at large N and strong coupling is a cousin, so holographic results on confinement or the quark-gluon plasma are analogies; and a proof: the strongest form remains a conjecture, spectacularly well tested (spectrum matching in Chapter 17, anomalies in Chapter 14, localization and integrability, black hole entropy counting) with no counterexample known.

- *Where we go next.* Chapter 13 makes (12.8) compute, extracting two- and three-point functions from the on-shell action and confronting the divergences at the AdS boundary with holographic renormalization. Chapters 15–16 then pay the outstanding debt: what string theory and D3-branes are, and why $\mathcal{N} = 4$ SYM and $\mathrm{AdS}_5 \times S^5$ are the same theory.

Chapter 13

Correlation Functions from Gravity

This chapter turns the GKPW formula into a working computational scheme. We solve the free massive scalar in Euclidean AdS in momentum and position space, construct the bulk-to-boundary propagator, reduce the on-shell action to a boundary term at $z = \epsilon$, and renormalize with local counterterms, *holographic renormalization*, obtaining $\langle \mathcal{O} \rangle = (2\Delta - d) \times$ (normalizable coefficient) and $\langle \mathcal{O}\mathcal{O} \rangle \propto (2\Delta - d)\, c_\Delta/|x|^{2\Delta}$, with attention to a normalization subtlety invisible to conformal symmetry but fatal to Ward identities. A momentum-space computation cross-checks the result and reveals the logarithms at integer $\Delta - d/2$. Cubic bulk couplings then produce conformally covariant three-point functions, and the perturbative solution of the bulk equations organizes into *Witten diagrams*, the Feynman rules of holography.

13.1 The recipe

We work in Euclidean signature (Lorentzian remarks in Section 13.6), on Euclidean AdS_{d+1} in Poincaré coordinates,

$$\mathrm{d}s^2 \;=\; L^2\,\frac{\mathrm{d}z^2 + \delta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu}{z^2}, \qquad z > 0, \tag{13.1}$$

which covers the whole of hyperbolic space, with conformal boundary $\mathbb{R}^d \cup \{\infty\} = S^d$ (Chapter 10). The semiclassical GKPW prescription reads

$$W[\phi_0] \;=\; \ln \left\langle \mathrm{e}^{\int \mathrm{d}^d x\, \phi_0\, \mathcal{O}} \right\rangle_{\mathrm{CFT}} \;=\; -\,S_{\mathrm{bulk}}[\hat{\phi}], \tag{13.2}$$

with $\hat{\phi}(z, x)$ the solution of the bulk equations that is regular in the interior and behaves as $\hat{\phi} \to z^{d-\Delta}\phi_0(x)$ at the boundary, and connected correlators follow by differentiation,

$$\langle \mathcal{O}(x_1) \cdots \mathcal{O}(x_n) \rangle_{\mathrm{c}} \;=\; \left.\frac{\delta^n W[\phi_0]}{\delta\phi_0(x_1) \cdots \delta\phi_0(x_n)}\right|_{\phi_0 = 0}. \tag{13.3}$$

The scheme has three steps, which structure the chapter.

1. *Extend*: solve the bulk equations with the source as boundary data. Since the solution vanishes when the sources are switched off, a term in the bulk action with more than n powers of the fields cannot contribute to an n-point function: the free action suffices for two-point functions, cubic couplings are needed for three-point functions, and so on. The extension is perturbative in the bulk couplings, which are of order $1/N$ (Chapter 11).

2. *Evaluate*: compute the on-shell action, which for the quadratic part reduces by an integration by parts to a boundary term.

3. *Renormalize*: the boundary term diverges as the cutoff surface $z = \epsilon$ approaches the boundary. The divergences are local functionals of the sources and are removed by boundary counterterms; the renormalized action generates the correlators. By the UV/IR connection of Section 12.5 this is the holographic image of UV renormalization in the CFT.

13.2 The free scalar in Euclidean AdS

Consider a single scalar of mass m with action

$$S \;=\; \frac{1}{2}\int \mathrm{d}^{d+1}x\sqrt{g}\,\Big(g^{MN}\partial_M\phi\,\partial_N\phi + m^2\phi^2\Big), \tag{13.4}$$

dual to a scalar operator $\mathcal{O}$ of dimension

$$\Delta \;=\; \frac{d}{2}+\nu, \qquad \nu \;\equiv\; \sqrt{\frac{d^2}{4}+m^2L^2} \;=\; \Delta - \frac{d}{2} \;\; (\geq 0). \tag{13.5}$$

From here on $L = 1$; the AdS radius is restored by dimensional analysis (each correlator acquires a prefactor L^{d-1}, which in the canonical example carries the factor N^2 via $L^{d-1}/G \sim N^2$, Section 12.5). With $\sqrt{g} = z^{-(d+1)}$ and $g^{MN} = z^2\delta^{MN}$, the equation of motion $(\Box_g - m^2)\phi = 0$ reads

$$z^{d+1}\,\partial_z\big(z^{1-d}\,\partial_z\phi\big) + z^2\,\partial^\mu\partial_\mu\phi \;=\; m^2\phi, \tag{13.6}$$

whose near-boundary solutions $\phi \sim z^{d-\Delta}, z^{\Delta}$ we know from Chapter 10.

13.2.1 Momentum space: Bessel functions

- *Radial equation.* The coefficients in (13.6) depend only on z, so Fourier transform along the boundary, $\phi(z,x) = f_p(z)\,\mathrm{e}^{\mathrm{i}p\cdot x}$ with $p^2 \geq 0$ Euclidean:

  $$z^2 f_p'' - (d-1)\,z f_p' - \big(p^2z^2 + m^2\big) f_p \;=\; 0. \tag{13.7}$$

  Substituting $f_p(z) = z^{d/2}\,y(u)$ with $u = pz$ removes the friction term and, using $m^2 = \nu^2 - d^2/4$, gives the modified Bessel equation

  $$u^2\,y'' + u\,y' - \big(u^2+\nu^2\big)\,y \;=\; 0, \qquad y \;=\; I_\nu(u) \;\text{ or }\; K_\nu(u). \tag{13.8}$$

- *Regularity.* The two solutions differ deep in the bulk,

  $$I_\nu(u) \;\sim\; \frac{\mathrm{e}^u}{\sqrt{2\pi u}}, \qquad K_\nu(u) \;\sim\; \sqrt{\frac{\pi}{2u}}\,\mathrm{e}^{-u}, \qquad u\to\infty, \tag{13.9}$$

  so regularity at $z\to\infty$, the single point at infinite distance in Euclidean AdS, selects

  $$\boxed{f_p(z) \;\propto\; z^{d/2}\,K_\nu(pz). \tag{13.10}}$$

- *Source and response.* Near the boundary, for non-integer ν (the integer case is in Section 13.4),

  $$K_\nu(u) \;=\; \frac{\Gamma(\nu)}{2}\Big(\frac{u}{2}\Big)^{-\nu}\big(1+O(u^2)\big) \;+\; \frac{\Gamma(-\nu)}{2}\Big(\frac{u}{2}\Big)^{\nu}\big(1+O(u^2)\big), \tag{13.11}$$

reproduces the two branches $z^{d/2\mp\nu} = z^{d-\Delta}, z^{\Delta}$ of Chapter 10 with a *fixed relative coefficient*: once regularity is imposed the normalizable coefficient is determined by the source,

$$\frac{\beta(p)}{\alpha(p)} = \frac{\Gamma(-\nu)}{\Gamma(\nu)}\Big(\frac{p}{2}\Big)^{2\nu}. \tag{13.12}$$

This ratio, the *response* of the field to its source, is non-analytic in p^2 and carries the entire physical content of the two-point function; everything analytic in p^2 is a contact term.

13.2.2 Position space: the bulk-to-boundary propagator

- *The kernel.* For an arbitrary source we need the kernel propagating boundary data into the bulk,

$$\hat{\phi}(z,x) = \int \mathrm{d}^d x'\, K_\Delta(z,x;x')\,\phi_0(x'), \tag{13.13}$$

characterized by

$$(\Box_g - m^2)K_\Delta = 0, \qquad K_\Delta(z,x;x') \xrightarrow{z\to 0} z^{\,d-\Delta}\,\delta^{(d)}(x-x'). \tag{13.14}$$

- *Witten's construction* [60]. The function $\phi = z^\Delta$ solves the equation of motion exactly. On the compactified boundary S^d it vanishes at every point of $\mathbb{R}^d$ and blows up at $z=\infty$: it is the field sourced by a delta function *at the point at infinity.* To move the source to a finite point, apply the isometry exchanging infinity with the origin, the *inversion*

$$z \to \frac{z}{z^2+x^2}, \qquad x^\mu \to \frac{x^\mu}{z^2+x^2}, \tag{13.15}$$

which preserves the metric (13.1) (it is the conformal inversion of Chapter 8 lifted to the bulk). Applied to z^Δ, followed by a translation of x', it gives

$$\boxed{K_\Delta(z,x;x') = c_\Delta\left(\frac{z}{z^2+|x-x'|^2}\right)^{\Delta}, \qquad c_\Delta = \frac{\Gamma(\Delta)}{\pi^{d/2}\,\Gamma(\Delta-\frac{d}{2})}. \qquad (13.16)}$$

- *Boundary condition and normalization.* As $z\to 0$ at fixed $x \neq x'$, $K_\Delta \sim z^\Delta \to 0$: the kernel is supported, in the limit, only at $x = x'$. Integrating against a test source and substituting $x' = x - zy$,

$$\begin{aligned}\int \mathrm{d}^d x'\, K_\Delta(z,x;x')\,\phi_0(x') &= c_\Delta\, z^{d-\Delta}\int \mathrm{d}^d y\,\frac{\phi_0(x-zy)}{(1+y^2)^\Delta} \\ &\xrightarrow{z\to 0}\; c_\Delta\, z^{d-\Delta}\,\phi_0(x)\int\frac{\mathrm{d}^d y}{(1+y^2)^\Delta},\end{aligned} \tag{13.17}$$

and the standard integral

$$\int\frac{\mathrm{d}^d y}{(1+y^2)^\Delta} = \frac{\pi^{d/2}\,\Gamma(\Delta-\frac{d}{2})}{\Gamma(\Delta)} = \frac{1}{c_\Delta} \tag{13.18}$$

fixes c_Δ as quoted in (13.16) and verifies $K_\Delta \to z^{d-\Delta}\delta^{(d)}(x-x')$.

- *Equation of motion.* ⋆ *Self-study.* Isometries map solutions to solutions, so the construction guarantees it; a direct check is instructive. Write $K = c_\Delta z^\Delta/\rho^\Delta$ with $\rho \equiv z^2 + r^2$, $r = |x - x'|$. Acting with the two parts of (13.6),

$$z^{d+1}\partial_z(z^{1-d}\partial_z K) = c_\Delta \left[\Delta(\Delta - d)\,\frac{z^\Delta}{\rho^\Delta} - 2\Delta(2\Delta - d + 2)\,\frac{z^{\Delta+2}}{\rho^{\Delta+1}} + 4\Delta(\Delta+1)\,\frac{z^{\Delta+4}}{\rho^{\Delta+2}}\right],$$
$$z^2\,\partial^\mu\partial_\mu K = c_\Delta \left[-2\Delta d\,\frac{z^{\Delta+2}}{\rho^{\Delta+1}} + 4\Delta(\Delta+1)\,\frac{z^{\Delta+2}\,r^2}{\rho^{\Delta+2}}\right]. \tag{13.19}$$

Adding the two lines and using $r^2 = \rho - z^2$, the $z^{\Delta+4}/\rho^{\Delta+2}$ pieces cancel, the $z^{\Delta+2}/\rho^{\Delta+1}$ pieces have total coefficient $-2\Delta(2\Delta - d + 2) - 2\Delta d + 4\Delta(\Delta + 1) = 0$, and what remains is $\Delta(\Delta - d)K = m^2 K$.

- *The response.* Expanding (13.13) in powers of z,

$$\hat\phi(z,x) \;=\; z^{d-\Delta}(\phi_0(x) + O(z^2)) \,+\, z^\Delta(\phi_1(x) + O(z^2)), \qquad \phi_1(x) \;=\; c_\Delta \int \mathrm{d}^d x'\, \frac{\phi_0(x')}{|x - x'|^{2\Delta}}, \tag{13.20}$$

the coefficient of z^Δ coming from the pointwise limit $K_\Delta \to c_\Delta z^\Delta/|x - x'|^{2\Delta}$. The response is a *non-local* functional of the source, the convolution with the conformal two-point kernel, while the $O(z^2)$ towers are local differential operators acting on ϕ_0 and ϕ_1; the first subleading source coefficient is

$$\phi_{0,2} \;=\; \frac{\partial^2\phi_0}{2\,(2\Delta - d - 2)}, \tag{13.21}$$

by matching powers of z in (13.6).[1] The source determines the entire solution, and the expectation-value mode ϕ_1 is a *computed* quantity: the one-point function in the presence of the source, as we now confirm.

13.3 The on-shell action and holographic renormalization

1. *Reduction to a boundary term.* Integrating (13.4) by parts,

$$S \;=\; \frac{1}{2}\int \mathrm{d}^{d+1}x\;\partial_M\big(\sqrt{g}\,\phi\, g^{MN}\partial_N\phi\big) \;-\; \frac{1}{2}\int \mathrm{d}^{d+1}x\sqrt{g}\;\phi\,(\Box_g - m^2)\phi, \tag{13.22}$$

the second term vanishes on shell and the first is a total derivative. Regulating the space by $z \geq \epsilon$, the boundary is the surface $z = \epsilon$ (the point $z \to \infty$ contributes nothing for the regular solution), the outward direction is $-\partial_z$, and $\sqrt{g}\, g^{zz} = z^{1-d}$:

$$S_{\rm reg}[\phi_0;\epsilon] \;=\; -\frac{1}{2}\int \mathrm{d}^d x\,\Big[z^{1-d}\,\hat\phi\,\partial_z\hat\phi\Big]_{z=\epsilon}. \tag{13.23}$$

The on-shell action of a free field in AdS is a pure boundary term, evaluated at the cutoff surface.

2. *Divergences.* Inserting $\hat\phi = z^{d-\Delta}(\phi_0 + \phi_{0,2}z^2 + \cdots) + z^\Delta(\phi_1 + \cdots)$,

$$z^{1-d}\,\hat\phi\,\partial_z\hat\phi\Big|_{z=\epsilon} = (d-\Delta)\,\phi_0^2\,\epsilon^{d-2\Delta} \;+\; \big(2(d-\Delta)+2\big)\,\phi_0\,\phi_{0,2}\,\epsilon^{d-2\Delta+2} + \cdots$$
$$+\; d\,\phi_0\,\phi_1 \;+\; O(\epsilon^2), \tag{13.24}$$

[1] Substitute $\hat\phi = z^{d-\Delta}(\phi_0 + \phi_{0,2}z^2)$ into (13.6) and use $z^{d+1}\partial_z(z^{1-d}\partial_z z^a) = a(a-d)z^a$. The coefficient of $z^{d-\Delta}$ gives $(d-\Delta)(-\Delta) = m^2$; that of $z^{d-\Delta+2}$ gives $(d-\Delta+2)(2-\Delta)\,\phi_{0,2} + \partial^2\phi_0 = m^2\phi_{0,2}$, and with $m^2 = \Delta(\Delta - d)$ the terms quadratic in Δ cancel, leaving $-2(2\Delta - d - 2)\,\phi_{0,2} = -\partial^2\phi_0$.

where the first line is a finite chain of divergent terms (as many as there are even integers between 0 and $2\Delta - d$), each *local* in the source (ϕ_0^2, $\phi_0\partial^2\phi_0$, ...), while the finite cross term $\phi_0\phi_1$ is non-local. (Both orderings contribute to $d\,\phi_0\phi_1$: $\phi_0 z^{d-\Delta}\,\partial_z(\phi_1 z^{\Delta})$ gives $\Delta\,\phi_0\phi_1$ and $\phi_1 z^{\Delta}\,\partial_z(\phi_0 z^{d-\Delta})$ gives $(d-\Delta)\,\phi_0\phi_1$.)

3. *Counterterms.* Divergent local functionals of the sources are what renormalization removes, and the same is done here with *local, covariant boundary counterterms* built from the induced metric $\gamma_{\mu\nu} = \delta_{\mu\nu}/\epsilon^2$ and the field at the cutoff surface,

$$S_{\rm ct} \;=\; \int \mathrm{d}^d x \sqrt{\gamma}\,\left(\frac{d-\Delta}{2}\,\phi^2 \;+\; \frac{1}{2(2\Delta-d-2)}\,\phi\,\Box_\gamma\phi \;+\; \cdots\right)_{z=\epsilon}, \tag{13.25}$$

one term per divergence. The first does its job and leaves a finite piece behind: with $\sqrt{\gamma} = \epsilon^{-d}$ and $\phi^2|_\epsilon = \phi_0^2\epsilon^{2(d-\Delta)} + 2\phi_0\phi_1\epsilon^d + \cdots$,

$$\int \mathrm{d}^d x \sqrt{\gamma}\,\frac{d-\Delta}{2}\,\phi^2 = \int \mathrm{d}^d x\left[\frac{d-\Delta}{2}\,\phi_0^2\,\epsilon^{\,d-2\Delta} + (d-\Delta)\,\phi_0\phi_1\right] + \cdots, \tag{13.26}$$

so that, since $-\frac{d}{2} + (d-\Delta) = -\frac{2\Delta-d}{2}$ and the derivative counterterms cancel the remaining divergences without touching the non-local finite part,

$$S_{\rm ren}[\phi_0] \;=\; \lim_{\epsilon\to 0}\Big(S_{\rm reg} + S_{\rm ct}\Big) \;=\; -\,\frac{2\Delta-d}{2}\int \mathrm{d}^d x\;\phi_0(x)\,\phi_1(x). \tag{13.27}$$

The finite leftovers of counterterms are the entire content of the normalization subtlety below.

4. *One-point function.* Differentiating $W = -S_{\rm ren}$ once at non-zero source (the derivative acts on both factors, which contribute equally since (13.20) has a symmetric kernel),

$$\boxed{\langle\mathcal{O}(x)\rangle_{\phi_0} \;=\; \frac{\delta W}{\delta\phi_0(x)} \;=\; (2\Delta-d)\;\phi_1(x)\,: \tag{13.28}}$$

the one-point function of $\mathcal{O}$ is $(2\Delta - d)$ times the normalizable coefficient of the on-shell field.

At zero source ϕ_1 vanishes, as scale invariance requires. But (13.28) is far more general than its derivation: in any asymptotically AdS configuration (a black hole, a domain wall, an excited state) the coefficient of z^Δ measures $\langle\mathcal{O}\rangle$ in the corresponding CFT state. This entry does heavy lifting in Chapters 14 and 18 and confirms the identification of normalizable modes with states of Chapter 12.

5. *Two-point function.* Differentiating once more and inserting (13.20):

$$\boxed{\langle\mathcal{O}(x)\,\mathcal{O}(x')\rangle \;=\; (2\Delta-d)\;c_\Delta\;\frac{1}{|x-x'|^{2\Delta}} \;=\; \frac{(2\Delta-d)\,\Gamma(\Delta)}{\pi^{d/2}\,\Gamma\big(\Delta-\frac{d}{2}\big)}\;\frac{1}{|x-x'|^{2\Delta}}. \tag{13.29}}$$

The form $|x-x'|^{-2\Delta}$ is the unique one allowed by conformal invariance (Chapter 9); the non-trivial information is the normalization, which must therefore be computed honestly.

13.3.1 The normalization subtlety

- *A tempting shortcut.* Evaluate (13.23) directly with (13.13), using $\partial_z K_\Delta \to (\Delta/z) K_\Delta$ at small z (correct pointwise at separated points) and discarding the divergent contact terms:

$$S \;\approx\; -\frac{\Delta}{2}\, c_\Delta \int \mathrm{d}^d x\, \mathrm{d}^d x' \,\frac{\phi_0(x)\,\phi_0(x')}{|x-x'|^{2\Delta}} \qquad \Longrightarrow \qquad \langle \mathcal{O}\mathcal{O}\rangle_{\text{naive}} \;=\; \frac{\Delta\, c_\Delta}{|x-x'|^{2\Delta}}, \tag{13.30}$$

  the right conformal structure with the *wrong coefficient*,

$$\langle \mathcal{O}\mathcal{O}\rangle_{\text{correct}} \;=\; \frac{2\Delta - d}{\Delta}\, \langle \mathcal{O}\mathcal{O}\rangle_{\text{naive}}. \tag{13.31}$$

- *Why it fails.* Dropping divergent terms is not subtracting covariant counterterms: the counterterms leave finite remainders, as in (13.27), and the pointwise limit of $\partial_z K$ misses contributions at coincident points that mix with them. The mistake is invisible to conformal symmetry but not to physics: two-point normalizations are tied by Ward identities to three-point functions of currents and to central charges, and only (13.29) satisfies them. It went unnoticed at first because the earliest examples were marginal, $\Delta = d$, where $(2\Delta - d)/\Delta = 1$.

- *Holographic renormalization.* The systematic procedure, regularize at $z = \epsilon$, subtract covariant local counterterms, remove the cutoff, was developed in generality (scalars in arbitrary backgrounds, the metric, the logarithmic cases) by Skenderis and collaborators [12], whose conventions we follow. Applied to the gravitational action itself it requires the counterterms

$$S^{\text{grav}}_{\text{ct}} \;=\; \frac{1}{8\pi G} \oint \mathrm{d}^d x \sqrt{\gamma}\left(\frac{d-1}{L} + \frac{L}{2(d-2)}\, R[\gamma] + \cdots \right), \tag{13.32}$$

  which reappear in Chapter 14 (in $d = 4$ the series ends in a *logarithmic* divergence whose coefficient is the Weyl anomaly) and in Chapter 18 (black hole free energies).

13.4 Momentum-space cross-check and logarithms ⋆

1. *Cutoff solution.* With the regular solution (13.10) and the cutoff boundary condition $\hat{\phi}_p(\epsilon) = \epsilon^{d-\Delta} \phi_0(p)$,

$$\hat{\phi}_p(z) \;=\; \epsilon^{\,d-\Delta}\, \frac{z^{d/2} K_\nu(pz)}{\epsilon^{d/2} K_\nu(p\epsilon)}\, \phi_0(p), \tag{13.33}$$

   and writing $\hat{\phi}_p \partial_z \hat{\phi}_p = \hat{\phi}_p^2\, \partial_z \ln \hat{\phi}_p$ in (13.23),

$$S_{\text{reg}} = -\frac{1}{2} \int \frac{\mathrm{d}^d p}{(2\pi)^d}\, \phi_0(p)\, \phi_0(-p)\, \mathcal{F}(p,\epsilon), \qquad \mathcal{F} \;=\; \epsilon^{\,d-2\Delta+1}\, \partial_z \ln\left(z^{d/2} K_\nu(pz)\right)\Big|_{z=\epsilon}. \tag{13.34}$$

2. *Expansion.* By (13.11),

$$\partial_u \ln\left(u^{d/2} K_\nu(u)\right) \;=\; \frac{d-\Delta}{u}\Big(1+\cdots\Big) \;+\; \frac{2\nu\,\Gamma(-\nu)}{\Gamma(\nu)}\, \frac{u^{2\nu-1}}{2^{2\nu}}\Big(1+\cdots\Big), \tag{13.35}$$

   with the ellipses integer powers of u^2, and since $u = pz$,

$$\mathcal{F} \;=\; \epsilon^{\,d-2\Delta+1}\, p \left[\partial_u \ln\left(u^{d/2} K_\nu(u)\right)\right]_{u=p\epsilon}. \tag{13.36}$$

The analytic powers of p^2 come with inverse powers of ϵ: they are the divergences removed by the counterterms (polynomials in p^2, i.e. contact terms). The non-analytic term carries $\epsilon^{d-2\Delta+2\nu} = \epsilon^0$: cutoff-independent, it survives renormalization untouched, so $\mathcal{F}_{\rm ren}(p)$ is that term alone.

3. *Result.* Differentiating twice as in (13.3) (the $\frac{1}{2}$ cancels against the two ways of hitting the sources) and stripping $(2\pi)^d\delta^{(d)}(p+p')$, with $2\nu = 2\Delta - d$,

$$\langle \mathcal{O}(p)\,\mathcal{O}(-p)\rangle \;=\; (2\Delta-d)\,\frac{\Gamma(-\nu)}{\Gamma(\nu)}\left(\frac{p}{2}\right)^{2\nu} \;=\; (2\Delta-d)\,\frac{\beta(p)}{\alpha(p)}, \tag{13.37}$$

a pure non-analytic power of momentum, as conformal invariance requires. Transforming back with

$$\int \frac{\mathrm{d}^d p}{(2\pi)^d}\,\mathrm{e}^{\mathrm{i}p\cdot x}\,p^{2\nu} \;=\; \frac{2^{2\nu}}{\pi^{d/2}}\,\frac{\Gamma(\frac{d}{2}+\nu)}{\Gamma(-\nu)}\,\frac{1}{|x|^{d+2\nu}}, \tag{13.38}$$

the $\Gamma(-\nu)$ cancel and, with $d/2+\nu=\Delta$,

$$\langle \mathcal{O}(x)\,\mathcal{O}(0)\rangle \;=\; (2\Delta-d)\,\frac{\Gamma(\Delta)}{\pi^{d/2}\,\Gamma(\Delta-\frac{d}{2})}\,\frac{1}{|x|^{2\Delta}} \;=\; \frac{(2\Delta-d)\,c_\Delta}{|x|^{2\Delta}}, \tag{13.39}$$

in exact agreement with (13.29), normalization included.

In momentum space the two-point function is $(2\Delta-d)$ times the source-to-response ratio $\beta(p)/\alpha(p)$ of the regular solution, (13.12): all dynamical information sits in the non-analytic momentum dependence $p^{2\nu}$, whose Fourier transform is $1/|x|^{2\Delta}$; analytic terms in p^2 are scheme-dependent contact terms.

- *Integer ν and logarithms.* When $\nu = \Delta - \frac{d}{2}$ is a non-negative integer, as for $\Delta = 4$, $d = 4$, the expansion (13.11) degenerates: $\Gamma(-\nu)$ has a pole and K_ν contains $u^{2\nu}\ln u$ in place of the naive power. The non-analytic dependence becomes $p^{2\nu}\ln p$ and the cutoff cannot be avoided even in principle, $\langle\mathcal{O}(p)\mathcal{O}(-p)\rangle \propto p^{2\nu}\ln(p\epsilon) + (\text{analytic})$, the scale in the logarithm reflecting the scheme. The classic example [59, 60, 2] is the massless scalar in AdS$_5$ ($\Delta = d = 4$, $\nu = 2$), dual to $\frac{1}{N}\,\mathrm{Tr}\,F^2$ of $\mathcal{N}=4$ SYM: $f_p \propto z^2 K_2(pz)$ and

$$u^2K_2(u) \;=\; 2 - \frac{u^2}{2} - \frac{u^4}{8}\,\ln u + (\text{analytic } u^4 \text{ terms}) + \cdots\,, \tag{13.40}$$

give[2]

$$\langle\mathcal{O}(p)\mathcal{O}(-p)\rangle \;\sim\; \frac{p^2}{\epsilon^2}\,(\text{contact}) \;+\; p^4\ln(p\,\epsilon) + \cdots \qquad\Longrightarrow\qquad \langle\mathcal{O}(x)\mathcal{O}(0)\rangle \;\propto\; \frac{1}{|x|^8}, \tag{13.41}$$

the Fourier transform of $p^4\ln p$ being $\propto |x|^{-8}$, the two-point function of a dimension-four operator. The logarithm is not an anomaly of the correlator's *form*, which is scale covariant at separated points, but signals the contact-term scheme dependence of even integer 2ν; the same mechanism produces the holographic Weyl anomaly in Chapter 14.

[2] Write $u^2K_2(u) = 2\big(1 - \frac{u^2}{4} - \frac{u^4}{16}\ln u + O(u^4)\big)$, so $\partial_u \ln\big(u^2K_2(u)\big) = -\frac{u}{2} - \frac{u^3}{4}\ln u + (\text{analytic})$. In (13.36) with $u = p\epsilon$ and $d - 2\Delta + 1 = -3$, the analytic term gives the contact term $-p^2/2\epsilon^2$ and the logarithmic term $-\frac{1}{4}p^4\ln(p\epsilon)$, whose power of ϵ cancels but whose logarithm retains the cutoff.

13.5 Interactions: three-point functions and Witten diagrams

- *Perturbative solution.* Let the bulk contain scalars ϕ_i of masses m_i, dual to $\mathcal{O}_i$ of dimensions Δ_i, with local interactions

$$S \;=\; \int \mathrm{d}^{d+1}x\sqrt{g}\,\left[\sum_i \left(\frac{1}{2}(\partial\phi_i)^2 + \frac{m_i^2}{2}\phi_i^2\right) + \frac{\lambda_{ijk}}{3}\,\phi_i\phi_j\phi_k + \cdots\right], \tag{13.42}$$

the couplings being outputs of a compactification (Chapter 17) or, here, inputs. The equations $(\Box - m_i^2)\phi_i = \lambda_{ijk}\phi_j\phi_k + \cdots$ are solved iteratively: $\phi_i^{(0)} = \int K_{\Delta_i}\phi_0^i$, and at each further order $(\Box - m^2)$ is inverted with the *bulk-to-bulk propagator*, the normalizable solution of

$$(\Box_g - m^2)\,G_\Delta(z,x;z',x') \;=\; \frac{1}{\sqrt{g}}\,\delta(z-z')\,\delta^{(d)}(x-x'), \tag{13.43}$$

a hypergeometric function of the chordal distance, not needed here, which reduces to $K_\Delta/(2\Delta - d)$ when one point goes to the boundary. Since an n-point function needs only n powers of the sources, the iteration terminates. The series has a graphical representation (Figure 13.1):

Witten diagram rules (Euclidean AdS, drawn as a disk whose circle is the boundary):

1. an n-point function is a sum of diagrams with n points on the boundary circle;
2. each boundary point x_i connects into the bulk via a bulk-to-boundary propagator $K_{\Delta_i}(z,w;x_i)$;
3. propagators meet at bulk vertices, one per interaction term, with coupling $-\lambda_{ijk\cdots}$; internal lines between two bulk vertices are bulk-to-bulk propagators G_Δ;
4. each vertex position is integrated over the bulk with the invariant measure $\int \mathrm{d}^{d+1}w\sqrt{g}$;
5. the diagram contributes at order N^{2-n} (in units where the two-point function is $O(1)$), with bulk loops suppressed by further powers of $1/N^2$, the semiclassical expansion of Chapter 12.

- *The three-point function.* Three sources must end on bulk interactions, and the only diagram with at most three powers of the sources is the cubic contact diagram,

$$\langle \mathcal{O}_1(x_1)\,\mathcal{O}_2(x_2)\,\mathcal{O}_3(x_3)\rangle \;=\; -\,\lambda_{123}\int \mathrm{d}^{d+1}w\sqrt{g}\;K_{\Delta_1}(w;x_1)\,K_{\Delta_2}(w;x_2)\,K_{\Delta_3}(w;x_3), \tag{13.44}$$

with $w = (z,\vec{w}\,)$ the bulk point. The integral is fixed by the same symmetry that built the propagators: under the inversion (13.15) each propagator transforms covariantly,[3]

$$K_\Delta(w';x') \;=\; |x|^{2\Delta}\,K_\Delta(w;x), \tag{13.45}$$

and the measure is invariant. Three moves [60, 61]:

[3] With $w^2 \equiv z^2 + |\vec{w}|^2$ the inversion reads $w' = w/w^2$. For any two points u, v of the half space, $(u'-v')^2 = (u-v)^2/(u^2v^2)$; taking $u = (z,\vec{w}\,)$ and the boundary point $v = (0,x)$, so that $v^2 = |x|^2$ and $v' = (0,x')$, gives $z'/\big(z'^2 + (\vec{w}' - x')^2\big) = |x|^2\,z/\big(z^2 + (\vec{w} - x)^2\big)$, hence $K_\Delta(w';x') = |x|^{2\Delta}K_\Delta(w;x)$ by (13.16). The factors of w^2 cancel: the transformation factor depends only on the boundary point, as conformal covariance requires.

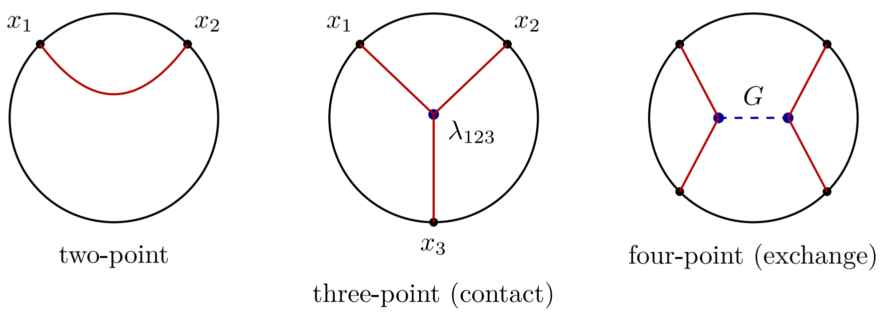


Figure 13.1: Witten diagrams. The disk is (Euclidean) AdS, its boundary circle the CFT spacetime. Solid red lines from the boundary are bulk-to-boundary propagators K_Δ; the dashed internal line is a bulk-to-bulk propagator G; bulk vertices (blue) are integrated over AdS. The two-point function needs no vertex; the three-point function is exhausted by the single contact diagram; four-point functions receive exchange diagrams in all channels plus a quartic contact term.

1. Set $x_3 = 0$ by translation and invert all points, $w = w'/w'^2$, $x_i = x'_i/|x'_i|^2$. The third propagator *collapses*, since $z/(z^2+|\vec{w}|^2) = z'$: $K_{\Delta_3}(w;0) = c_{\Delta_3} z'^{\Delta_3}$, its boundary point having gone to infinity, while the other two produce their covariance factors,
$$\langle \mathcal{O}_1\mathcal{O}_2\mathcal{O}_3\rangle \;=\; -\lambda_{123}\, c_{\Delta_3}\, |x'_1|^{2\Delta_1}\, |x'_2|^{2\Delta_2} \times \int \mathrm{d}^{d+1}w'\; z'^{\,\Delta_3-d-1} \prod_{i=1,2} c_{\Delta_i}\Big(\frac{z'}{z'^2+(\vec{w}'-x'_i)^2}\Big)^{\Delta_i}. \tag{13.46}$$
2. The surviving integral is fixed by symmetry up to a constant: translations make it a function of x'_{12} only, and under $w' \to \lambda w'$, $x'_i \to \lambda x'_i$ it picks up $\lambda^{\Delta_3-\Delta_1-\Delta_2}$ (invariant measure, $+\Delta_3$ from z'^{Δ_3}, $-\Delta_i$ from each propagator), so
$$\int \mathrm{d}^{d+1}w'\,(\cdots) \;=\; \frac{C_{123}}{|x'_{12}|^{\Delta_1+\Delta_2-\Delta_3}}, \tag{13.47}$$
with C_{123} a pure number, evaluated by Schwinger parameters in [61].
3. Undo the inversion with $|x'_i| = 1/|x_i|$ and $|x'_{12}| = |x_{12}|/(|x_1||x_2|)$: the powers reassemble as $(\Delta_1+\Delta_2-\Delta_3)-2\Delta_1 = -(\Delta_3+\Delta_1-\Delta_2)$ and $(\Delta_1+\Delta_2-\Delta_3)-2\Delta_2 = -(\Delta_2+\Delta_3-\Delta_1)$, and restoring x_3 by translation ($|x_1| \to |x_{31}|$, $|x_2| \to |x_{23}|$):

$$\langle \mathcal{O}_1\mathcal{O}_2\mathcal{O}_3\rangle \;=\; \frac{a_{123}}{|x_{12}|^{\Delta_1+\Delta_2-\Delta_3}\,|x_{23}|^{\Delta_2+\Delta_3-\Delta_1}\,|x_{31}|^{\Delta_3+\Delta_1-\Delta_2}}, \qquad x_{ij} \equiv x_i - x_j, \tag{13.48}$$

exactly the coordinate dependence conformal invariance dictates for three primaries (Chapter 9), with structure constant

$$a_{123} \;=\; -\lambda_{123}\, \frac{\Gamma\big[\frac{\delta_{12}}{2}\big]\,\Gamma\big[\frac{\delta_{23}}{2}\big]\,\Gamma\big[\frac{\delta_{31}}{2}\big]}{2\,\pi^d\,\Gamma\big[\Delta_1-\frac{d}{2}\big]\,\Gamma\big[\Delta_2-\frac{d}{2}\big]\,\Gamma\big[\Delta_3-\frac{d}{2}\big]}\;\Gamma\Big[\tfrac{1}{2}\Big(\textstyle\sum_i \Delta_i - d\Big)\Big], \tag{13.49}$$

where $\delta_{ij} \equiv \Delta_i + \Delta_j - \Delta_k$ $(k \neq i,j)$ are the exponents in (13.48).

- *Kinematics and dynamics.* The coordinate dependence is guaranteed by the isometries; the OPE coefficient a_{123} is computed by the bulk coupling. Four-point functions, built from exchange and contact diagrams, are not fixed by symmetry (they depend on two cross-ratios) and are genuine dynamical checks: the holographic answers, expressible through "D-functions", integrals of four K's, expand in conformal blocks consistently with an OPE containing the single-trace $\mathcal{O}_i$ and the *double-trace* composites $\mathcal{O}_i \partial^{2n} \partial_{\mu_1} \cdots \partial_{\mu_\ell} \mathcal{O}_j$ of dimensions $\Delta_i + \Delta_j + 2n + \ell + O(1/N^2)$, the CFT image of two-particle states in AdS. That GKPW passes such checks is strong evidence that it defines bona fide CFT correlators.

- *Higher spins.* The same machinery, with more indices, computes correlators of currents and stress tensors from bulk gauge fields and gravitons. The current dual to a massless vector in AdS_5 comes out transverse,

$$\langle J_\mu(x) J_\nu(0) \rangle \;\propto\; \frac{1}{|x|^6} \left(\delta_{\mu\nu} - 2\, \frac{x_\mu x_\nu}{|x|^2} \right), \tag{13.50}$$

with $\Delta = 3$ exactly: bulk gauge invariance has produced a conserved current of protected dimension. The normalizations of (13.50) and of the stress-tensor correlator give the current central charges and c; getting them right requires the holographic renormalization of Section 13.3, and comparing with field theory gives the tests of Chapter 14.

13.6 Lorentzian remarks

Everything above was Euclidean, where the regular extension of boundary data is unique and the generating functional computes vacuum correlators, to be continued to any Lorentzian ordering afterwards. Direct Lorentzian computations bring three novelties, developed where needed (Chapters 18–21).

- *Choices at the horizon.* The interior condition is imposed at the Poincaré horizon, and different choices (in-falling, out-going, iϵ deformations of the Euclidean prescription) yield retarded, advanced or time-ordered Green functions. In-falling conditions compute retarded correlators, hence transport coefficients: the basis of holographic hydrodynamics.

- *Normalizable modes are states.* They are genuine propagating excitations (Chapter 10), and turning them on changes the state in which correlators are evaluated: the bulk configuration is no longer unique given the sources, and (13.28) acquires its full force.

- *Global versus Poincaré.* Global AdS computes correlators of the CFT on the cylinder, with the AdS spectrum matching the cylinder energies, i.e. operator dimensions, the state–operator map of Chapter 9 realized geometrically: normal modes of a scalar in global AdS have $\omega L = \Delta + \ell + 2n$, the tower of descendants.

13.7 The dictionary, established

Everything in the following box has now been either derived within the semiclassical bulk description or defined and checked for consistency.

The holographic dictionary (Euclidean signature, semiclassical bulk):

CFT$_d$	**gravity on** AdS_{d+1}
generating functional $W[\phi_0]$	on-shell action: $W = -S_{\mathrm{ren}}[\phi \to z^{d-\Delta}\phi_0]$
single-trace primary $\mathcal{O}$, dimension Δ	bulk field ϕ, $m^2L^2 = \Delta(\Delta - d)$ (scalar)
source ϕ_0 of $\mathcal{O}$	non-normalizable coefficient ($z^{d-\Delta}$)
$\langle\mathcal{O}\rangle$	$(2\Delta - d)\times$ normalizable coefficient (z^{Δ})
$\langle\mathcal{O}\mathcal{O}\rangle = \dfrac{(2\Delta - d)\, c_\Delta}{\lvert x\rvert^{2\Delta}}$	free field: bulk-to-boundary propagator K_Δ, holographically renormalized
OPE coefficients a_{ijk}	bulk cubic couplings λ_{ijk} via (13.49)
n-point correlators	Witten diagrams: K's, G's, vertices integrated over AdS
UV divergences, counterterms, scheme	$z = \epsilon$ cutoff, covariant boundary counterterms, finite counterterm ambiguity
conserved current J_μ ($\Delta = d - 1$)	massless bulk gauge field
stress tensor $T_{\mu\nu}$ ($\Delta = d$)	graviton
$1/N^2$ corrections	bulk quantum loops
$1/\sqrt{\lambda}$ corrections	worldsheet (α') corrections

What the dictionary still lacks is a microscopic derivation, an argument that a specific CFT and a specific bulk string theory are the same object, fixing the couplings that entered here as free parameters. That is the business of Chapters 15–17: string theory, D3-branes, the Maldacena limit, and the spectral match between $\mathcal{N} = 4$ SYM and Type IIB supergravity on $\mathrm{AdS}_5 \times S^5$.

Chapter 14

Holographic Anomalies

We now put the dictionary to work on the observables best suited for a *quantitative test* of the correspondence: anomalies, which are one-loop exact and coupling-independent, so a weak-coupling field-theory computation can legitimately be compared with strong-coupling supergravity. We carry out the holographic Weyl anomaly computation, finding $a = c = \pi L^3/8G_5$ in exact agreement with the $(N^2-1)/4$ of $\mathcal{N}=4$ SYM at large N, and indicate how the R-symmetry triangle anomaly is matched by a five-dimensional Chern–Simons term.

14.1 Anomalies as precision tests of a duality

- *The obstacle.* AdS/CFT is a strong/weak coupling duality, so a generic observable, say the coefficient of the two-point function of an unprotected operator, is a nontrivial function of $\lambda = g_{\rm YM}^2 N$; computing it in perturbation theory and in supergravity produces two numbers with no reason to coincide, and their disagreement tests nothing. To test the duality we need *coupling-independent* observables, computable exactly at weak coupling and predicted by the gravity side.

- *Anomalies are the prime example.* Anomalies of *global* symmetries arise from triangle diagrams with one loop of chiral fermions, and the Adler–Bardeen theorem guarantees that the one-loop triangle is exact: the coefficient is a pure number. 't Hooft's matching argument shows moreover that it is invariant along renormalization-group flows (weakly gauge the symmetry, add free spectator fermions cancelling the anomaly; the resulting gauge theory must stay consistent at all scales, so the anomaly of the dynamical sector cannot change). An anomaly of a global symmetry is not a sickness: it signals that the current, conserved in the vacuum, fails to be conserved once the theory is coupled to *background* fields, external gauge fields for flavour symmetries or an external metric for conformal symmetry. If two theories are dual, their anomalies must match on the nose.

- *In the dictionary.* Both anomalies concern the generating functional of current and stress-tensor correlators,

$$\mathrm{e}^{W[A_\mu,\,g_{\mu\nu}]} = \Big\langle \exp \int \mathrm{d}^4x\sqrt{g}\,\big(\tfrac12 g_{\mu\nu}T^{\mu\nu} + A^i_\mu J^{i\,\mu}\big)\Big\rangle_{\rm CFT}, \tag{14.1}$$

  which holography computes as the on-shell bulk action with boundary conditions A^i_μ, $g_{\mu\nu}$ for the bulk gauge fields and metric. The anomaly is the failure of W to be invariant under gauge transformations of A^i_μ (flavour anomaly) or Weyl rescalings of $g_{\mu\nu}$ (trace anomaly); since holography gives W directly, both failures are computable in the bulk.

14.2 The trace anomaly in four dimensions

- *The general form.* A CFT in flat space has $T^\mu{}_\mu = 0$ as an operator equation. On a curved background the trace acquires an anomalous expectation value, restricted by general covariance and the Wess–Zumino consistency conditions to

$$\langle T^\mu{}_\mu \rangle = \frac{c}{16\pi^2} W_{\mu\nu\rho\sigma} W^{\mu\nu\rho\sigma} - \frac{a}{16\pi^2} E_4, \tag{14.2}$$

with E_4 the Euler density and W^2 the square of the Weyl tensor,

$$E_4 = R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} - 4 R_{\mu\nu} R^{\mu\nu} + R^2, \tag{14.3}$$

$$W_{\mu\nu\rho\sigma} W^{\mu\nu\rho\sigma} = R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} - 2 R_{\mu\nu} R^{\mu\nu} + \tfrac{1}{3} R^2. \tag{14.4}$$

(A $\Box R$ term is scheme-dependent, removed by the counterterm $\int \sqrt{g}\, R^2$; (14.3)–(14.4) cannot be cancelled by any local counterterm.) The numbers a and c are the *central charges*, generalizing the Virasoro c of two-dimensional CFT, where $\langle T^\mu{}_\mu \rangle = -\frac{c}{24\pi} R$. Unitarity forces both positive, and the a-theorem (conjectured by Cardy [62], proven by Komargodski and Schwimmer [63]) states that a decreases along RG flows: a counts degrees of freedom.

- *Free fields.* Heat-kernel methods give, per massless field [4, §6.3], [64],

$$\begin{array}{l|cc} & a & c \\ \hline \text{real scalar} & \frac{1}{360} & \frac{1}{120} \\ \text{Weyl fermion} & \frac{11}{720} & \frac{1}{40} \\ \text{vector field} & \frac{31}{180} & \frac{1}{10} \end{array} \tag{14.5}$$

so a theory with N_0 real scalars, $N_{1/2}$ Weyl fermions and N_1 vectors has

$$a = \frac{2N_0 + 11 N_{1/2} + 124 N_1}{720}, \qquad c = \frac{N_0 + 3N_{1/2} + 12 N_1}{120}. \tag{14.6}$$

- *$\mathcal{N} = 4$ SYM with gauge group* SU(N). Per generator, one vector, four Weyl fermions and six real scalars, all adjoint, so with multiplicity $N^2 - 1$,

$$a = c = (N^2 - 1)\, \frac{2 \cdot 6 + 11 \cdot 4 + 124}{720} = (N^2 - 1)\, \frac{180}{720} = \frac{N^2 - 1}{4}. \tag{14.7}$$

Two features. First, this free-field computation is *exact* for all couplings: g_{YM} is exactly marginal (Chapter 16), so the theory at any λ is a CFT with well-defined a and c, and supersymmetry relates the trace anomaly to triangle anomalies of the R-current ($a, c \sim \operatorname{Tr} R^3, \operatorname{Tr} R$), one-loop exact by Adler–Bardeen; the free-field values therefore hold at $\lambda = \infty$, exactly where supergravity is reliable. Second, $a = c$ exactly, a peculiarity of $\mathcal{N} = 4$ SYM and a special class of theories; generic CFTs have $a \neq c$, and gravity will have something sharp to say about this.

14.3 The holographic Weyl anomaly

We compute the same quantity from the bulk, following Henningson and Skenderis [65]. The logic continues Chapter 13: $W[g]$ is the on-shell gravitational action with boundary metric g, it diverges near the boundary and is renormalized by local counterterms; the new ingredient is a *logarithmic* divergence whose coefficient no local covariant counterterm can remove, the fingerprint of the anomaly.

1. *Fefferman–Graham expansion* [66]. Consider Euclidean gravity in five dimensions,

$$S = -\frac{1}{16\pi G_5}\int \mathrm{d}^5x\,\sqrt{g}\left(R + \frac{12}{L^2}\right) - \frac{1}{8\pi G_5}\oint \mathrm{d}^4x\,\sqrt{\gamma}\,K, \tag{14.8}$$

with the Gibbons–Hawking term of Chapter 7, and the solution asymptotic to AdS_5 with prescribed conformal boundary metric $g_{(0)ij}(x)$. Near the boundary any such solution can be brought to the form

$$\mathrm{d}s^2 = L^2\left(\frac{\mathrm{d}\rho^2}{4\rho^2} + \frac{1}{\rho}\,g_{ij}(x,\rho)\,\mathrm{d}x^i\mathrm{d}x^j\right), \qquad g(x,\rho) = g_{(0)} + \rho\,g_{(2)} + \rho^2 g_{(4)} + \rho^2\ln\rho\;h_{(4)} + \dots \tag{14.9}$$

($\rho = z^2/L^2$ for $g_{(0)} = \delta$ recovers the Poincaré metric). Einstein's equations determine the subleading coefficients *algebraically* in terms of $g_{(0)}$; at first order

$$g_{(2)ij} = -\frac{1}{2}\Big(R_{ij}[g_{(0)}] - \frac{1}{6}\,R[g_{(0)}]\,g_{(0)ij}\Big), \tag{14.10}$$

and at second order they fix $h_{(4)}$ and the trace and divergence of $g_{(4)}$ but not the full tensor $g_{(4)}$: this is the normalizable mode, the choice of state, encoding $\langle T_{ij}\rangle$ by the dictionary of Chapter 12. The logarithmic term $h_{(4)}$, present in even boundary dimensions, is the seed of the anomaly.

2. *The logarithmic divergence.* On shell $R = -20/L^2$, so the bulk integrand is a constant times the volume form and all divergences come from the infinite volume near $\rho = 0$. Cutting off at $\rho = \epsilon$ and evaluating (14.8) on (14.9), the measure $\sqrt{g} \sim \rho^{-3}\sqrt{\det g(x,\rho)}$ gives

$$S_{\text{reg}} = \frac{L^3}{16\pi G_5}\int \mathrm{d}^4x\sqrt{g_{(0)}}\left(\frac{b_{(0)}}{\epsilon^2} + \frac{b_{(2)}}{\epsilon} - \ln\epsilon\;\mathcal{A}(x)\right) + S_{\text{finite}}, \tag{14.11}$$

with $b_{(0)}$ a number and $b_{(2)} \propto R[g_{(0)}]$, removed by the counterterms

$$S_{\text{ct}} = \frac{1}{8\pi G_5}\oint \mathrm{d}^4x\sqrt{\gamma}\left(\frac{3}{L} + \frac{L}{4}\,R[\gamma]\right) \tag{14.12}$$

the gravitational counterterms (13.32) of Chapter 13 for $d = 4$. The logarithm is different in kind: under a boundary Weyl rescaling $g_{(0)} \to \mathrm{e}^{2\sigma} g_{(0)}$ the cutoff surface moves, $\epsilon \to \mathrm{e}^{2\sigma}\epsilon$, so $\ln\epsilon$ shifts additively and no local covariant functional of γ can cancel it. Its coefficient measures the failure of Weyl invariance of

$$W[g_{(0)}] \;\equiv\; -S_{\text{ren}}[g_{(0)}] \;=\; \ln Z_{\text{CFT}}[g_{(0)}], \qquad S_{\text{ren}} = \lim_{\epsilon\to 0}(S_{\text{reg}} + S_{\text{ct}}), \tag{14.13}$$

the metric analogue of (13.2), so that

$$\langle T^\mu{}_\mu\rangle = \frac{2}{\sqrt{g_{(0)}}}\,g_{(0)ij}\frac{\delta W}{\delta g_{(0)ij}} = \frac{L^3}{16\pi G_5}\,\mathcal{A}(x). \tag{14.14}$$

Computing $\mathcal{A}$ means expanding $\sqrt{\det g(x,\rho)}$ to order ρ^2 with (14.10) and the trace of $g_{(4)}$ and collecting the $1/\rho$ term; every step is elementary, and carrying it out once is recommended as self-study [65, 12]. The result is

$$\langle T^\mu{}_\mu\rangle = \frac{L^3}{64\pi G_5}\left(R_{\mu\nu}R^{\mu\nu} - \frac{1}{3}R^2\right), \tag{14.15}$$

with all curvatures those of $g_{(0)}$.

3. *Reading off a and c.* The holographic answer contains no $R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$ term. Since E_4 and W^2 each contain Riemann squared with coefficient one, it cancels in $cW^2 - aE_4$ iff $a = c$, in which case
$$\frac{c\,W^2 - a\,E_4}{16\pi^2}\bigg|_{a=c} = \frac{a}{8\pi^2}\left(R_{\mu\nu}R^{\mu\nu} - \frac{1}{3}R^2\right), \tag{14.16}$$
and matching against (14.15):

Any CFT whose holographic dual is governed by the two-derivative Einstein action (14.8) has central charges $$a = c = \frac{\pi L^3}{8G_5}.$$

The equality $a = c$ holds for *every* theory with a weakly curved Einstein-gravity dual, the first sharp restriction on which CFTs can have one; the common value is fixed by the AdS radius in Planck units.

4. *The match.* For Type IIB on $\mathrm{AdS}_5 \times S^5$ use two facts derived in Chapters 15–16: the ten-dimensional Newton constant
$$16\pi G_{10} = (2\pi)^7 g_s^2\,\alpha'^4, \tag{14.17}$$
and the radius produced by N units of five-form flux, $L^4 = 4\pi g_s N\alpha'^2$. Reducing on the S^5 of radius L, $\mathrm{Vol}(S^5) = \pi^3 L^5$,
$$\frac{1}{G_5} = \frac{\mathrm{Vol}(S^5)}{G_{10}} = \frac{\pi^3 L^5}{G_{10}}, \qquad \Longrightarrow \qquad a = c = \frac{\pi L^3}{8G_5} = \frac{\pi^4 L^8}{8\,G_{10}} = \frac{L^8}{64\pi^2 g_s^2\alpha'^4}, \tag{14.18}$$
and inserting $L^8 = 16\pi^2 g_s^2 N^2\alpha'^4$,
$$a = c = \frac{N^2}{4}. \tag{14.19}$$
This agrees with the exact result (14.7) at leading order in N, *including the coefficient* $1/4$: a classical gravitational computation, an expansion of $\sqrt{\det g}$ near the boundary of AdS_5, has reproduced the number of degrees of freedom of a gauge theory counted by a one-loop triangle. Historically one of the first precision tests of the correspondence.

- *The $1/N^2$ correction.* The mismatch N^2 versus $N^2 - 1$ is not a failure: the classical action scales as $1/G_5 \sim N^2$, and $O(N^0)$ effects come from one-loop corrections in the bulk, here the graviton multiplet. These reproduce the -1 [67] (and [68] for the R-symmetry anomaly): the duality is expected, and verified where checked, to hold order by order in $1/N$.

- *Theories with $a \neq c$* cannot be dual to a weakly curved two-derivative bulk; $a - c$ is encoded in higher-derivative terms, suppressed by $\alpha'/L^2 \sim \lambda^{-1/2}$ in string theory. Many SCFTs have $a - c = O(1)$, subleading at large N; when $a - c \sim N^2$ the dual is irreducibly stringy. The smallness of $|a - c|/a$ is a diagnostic for a gravity-like dual.

- *The two-dimensional analogue* in AdS_3 gives $c = 3L/2G_3$, the Brown–Henneaux central charge [69], met again in Part III.

- *The R-symmetry anomaly: field theory.* The second anomaly test involves the global symmetries. The four Weyl fermions ψ_a of $\mathcal{N} = 4$ SYM transform in the $\mathbf{4}$ of $\mathrm{SU}(4)_R \simeq \mathrm{SO}(6)$ and the six scalars in the $\mathbf{6}$. Being chiral, the fermions generate a cubic 't Hooft anomaly: in background SU(4) gauge fields $A_\mu = A_\mu^a T^a$ the covariant divergence of the R-current is

$$D_\mu J^{\mu\, a} \;\propto\; d^{abc}\, \epsilon^{\mu\nu\rho\sigma} \partial_\mu A_\nu^b \partial_\rho A_\sigma^c + \dots, \qquad d^{abc} = \mathrm{Tr}_{\mathbf{4}}\, T^a\{T^b, T^c\} \times (N^2-1), \tag{14.20}$$

the factor N^2-1 counting the adjoint multiplicity, the dots the non-abelian completion. By Adler–Bardeen the coefficient is one-loop exact: another coupling-independent number the bulk must reproduce.

- *The bulk Chern–Simons term.* By the dictionary the R-currents couple to $\mathrm{SU}(4) \simeq \mathrm{SO}(6)$ gauge fields in AdS_5, the Kaluza–Klein gauge bosons of the S^5 isometries. Their effective action contains, besides the Yang–Mills term, a five-dimensional *Chern–Simons term*; for abelian fields [2]
$$S_{\mathrm{CS}} = -f_{ijk} \int_{\mathrm{AdS}_5} \mathrm{d}^5 x\; \epsilon^{\mu\nu\rho\sigma\tau} A_\mu^i\, F_{\nu\rho}^j\, F_{\sigma\tau}^k, \tag{14.21}$$
which is gauge invariant only up to a total derivative: under $\delta A_\mu^i = \partial_\mu \alpha^i$,
$$\delta S_{\mathrm{CS}} = -f_{ijk} \int_{\mathrm{AdS}_5} \mathrm{d}^5 x\; \epsilon^{\mu\nu\rho\sigma\tau} \partial_\mu (\alpha^i F_{\nu\rho}^j F_{\sigma\tau}^k) = -f_{ijk} \int_{\mathrm{bdy}} \mathrm{d}^4 x\; \epsilon^{\nu\rho\sigma\tau} \alpha^i F_{\nu\rho}^j F_{\sigma\tau}^k. \tag{14.22}$$

- *Anomaly inflow.* The variation is a pure boundary term. But the on-shell bulk action *is* the generating functional $W[A]$ of the boundary theory, so (14.22) says that W shifts under gauge transformations of the background fields by exactly the form of a four-dimensional triangle anomaly, $\partial_\mu \langle J^{\mu\, i}\rangle = f_{ijk}\, \epsilon^{\nu\rho\sigma\tau} F_{\nu\rho}^j F_{\sigma\tau}^k$. The classical bulk theory is perfectly gauge invariant; the anomaly flows from the bulk to the boundary. The non-abelian case works identically with the SU(4) Chern–Simons five-form, whose coefficient is an integer level k multiplying $\frac{1}{24\pi^2}\,\mathrm{Tr}(A\, \mathrm{d}A\, \mathrm{d}A + \dots)$.

- *Where the term and its level come from.* The Chern–Simons term descends from the Kaluza–Klein reduction of Type IIB supergravity on S^5 with N units of five-form flux: the self-dual F_5 has a background value proportional to N on the S^5 and, by self-duality, on AdS_5, and reducing the ten-dimensional coupling cubic in the form fields on this background soaks up two powers of the flux, leaving a five-dimensional term cubic in the SO(6) gauge fields with coefficient $\propto N^2$. Pure five-dimensional reasoning gives the same: gauged supergravity ties the Chern–Simons coefficient to $1/16\pi G_5 \propto N^2$ by supersymmetry, and a careful normalization [60, 65] gives
$$k_{\mathrm{bulk}} = N^2, \qquad \text{versus} \qquad k_{\mathrm{CFT}} = N^2 - 1, \tag{14.23}$$
the exact one-loop anomaly at leading order in N, with the -1 again a one-loop effect in the bulk [68].

- *Division of labour.* On the field-theory side the anomaly is quantum mechanical, a fermion loop; on the gravity side it is classical, a topological term in the Lagrangian. Mixed anomalies involving the R-current and two stress tensors are matched by mixed gauge–gravitational Chern–Simons couplings, and in supersymmetric theories the current anomalies are tied to a and c, so the two tests of this chapter are secretly one.

Chapter 15

Strings and D-Branes: A Crash Course

The next chapter derives the correspondence from D3-branes; this one provides the minimum string-theory background needed to follow that argument, assuming no prior exposure. Nothing here is derived in full. Each fact is stated with the logic that makes it plausible and with a pointer to where it is derived: Tong's lecture notes [8], cited by section, for everything concerning the bosonic string, D-branes and the low-energy action, and Polchinski [9, 10] for superstrings, RR fields and Type II supergravity. What Chapter 16 will use is: the string scale α'; gauge fields from open strings and gravity from closed strings; the string coupling g_s and the genus expansion; the Ramond–Ramond fields of Type IIB and the p-branes charged under them; D-branes, the $\mathrm{U}(N)$ gauge theory on a stack, and $g^2_{\mathrm{YM}} = 4\pi g_s$.

15.1 The relativistic string

A point particle of mass m sweeps out a worldline and its action is $-m$ times the proper length. A string sweeps out a two-dimensional *worldsheet*, described by embedding functions $X^M(\tau, \sigma)$, $M = 0, \ldots, d-1$.

- *Nambu–Goto action.* The action is proportional to the invariant area, computed from the induced metric $\gamma_{ab} = \partial_a X^M \partial_b X^N \eta_{MN}$:

$$S_{\mathrm{NG}} = -T \int \mathrm{d}\tau\, \mathrm{d}\sigma\, \sqrt{-\det \gamma_{ab}}, \qquad T = \frac{1}{2\pi\alpha'}, \qquad \ell_s = \sqrt{\alpha'}. \tag{15.1}$$

  The tension T is the energy per unit length; it is traded for the Regge slope α' and the string length ℓ_s, the one dimensionful parameter of string theory.

- *Open and closed.* An open string is an interval, $\sigma \in [0, \pi]$, whose worldsheet is a strip; a closed string is a circle, whose worldsheet is a tube.

- *Transverse degrees of freedom.* The action is invariant under reparametrizations of the worldsheet, so two of the X^M are pure gauge and the physical degrees of freedom are the $d-2$ transverse displacements.

- *Polyakov action* [8, §1.3]. An auxiliary worldsheet metric h_{ab} removes the square root:

$$S_{\mathrm{P}} = -\frac{1}{4\pi\alpha'} \int \mathrm{d}^2\sigma\, \sqrt{-h}\, h^{ab}\, \partial_a X^M \partial_b X^N\, \eta_{MN}, \tag{15.2}$$

  classically equivalent to (15.1) and quadratic in X^M. It has one more local symmetry, Weyl rescaling of h_{ab}, and the two together allow the gauge $h_{ab} = \eta_{ab}$, in which X^M obeys the free wave equation $(\partial_\tau^2 - \partial_\sigma^2) X^M = 0$ subject to the constraints $T_{ab} = 0$.

- *What quantization means.* Quantizing the string is quantizing this constrained two-dimensional conformal field theory. Preserving Weyl invariance at the quantum level is the source of every consistency condition of string theory, the critical dimension first among them [8, §5.3].

15.2 The spectrum: gauge fields and gravitons

The solution of the wave equation is a sum of harmonics, with amplitudes α_n^M for the open string and two independent towers $\alpha_n^M, \tilde{\alpha}_n^M$ for the closed string, on top of the centre-of-mass position and momentum.

- *States are particles.* Upon quantization the amplitudes become oscillator ladder operators, and a string state is a collection of excited harmonics acting on a ground state $|0; p\rangle$. Each such state behaves as a particle in spacetime, with mass and spin fixed by the oscillators that act: one object, infinitely many particle species.

- *Light-cone gauge* [8, §2.2]. The constraints remove the timelike oscillators, and only the $d-2$ transverse oscillators α_n^i act on physical states.

- *Mass formulas* [8, §§2.2.2, 2.3].

$$\alpha' m^2 = N - a \quad \text{(open)}, \qquad \frac{\alpha' m^2}{4} = N - a = \tilde{N} - a \quad \text{(closed)}, \qquad N = \sum_{n\geq 1}\sum_{i=1}^{d-2} \alpha_{-n}^i \alpha_n^i, \tag{15.3}$$

  where the level N counts harmonics with multiplicity, and $a = -\frac{d-2}{2}\sum_n n = \frac{d-2}{24}$ is the zero-point energy of the transverse oscillators, the divergent sum defined by $\zeta(-1) = -\frac{1}{12}$ as for a Casimir energy.

- *The critical dimension.* The first excited open-string state $\alpha_{-1}^i|0; p\rangle$ is a vector with $d-2$ components, which is the number of polarizations of a *massless* vector, so Lorentz invariance requires $a = 1$, i.e. $d = 26$ for the bosonic string.

- *Low-energy limit.* The mass spacing is $1/\alpha'$, so at energies far below $M_s = 1/\ell_s$ only the massless level survives. This limit is taken repeatedly below.

- *The tachyon.* The ground state has $\alpha' m^2 = -1$: the bosonic string is unstable and will be replaced by the superstring in Section 15.4.

The massless levels dominate low-energy physics.

- *Open string.* The level-one state is a massless vector A_M, whose consistent self-interactions force gauge invariance: the low-energy theory is Maxwell theory, or Yang–Mills theory once open strings carry charges (Section 15.6).

- *Closed string.* The level-one states $\xi_{ij}\alpha_{-1}^i\tilde{\alpha}_{-1}^j|0; p\rangle$ decompose into

$$\underbrace{g_{MN}}_{\text{symmetric traceless}} \qquad \underbrace{B_{MN}}_{\text{antisymmetric}} \qquad \underbrace{\phi}_{\text{trace}}, \tag{15.4}$$

  a massless spin-two field, a two-form gauge potential (the Kalb–Ramond field, under which strings are the charged objects), and a scalar, the dilaton. An interacting massless spin-two particle must couple to a conserved stress tensor, and consistency of its self-coupling forces diffeomorphism invariance: at low energy it is general relativity.

The two golden facts of the string spectrum. The massless level of the *open* string is a gauge field A_M; gauge theory is the low-energy limit of open strings. The massless level of the *closed* string contains a graviton g_{MN}, a two-form B_{MN} and the dilaton ϕ; Einstein gravity is the low-energy limit of closed strings. Two open-string endpoints can always join, so every string theory contains closed strings and is a theory of quantum gravity.

Interactions of strings are smeared over the length ℓ_s, which is why the ultraviolet divergences of perturbative quantum gravity are absent [8, §6.4.4].

15.3 Gravity from strings and the genus expansion

- *Strings in a background.* Replace η_{MN} in (15.2) by $G_{MN}(X)$ and couple the other massless fields:

$$S = -\frac{1}{4\pi\alpha'}\int \mathrm{d}^2\sigma\,\sqrt{-h}\,\Big(h^{ab}\,\partial_a X^M \partial_b X^N\,G_{MN}(X) + \epsilon^{ab}\,\partial_a X^M \partial_b X^N\,B_{MN}(X) - \alpha' R^{(2)}\,\phi(X)\Big). \tag{15.5}$$

  This is an interacting two-dimensional field theory whose couplings are the spacetime fields.

- *Einstein's equations as beta functions* [8, §§7.1.1, 7.2.3]. Weyl invariance requires the beta functions to vanish. To leading order in α'/R_c^2, with R_c the curvature radius [70],

$$\beta^G_{MN} = \alpha'\Big(R_{MN} + 2\nabla_M\nabla_N\phi - \tfrac{1}{4}H_{MPQ}H_N{}^{PQ}\Big) + O(\alpha'^2) = 0, \qquad H = \mathrm{d}B, \tag{15.6}$$

  with analogous equations for B and ϕ: backgrounds on which strings propagate consistently are solutions of gravity, corrected by higher-curvature terms in powers of α'.

- *The low-energy effective action* [8, §7.3]. The conditions $\beta = 0$ are the equations of motion of

$$S_{\text{eff}} = \frac{1}{2\kappa_0^2}\int \mathrm{d}^d x\,\sqrt{-G}\;\mathrm{e}^{-2\phi}\Big(R + 4\,\partial_M\phi\,\partial^M\phi - \tfrac{1}{12}H_{MNP}H^{MNP}\Big) + O(\alpha'), \tag{15.7}$$

  with $2\kappa_0^2 = (2\pi)^7\alpha'^4$ in ten dimensions, fixed by comparison with string amplitudes. The α' corrections matter only when the geometry varies on the scale ℓ_s; this is the origin of the condition $L \gg \ell_s$ met since Chapter 12.

- *The string coupling.* On a Euclidean worldsheet, $\frac{1}{4\pi}\int\sqrt{h}\,R^{(2)}$ is the Euler characteristic $\chi = 2-2g-b$, with g handles and b boundaries, so a constant dilaton weights a worldsheet by $\mathrm{e}^{-\chi\langle\phi\rangle}$. Define

$$g_s \equiv \mathrm{e}^{\langle\phi\rangle}\ : \tag{15.8}$$

  a handle costs g_s^2, a boundary costs g_s. The coupling is not a parameter but the expectation value of a field.

- *The genus expansion.* Closed-string perturbation theory is a sum over topologies with one surface per order (Figure 15.1),

$$\mathcal{A} = \sum_{g=0}^{\infty} g_s^{2g-2}\,F_g(\alpha'). \tag{15.9}$$

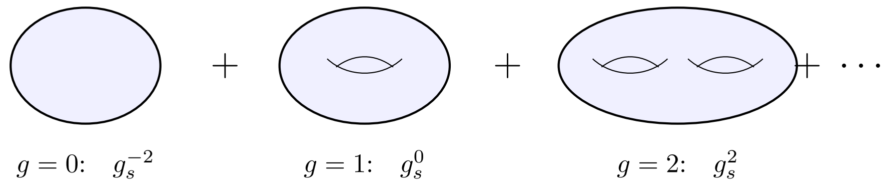


Figure 15.1: Closed-string perturbation theory as a genus expansion (15.9): one surface of genus g, weighted by g_s^{2g-2}, replaces the Feynman diagrams of a field theory at each loop order. This is the topological expansion of large-N gauge theory in Chapter 11, with $g_s^2 \leftrightarrow 1/N^2$.

- *Large N is a string theory.* Comparing with the 't Hooft expansion $\ln Z = \sum_g N^{2-2g} f_g(\lambda)$ of Chapter 11 identifies a large-N gauge theory as a weakly coupled string theory with

$$g_s \;\propto\; \frac{1}{N} \quad \text{at fixed } \lambda, \tag{15.10}$$

  the constant being fixed for D3-branes in Section 15.7. Which string theory it is took twenty-three years to answer.

15.4 Superstrings, RR fields, and Type IIB supergravity

The bosonic string has a tachyon and no spacetime fermions. Both are cured by the *superstring*, which adds to the X^M an equal number of worldsheet fermions ψ^M related to them by worldsheet supersymmetry [8, §2.5], [10].

- The critical dimension becomes $d = 10$.
- Worldsheet fermions admit antiperiodic or periodic boundary conditions, defining the Neveu–Schwarz (NS) sector, whose states are spacetime bosons, and the Ramond (R) sector, whose ground state is a spacetime spinor and gives spacetime fermions.
- Consistency (modular invariance) imposes the GSO projection, which removes the tachyon and leaves equal numbers of bosons and fermions at every level: the theory is supersymmetric in spacetime.
- *Supersymmetry and supercharges.* Spacetime supersymmetry is a symmetry whose generators, the *supercharges* Q, are fermionic: they carry a spinor index and obey anticommutation relations, the essential one being

$$\{Q, \bar{Q}\} \;\sim\; \gamma^M P_M, \tag{15.11}$$

  so that two supersymmetry transformations compose to a translation. A supercharge maps bosons to fermions of the same mass, so states organize into *supermultiplets* with equal numbers of bosonic and fermionic states. The amount of supersymmetry is counted by the number of real components of the supercharges. A Majorana–Weyl spinor in ten dimensions has 16 real components, and the Type II theories have two such spinors, hence 32 supercharges. Reduced to four dimensions, where a Weyl spinor Q_α, $\alpha = 1, 2$, has 4 real components, 16 supercharges are four Weyl spinors Q^a_α, $a = 1, \ldots, 4$, together with their conjugates $\bar{Q}_{a\dot{\alpha}}$: this is what $\mathcal{N} = 4$ supersymmetry means, and it is the supersymmetry of the D3-brane worldvolume theory of Chapter 16. We use only this counting and the BPS bound below; for the algebra and its representations see [10].

- Assembling left- and right-movers consistently allows exactly five theories in ten dimensions. Only the closed-string theories of Type IIA and Type IIB, with 32 supercharges, concern us, and chiefly IIB.

The massless bosonic fields of the Type II theories come in two families, according to the boundary conditions on the two sides of the closed string.

- *NS–NS sector*: the trio (15.4). Everything in Section 15.3 carries over.
- *R–R sector*: antisymmetric tensor gauge potentials C_p, with field strengths $F_{p+1} = \mathrm{d}C_p$ and gauge invariance $C_p \to C_p + \mathrm{d}\Lambda_{p-1}$,

$$\text{IIA:} \quad C_1,\ C_3; \qquad\qquad \text{IIB:} \quad C_0,\ C_2,\ C_4 \ \ (\text{with } F_5 = *F_5). \tag{15.12}$$

- *Who is charged.* Just as a point particle couples to A_1 by $q\int A_1$ along its worldline, the object charged under C_{p+1} has a $(p+1)$-dimensional worldvolume, a *p-brane*, with coupling $\mu_p \int C_{p+1}$. No perturbative string state carries RR charge, so the RR charges must be carried by nonperturbative objects.
- *Type II supergravity.* At energies below M_s each Type II theory reduces to the unique two-derivative theory with 32 supercharges, whose NS–NS part is (15.7) and whose RR part contains kinetic terms for the C_p, without the factor $\mathrm{e}^{-2\phi}$, and a Chern–Simons term $\int C_4 \wedge H_3 \wedge F_3$ [10].
- *Newton's constant.* Since ϕ enters the Einstein term only through $\mathrm{e}^{-2\phi}$, the physical gravitational coupling is $\kappa^2 = \kappa_0^2 g_s^2$:

The ten-dimensional Newton constant of Type II string theory is

$$16\pi G_{10} = 2\kappa^2 = (2\pi)^7\, g_s^2\, \alpha'^4,$$

the equation promised in eqn. (14.17). The Planck length $\ell_P^{(10)} \sim g_s^{1/4}\ell_s$ is parametrically smaller than the string length at weak coupling.

- *Validity.* Supergravity is trustworthy when both expansions are under control: weak coupling, $g_s \ll 1$, and weak curvature, $R_c \gg \ell_s$. This double condition becomes the validity regime of the gravity side of AdS/CFT in Chapter 16.

15.5 p-branes: black solutions with RR charge

Heavy extended objects in gravity are described by their backreacted geometry [10].

- *Source and charge.* Type II supergravity admits solutions sourced by an object filling p spatial directions with worldvolume action

$$S_{p\text{-brane}} = -\tau_p \int \mathrm{d}^{p+1}x\,\sqrt{-\det g_{ab}} \;+\; \mu_p \int C_{p+1}, \tag{15.13}$$

whose charge is the flux $\int_{S^{8-p}} *F_{p+2} = N \in \mathbb{Z}$ through the sphere surrounding it in the transverse $\mathbb{R}^{9-p}$.

- *The BPS bound.* As for Reissner–Nordström there is a bound,

$$\tau \;\geq\; \frac{N}{(2\pi)^p\, g_s\, \alpha'^{(p+1)/2}}, \tag{15.14}$$

below which the singularity is naked. In a supersymmetric theory it is the BPS bound: in the presence of a brane the algebra (15.11) acquires a central term proportional to the charge N, $\{Q,\bar{Q}\} \sim \gamma^M P_M + Z$, and positivity of the norm of $Q|\text{state}\rangle$ gives mass $\geq |Z|$. Configurations saturating it are annihilated by half of the supercharges, preserving 16 of the 32, and their tension–charge relation is protected against all corrections. Note the tension $\propto N/g_s$, between a perturbative string state (g_s^0) and a conventional soliton ($1/g_s^2$).

- *The extremal solutions* [71]. With worldvolume coordinates x^μ and transverse y^i, $r^2 = y^i y^i$,

$$\mathrm{d}s^2 = H^{-1/2}\,\eta_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu + H^{1/2}\,\mathrm{d}y^i \mathrm{d}y^i, \qquad \mathrm{e}^\phi = g_s\,H^{(3-p)/4}, \qquad C_{01\cdots p} = H^{-1} - 1, \tag{15.15}$$

  with $H = 1 + c_p g_s N \alpha'^{(7-p)/2}/r^{7-p}$ a harmonic function on $\mathbb{R}^{9-p}$, $c_p = (2\sqrt{\pi})^{5-p}\Gamma(\frac{7-p}{2})$.

- *No force and moduli space.* Since the Laplace equation is linear,

$$H \;=\; 1 + c_p\,g_s\,\alpha'^{(7-p)/2} \sum_a \frac{N_a}{|\vec{y} - \vec{y}_a|^{7-p}} \tag{15.16}$$

  is also a solution, for any positions $\vec{y}_a$ of the centres: parallel extremal branes exert no force on one another, their gravitational and dilaton attraction being cancelled by RR repulsion because tension equals charge. The positions are therefore not fixed by the equations of motion but are free parameters of the solution, all with the same energy, and the space they parametrize, N unordered points in $\mathbb{R}^{9-p}$, is the *moduli space* of the configuration. In Section 16.2.1 the same space reappears as the flat directions of the gauge theory on the branes.

- *The D3 case.* For $p = 3$ the dilaton is constant and the horizon $r = 0$ is regular, at infinite proper distance down a smooth throat:

$$\mathrm{d}s^2 = H^{-1/2}(-\mathrm{d}t^2 + \mathrm{d}\vec{x}^2) + H^{1/2}(\mathrm{d}r^2 + r^2\,\mathrm{d}\Omega_5^2), \qquad H = 1 + \frac{L^4}{r^4}, \quad L^4 = 4\pi g_s N \alpha'^2, \tag{15.17}$$

  the geometry whose throat Chapter 16 shows to be $\mathrm{AdS}_5 \times S^5$.

What these objects are in string theory is the question answered next.

15.6 D-branes

- *Boundary conditions.* Varying the Polyakov action for an open string leaves a boundary term $-\frac{1}{2\pi\alpha'}\int \mathrm{d}\tau[\partial_\sigma X_M \delta X^M]_0^\pi$, which vanishes for either of two conditions at each endpoint and direction:

$$\underbrace{\partial_\sigma X^M = 0}_{\text{Neumann}} \qquad \text{or} \qquad \underbrace{\delta X^M = 0 \;\; (X^M \text{ fixed})}_{\text{Dirichlet}}. \tag{15.18}$$

- *Definition.* Impose Neumann conditions along $x^0, \ldots, x^p$ and Dirichlet conditions $X^i = y^i$ in the remaining $9-p$ directions: every open-string endpoint is confined to the hyperplane $\{X^i = y^i\}$, a *Dp-brane.* The momentum flowing off the string ends is absorbed by the brane, which is therefore a dynamical object with a tension and fluctuations.

- *They are unavoidable* [8, §8.3.2]. On a circle of radius R the string spectrum is invariant under $R \to \alpha'/R$ with momentum and winding exchanged (T-duality), and this exchanges Neumann and Dirichlet conditions.

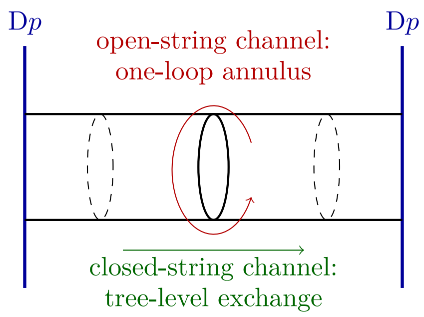


Figure 15.2: The annulus amplitude between two parallel D-branes, read in two channels. With worldsheet time running *around* the cylinder (red): a one-loop vacuum amplitude of the open strings stretched between the branes. With time running *along* the cylinder (green): tree-level exchange of a single closed string. Equating the two readings of one and the same amplitude determines which closed-string fields the brane sources, i.e. its tension and RR charge.

- *Polchinski's identification* [72]. The one-loop amplitude of open strings stretched between two parallel D-branes is an annulus. Read with worldsheet time running along the annulus instead of around it, the same diagram is the tree-level exchange of a single closed string between the branes (Figure 15.2). Equating the two readings shows which closed-string fields the brane sources and how strongly [8, §6.3.2].

> **D-branes are the RR-charged objects** [72]. The Dp-brane of Type II string theory carries one unit of charge under C_{p+1} and tension
>
> $$\tau_p = \frac{1}{(2\pi)^p\, g_s\, \alpha'^{(p+1)/2}},$$
>
> saturating the BPS bound (15.14). The extremal black p-brane (15.15) with flux N is the long-distance description of N coincident Dp-branes.

- *Why it all fits.* The $1/g_s$ in the tension is the weight of a disc, the worldsheet with one boundary, $\chi = 1$. It also explains why no perturbative string state carries RR charge: a D-brane has mass proportional to $1/g_s$, which diverges in the limit $g_s \to 0$ around which perturbation theory expands, so it never appears in the perturbative spectrum, whose masses are of order g_s^0, just as the 't Hooft–Polyakov monopole, of mass $\sim 1/g^2$, is absent from the perturbative spectrum of a gauge theory. And because D-branes are BPS the identification holds at weak and at strong coupling, which is what allows Chapter 16 to equate two descriptions of one stack.

The gauge theory on a stack.

- *One brane* [8, §§3.1.2, 7.7.1]. The massless open strings on a single Dp-brane give a gauge field A_a along the brane, $9-p$ scalars Φ^i from the transverse oscillators, and fermions completing the vector multiplet with 16 supercharges. The scalars are the collective coordinates of the brane: a vacuum expectation value $\langle \Phi^i \rangle$ is a transverse displacement,

$$y^i = 2\pi\alpha'\, \Phi^i. \tag{15.19}$$

- *N branes.* An open string carries two Chan–Paton labels, the branes on which it starts and ends, so the fields become $N \times N$ matrices (Figure 15.3). A string stretched between

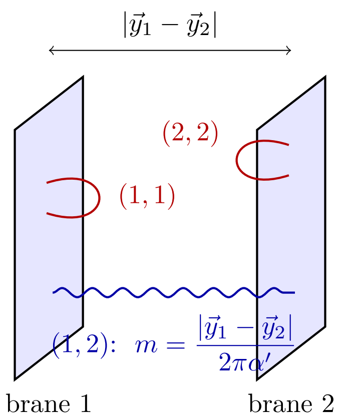


Figure 15.3: Open strings on a pair of parallel D-branes. Strings with both ends on the same brane give massless U(1) vector multiplets; strings stretched between different branes give vector multiplets of mass (15.20). When the branes coincide the gauge symmetry is enhanced to U(2); for N branes the N^2 sectors fill the adjoint of U(N).

branes at separation $|\vec{y}_m - \vec{y}_n|$ has minimal energy tension times length, so its lightest states have mass

$$m_{(mn)} = \frac{|\vec{y}_m - \vec{y}_n|}{2\pi\alpha'}. \tag{15.20}$$

- *Gauge group.* For coincident branes all N^2 vectors are massless and the worldvolume theory is a U(N) gauge theory with all fields in the adjoint; for separated branes the off-diagonal vectors are massive with the masses (15.20). This is the Higgs mechanism of $\mathrm{U}(N) \to \mathrm{U}(1)^N$ by an adjoint scalar VEV, and Section 16.2.1 verifies that the gauge theory reproduces (15.20) with the dictionary (15.19).

- *The Lagrangian* [8, §7.7]. The two-derivative theory on N Dp-branes is fixed by its 16 supercharges: maximally supersymmetric Yang–Mills in $p+1$ dimensions, the dimensional reduction of ten-dimensional super Yang–Mills. For $p = 3$ it is $\mathcal{N} = 4$ super Yang–Mills with gauge group U(N), with the SO(6) rotations of the transverse $\mathbb{R}^6$ as its R-symmetry. Uniquely among Dp-branes its coupling does not run, just as, uniquely among black p-branes, its dilaton is constant.

15.7 The D-brane action and $g^2_{\mathrm{YM}} = 4\pi g_s$

- *The action* [8, §§7.5, 7.6]. The effective action for the massless fields of a single Dp-brane in a closed-string background is

$$S_{\mathrm{D}p} = \underbrace{-\,T_p \int \mathrm{d}^{p+1}\xi \; \mathrm{e}^{-\phi} \sqrt{-\det\left(g_{ab} + B_{ab} + 2\pi\alpha'\, F_{ab}\right)}}_{\text{Dirac–Born–Infeld}} \;+\; \underbrace{\mu_p \int_{\mathrm{wv}} \Big(\sum_k C_k\Big) \wedge \mathrm{e}^{\,B + 2\pi\alpha' F}}_{\text{Wess–Zumino}}, \tag{15.21}$$

with $T_p = \mu_p = (2\pi)^{-p}\alpha'^{-(p+1)/2}$, so that the physical tension T_p/g_s is the one in the keybox above; g_{ab} and B_{ab} are pullbacks to the worldvolume and F_{ab} is the worldvolume field strength.

- *What the two terms are.* The Dirac–Born–Infeld term is the Nambu–Goto idea one level up, the area of the brane corrected by the gauge field it carries; expanding it in static gauge produces the scalar kinetic terms with the normalization (15.19). The Wess–Zumino term contains the RR coupling $\mu_p \int C_{p+1}$ and, on a D3-brane, the term $\mu_3(2\pi\alpha')^2 \frac{1}{2} \int C_0 F \wedge F$, which identifies the axion C_0 as the theta angle.

- *The Yang–Mills coupling.* In a flat background, using $\sqrt{-\det(\eta + M)} = 1 + \frac{1}{4} \operatorname{Tr} M^2 + O(M^4)$ for antisymmetric $M_{ab} = 2\pi\alpha' F_{ab}$,

$$\mathcal{L} = -\frac{T_p\,(2\pi\alpha')^2}{4\,g_s} \operatorname{Tr} F_{ab}F^{ab} \equiv -\frac{1}{2g_{\rm YM}^2} \operatorname{Tr} F_{ab}F^{ab}, \qquad g_{\rm YM}^2 = 2\,(2\pi)^{p-2}\, g_s\, \alpha'^{(p-3)/2}, \tag{15.22}$$

  which has the mass dimension $3 - p$ required in $p + 1$ dimensions.

- *The D3 case.* For $p = 3$ the α' dependence cancels:

$$\boxed{g_{\rm YM}^2 = 4\pi g_s\ , \qquad \theta = 2\pi C_0\ ,} \qquad \text{i.e.} \qquad \frac{L^4}{\alpha'^2} = 4\pi g_s N = g_{\rm YM}^2 N = \lambda. \tag{15.23}$$

  The coupling of the D3-brane gauge theory is the string coupling, its theta angle is the RR axion, and the gravitational size of the stack in string units is the 't Hooft coupling. This is the origin of $g_s = \lambda/4\pi N$ in (15.10), of the $\sqrt{\lambda}$ in the Wilson loops of Chapter 17, and of the Newton constant that fed the anomaly match of Chapter 14.

15.8 The two faces of a D3-brane stack

We have two descriptions of N coincident D3-branes in flat ten-dimensional Type IIB string theory. Open-string perturbation theory is an expansion in $g_s N$, each extra boundary costing g_s and gaining N; supergravity requires $L \gg \ell_s$, i.e. $g_s N \gg 1$.

The two faces of N coincident D3-branes.

Gauge-theory face (open strings)	*Gravity face (closed strings)*
hyperplane where open strings end	extremal black 3-brane sourcing g_{MN}, C_4 [71, 72]
U(N) $\mathcal{N} = 4$ SYM: A_μ, six adjoint scalars Φ^i, four adjoint Weyl fermions	$\mathrm{d}s^2 = H^{-1/2}\mathrm{d}x_\parallel^2 + H^{1/2}\mathrm{d}y^2$, $H = 1 + L^4/r^4$
coupling $g_{\rm YM}^2 = 4\pi g_s$, $\lambda = g_{\rm YM}^2 N$	size $L^4 = 4\pi g_s N \alpha'^2 = \lambda\,\alpha'^2$
scalar VEVs = brane positions: $\vec{y} = 2\pi\alpha'\,\vec{\phi}$	multi-centre harmonic H; no force between branes
16 supercharges (BPS)	16 supercharges (extremal)
valid for $\lambda = 4\pi g_s N \ll 1$	valid for $\lambda = 4\pi g_s N \gg 1$

One object, two descriptions, complementary regimes of validity. Chapter 16 takes the low-energy limit of both at once: the gauge theory decouples from the bulk, the geometry develops an $\mathrm{AdS}_5 \times S^5$ throat, and the two faces become the two sides of the AdS/CFT correspondence.

Chapter 16

D3-Branes and the Maldacena Limit

Everything is now in place for the central construction of the course. Maldacena's logic [58] is to view one and the same object, a stack of N coincident D3-branes, through two pairs of glasses: as the carrier of $\mathcal{N} = 4$ super Yang–Mills with gauge group $\mathrm{SU}(N)$, and as the extremal black 3-brane of Type IIB supergravity. A common low-energy limit, the decoupling or near-horizon limit, strips both descriptions to their interacting cores, and declaring the cores equal is the conjecture: $\mathcal{N} = 4$ SYM *is* Type IIB string theory on $\mathrm{AdS}_5 \times S^5$. Along the way we derive the parameter dictionary, including the relations $L^4 = 4\pi g_s N\alpha'^2$ and $L^3/G_5 = 2N^2/\pi$ used on faith so far, and map out when each description is weakly coupled, the origin of the strong/weak character of the duality.

16.1 One object, two descriptions

Recall from Chapter 15 the two characterizations of D-branes (Figure 16.1).

- *Perturbative string theory.* A Dp-brane is a $(p + 1)$-dimensional hyperplane on which open strings can end [72]; spacetime remains flat. Open strings, glued to the brane, have massless modes that are fields living *on* the brane; closed strings propagate in the bulk and their massless modes are the supergravity fields; open–closed interactions are controlled by g_s.

- *Supergravity.* The same object is a solitonic solution, a black p-brane [71], sourcing the metric, the dilaton and the appropriate Ramond–Ramond form. There are no open strings; the brane is geometry plus flux, and everything is closed strings in this curved background.

Neither description is more fundamental; they are useful in different regimes (Section 16.4), and the engine of the correspondence is that a single low-energy limit, applied to both, produces two theories that must describe the same physics.

16.2 The worldvolume theory: $\mathcal{N} = 4$ super Yang–Mills

- *Field content from the open-string spectrum.* A single D3-brane along $x^0, \ldots, x^3$, at a point in the six transverse directions $y^1, \ldots, y^6$, carries at the massless level (Chapter 15) a gauge field A_μ, the components of the open-string vector tangent to the brane; six real scalars ϕ^i, the transverse components, geometrically the fluctuations of the brane's position; and four Weyl fermions ψ_a, $a = 1, \ldots, 4$, required by the sixteen supersymmetries the brane preserves. Massive open-string excitations, $m^2 \sim 1/\alpha'$, play no role far below the string scale. That this is the right multiplet needs no string computation: an extremal D3-brane is BPS, preserving 16 of the 32 supercharges of Type IIB, so its worldvolume

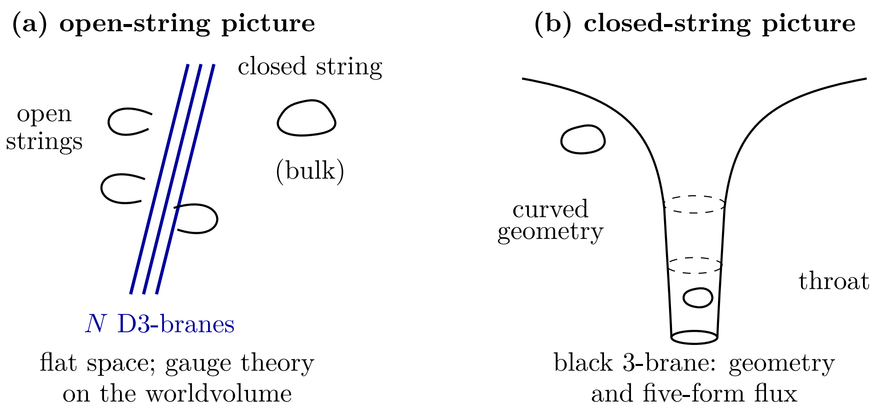


Figure 16.1: Two descriptions of the same stack of N D3-branes. (a) Hyperplanes in flat space on which open strings end; the massless open-string modes form a gauge theory on the worldvolume, and closed strings propagate in the flat bulk. (b) A solitonic solution of Type IIB supergravity: the branes have been replaced by their gravitational and Ramond–Ramond field, and only closed strings remain, moving in a curved geometry with a deepening throat.

theory is an $\mathcal{N}=4$ theory, and the unique $\mathcal{N}=4$ multiplet with spin at most one, the vector multiplet, contains exactly one vector, six scalars and four Weyl fermions. The six is forced to match the number of transverse directions.

- *N branes.* Open strings from brane p to brane q give each field two indices, assembling into $N \times N$ hermitian matrices. For separated branes a stretched string has minimal mass

$$m_{pq} = \frac{|\vec{y}_p - \vec{y}_q|}{2\pi\alpha'}, \tag{16.1}$$

tension times length. When the branes coincide all these states become massless and the N^2 vectors fill the adjoint of $\mathrm{U}(N)$; the overall $\mathrm{U}(1)$ is the free centre-of-mass motion and decouples, and the interacting $\mathrm{SU}(N)$ theory is what enters the duality.

- *The Lagrangian.* The two-derivative action is fixed by the 16 supercharges: $\mathcal{N}=4$ SYM, the reduction of ten-dimensional SYM on a six-torus,

$$\begin{aligned} S = \frac{1}{g_{\mathrm{YM}}^2} \int \mathrm{d}^4x \ \mathrm{Tr}\Big(&- \tfrac{1}{2}F_{\mu\nu}F^{\mu\nu} - D_\mu\phi^i D^\mu\phi^i - \mathrm{i}\,\bar{\psi}^a \not{D}\psi_a \\ &+ \tfrac{1}{2}\sum_{i,j} \big[\phi^i,\phi^j\big]^2 + C_i^{ab}\,\bar{\psi}_a[\phi^i,\psi_b]\Big) + \frac{\theta}{8\pi^2}\int \mathrm{Tr}\ F\wedge F, \end{aligned} \tag{16.2}$$

with $D_\mu = \partial_\mu + \mathrm{i}[A_\mu,\cdot\,]$ and fixed Clebsch–Gordan coefficients C_i^{ab}: a *single* coupling g_{YM} and the angle θ, set by the closed-string background through the Dirac–Born–Infeld expansion of Chapter 15,

$$g_{\mathrm{YM}}^2 = 4\pi g_s, \qquad\qquad \theta = 2\pi C_0. \tag{16.3}$$

- *R-symmetry.* The six scalars enter symmetrically, so the Lagrangian is invariant under $\mathrm{SO}(6)$ rotations $\phi^i \to R^i{}_j\phi^j$ with the covering group $\mathrm{SU}(4) \simeq \mathrm{Spin}(6)$ acting on the fermions in the $\mathbf{4}$ (scalars in the $\mathbf{6}$, gauge field a singlet). This $\mathrm{SU}(4)_R$ rotates the supercharges into one another, and its geometric meaning is worth remembering:

$$\mathrm{SU}(4)_R \simeq \mathrm{SO}(6) \ = \ \text{rotations of the six directions transverse to the D3-branes}, \tag{16.4}$$

a global symmetry of the gauge theory that is a spacetime symmetry of the brane configuration, the key to matching symmetries in Chapter 17.

- *Vanishing beta function.* The property that makes the story work: *the beta function of $\mathcal{N}=4$ SYM vanishes identically, and the theory is an exact interacting CFT for every $g_{\rm YM}$ and N.* At one loop, for Weyl fermions in representations R_f and real scalars in R_s,
$$\beta(g) = -\frac{g^3}{16\pi^2}\,b_0, \qquad b_0 = \frac{11}{3}\,C_2(G) - \frac{2}{3}\sum_f T(R_f) - \frac{1}{6}\sum_s T(R_s), \tag{16.5}$$
and with all matter adjoint, $C_2(\mathrm{SU}(N)) = T(\mathrm{adj}) = N$,
$$b_0 = \frac{11}{3}N - \frac{2}{3}\cdot 4N - \frac{1}{6}\cdot 6N = \frac{11-8-3}{3}\,N = 0. \tag{16.6}$$
Two- and three-loop coefficients vanish by direct computation, and several arguments establish the vanishing to all orders and non-perturbatively. So $g_{\rm YM}$ is *exactly marginal* and $\mathcal{N}=4$ SYM is a two-parameter family of CFTs labelled by $\tau = \frac{\theta}{2\pi} + \frac{4\pi \mathrm{i}}{g^2_{\rm YM}}$.

- *Superconformal algebra.* Just as $\mathfrak{so}(2,4)$ contains K_μ beyond translations, closing the algebra of the 16 Q^a_α with the K_μ requires 16 conformal supercharges S^α_a: the symmetry algebra of $\mathcal{N}=4$ SYM has 32 fermionic generators and bosonic part $\mathrm{SO}(2,4)\times\mathrm{SU}(4)_R$, the superconformal algebra $\mathfrak{su}(2,2|4)$, to be matched in Chapter 17 against the 32 supersymmetries of $\mathrm{AdS}_5\times S^5$.

16.2.1 The Coulomb branch: vacua are brane positions

- *Moduli space.* The potential $V \propto -\operatorname{Tr}\sum_{i<j}[\phi^i,\phi^j]^2 \geq 0$ vanishes when the six matrices commute and can be simultaneously diagonalized,
$$\phi^i = \mathrm{diag}\left(\phi^i_1, \phi^i_2, \ldots, \phi^i_N\right), \qquad i = 1,\ldots,6, \tag{16.7}$$
so the vacua are parametrized by N points $\vec\phi_p \in \mathbb{R}^6$ up to permutations, the *Coulomb branch*. Generically the gauge symmetry is broken to $\mathrm{U}(1)^{N-1}$ and the off-diagonal components acquire masses $\propto |\vec\phi_p - \vec\phi_q|$ by the Higgs mechanism.

- *Brane picture.* The transverse scalars are the brane positions, $\vec y_p = 2\pi\alpha'\,\vec\phi_p$; separating branes moves along the Coulomb branch, and the Higgsed W bosons are the strings stretched between branes, with (16.1) matching the Higgs mass: vacua of the gauge theory $\leftrightarrow$ arrangements of the branes, exactly, including the masses above each vacuum. The no-force condition between BPS branes of Chapter 15 is the spacetime face of the exact flatness of the Coulomb branch, guaranteed by the 16 supercharges.

- *Seed of the limit.* A scalar expectation value is an energy, a brane position a length, the conversion factor $1/2\pi\alpha'$. Holding $|\vec\phi|$ fixed while $\alpha' \to 0$ forces $|\vec y| \to 0$: keeping field-theory quantities finite zooms onto the branes (Section 16.5).

16.3 The gravitational description: the extremal black 3-brane

A stack of N D3-branes is a heavy object of tension $N\tau_3$, charged under the Ramond–Ramond four-form C_4, and curves spacetime. The solution of Type IIB supergravity is the *extremal black 3-brane* [71], the $p=3$ case of Chapter 15:

$$\boxed{\begin{aligned} &\mathrm{d}s^2 = H(r)^{-1/2}\,\eta_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu + H(r)^{1/2}\left(\mathrm{d}r^2 + r^2\mathrm{d}\Omega_5^2\right),\\ &F_5 = (1+\star)\,\mathrm{d}H^{-1}\wedge\mathrm{d}x^0\wedge\mathrm{d}x^1\wedge\mathrm{d}x^2\wedge\mathrm{d}x^3,\\ &\mathrm{e}^\phi = g_s \ \ (\text{constant}), \qquad\qquad H(r) = 1 + \frac{L^4}{r^4}.\end{aligned}} \tag{16.8}$$

Here r is the radial distance in the transverse $\mathbb{R}^6$, $\star$ the ten-dimensional Hodge dual enforcing $F_5 = \star F_5$, and H is harmonic in $\mathbb{R}^6$ like the potential of a point charge, the constant 1 making the solution asymptotically flat.

- *The dilaton is constant.* For a general extremal p-brane $\mathrm{e}^{\phi} = g_s H^{(3-p)/4}$, r-independent only for $p = 3$: string perturbation theory never breaks down on account of the dilaton. Via (16.3) this is the geometric counterpart of the vanishing beta function: the coupling does not run, the dilaton does not flow.

- *The solution is regular.* For $p \neq 3$ the horizon collapses onto a singular source; for $p = 3$ the surface $r = 0$ is at infinite proper distance down a smooth throat, where the geometry approaches $\mathrm{AdS}_5 \times S^5$.

- *BPS.* The solution preserves 16 supercharges and saturates the BPS bound, tension equal to charge density. This underlies the no-force property and the multi-centre solutions $H = 1 + \sum_a L_a^4/|\vec{y} - \vec{y}_a|^4$, the supergravity images of the Coulomb branch (16.7).

16.3.1 Fixing L: flux quantization

The parameter L must be fixed by N. Two statements do it, the geometry must carry N units of five-form flux and the mass of N D3-branes, and their agreement is the BPS property at work.

1. *The flux through the five-sphere.* In the orthonormal frame

$$e^{\hat{\mu}} = H^{-1/4}\,\mathrm{d}x^{\mu}, \qquad e^{\hat{r}} = H^{1/4}\,\mathrm{d}r, \qquad e^{\hat{\alpha}} = H^{1/4}\,r\,\hat{e}^{\alpha}, \tag{16.9}$$

with $\hat{e}^{\alpha}$ a frame on the unit S^5, the electric piece reads $\mathrm{d}H^{-1} \wedge \mathrm{d}^4x = -H^{-5/4}H'\,e^{\hat{r}} \wedge e^{\hat{0}} \wedge e^{\hat{1}} \wedge e^{\hat{2}} \wedge e^{\hat{3}}$, and its dual is the same coefficient times the sphere directions,

$$\star(\mathrm{d}H^{-1} \wedge \mathrm{d}^4x) = -H'\,r^5\,\omega_5 = 4L^4\,\omega_5, \tag{16.10}$$

with ω_5 the volume form of the unit sphere and $H' = -4L^4/r^5$: a constant, the same flux through every sphere as Gauss's law demands. Integrating, with $\mathrm{Vol}(S^5) = \pi^3$,

$$\int_{S^5} F_5 = 4L^4\ \mathrm{Vol}(S^5) = 4\pi^3 L^4. \tag{16.11}$$

Dirac quantization for Ramond–Ramond fields forces this to be N times the flux of a single D3-brane, so $L^4 \propto N$ with a constant we now pin down by the second, entirely low-energy method.

2. *Matching the tension.* At large r,

$$g_{\mu\nu} \simeq \Big(1 - \frac{L^4}{2r^4}\Big)\,\eta_{\mu\nu}, \qquad g_{mn} \simeq \Big(1 + \frac{L^4}{2r^4}\Big)\,\delta_{mn}, \tag{16.12}$$

(m, n transverse), and linearized gravity relates the tails to the source. Model the stack as

$$T_{\mu\nu} = -\tau\,\eta_{\mu\nu}\,\delta^6(\vec{y}), \qquad T_{mn} = 0, \qquad \tau = N\tau_3, \tag{16.13}$$

energy density $T_{00} = +\tau\delta^6$ and worldvolume pressure equal and opposite to the tension, characteristic of a BPS brane. The linearized equations in harmonic gauge, $\nabla_{\perp}^2 h_{MN} = -16\pi G_{10}(T_{MN} - \frac{1}{8}\eta_{MN}T)$ with $T = -4\tau\delta^6$, have trace-reversed source $-\frac{\tau}{2}\eta_{\mu\nu}\delta^6$ along the worldvolume and $+\frac{\tau}{2}\delta_{mn}\delta^6$ transverse; with the Green's function $\nabla^2(\frac{1}{4\Omega_5 r^4}) = -\delta^6(\vec{y})$, $\Omega_5 = \pi^3$,

$$h_{\mu\nu} = -\,\frac{16\pi G_{10}\,\tau}{2}\,\frac{1}{4\pi^3 r^4}\,\eta_{\mu\nu} = -\,\frac{2G_{10}\,\tau}{\pi^2 r^4}\,\eta_{\mu\nu}, \qquad h_{mn} = +\,\frac{2G_{10}\,\tau}{\pi^2 r^4}\,\delta_{mn}. \tag{16.14}$$

Both components match (16.12), a check that the source really is a BPS brane, provided

$$\frac{L^4}{2} = \frac{2G_{10}\, N\tau_3}{\pi^2}. \tag{16.15}$$

3. *The result.* Insert the two string-theory inputs of Chapter 15 [72],

$$16\pi G_{10} = (2\pi)^7 g_s^2\, \alpha'^4 \implies G_{10} = 8\pi^6 g_s^2 \alpha'^4, \qquad \tau_3 = \frac{1}{(2\pi)^3\, g_s\, \alpha'^2}, \tag{16.16}$$

so that $L^4 = \frac{4}{\pi^2} \cdot 8\pi^6 g_s^2 \alpha'^4 \cdot \frac{N}{8\pi^3 g_s \alpha'^2}$, i.e.

$$\boxed{L^4 = 4\pi\, g_s\, N\, \alpha'^2} \tag{16.17}$$

the relation quoted on faith in Chapter 12. In these units the flux (16.11) is $16\pi^4 g_s N \alpha'^2$, exactly N times the single-brane quantum: charge and mass give the same L, as BPS saturation guarantees.

16.4 Which description, when?

- *Gravity side.* The solution deviates from flat space for $r \lesssim L$, with curvature $\sim 1/L^2$. Classical supergravity is the leading term of the double expansion of Chapter 15, in α' and in g_s. The first requires

$$\frac{L^2}{\alpha'} = \sqrt{4\pi g_s N} = \sqrt{g_{\mathrm{YM}}^2 N} = \sqrt{\lambda} \gg 1, \tag{16.18}$$

  by (16.17) and (16.3): the geometry is weakly curved precisely when the 't Hooft coupling is *large*. The second, $g_s \ll 1$, is arranged at fixed λ by taking N large.

- *Open-string side.* Each additional open-string loop can end on any of the N branes, so the loop-counting parameter is $g_s N \sim \lambda$, the planar counting of Chapter 11, and perturbation theory is useful when

$$g_s N \sim \lambda \ll 1. \tag{16.19}$$

  Geometrically, then $L \ll \ell_s$: the backreaction is confined to a sub-stringy neighbourhood and the branes are rigid hyperplanes in flat space.

- *Maldacena's insight.* The two regimes are complementary and non-overlapping: *the geometry is reliable exactly where perturbation theory is not, and vice versa.* This looks like an obstruction to ever comparing them; Maldacena turned it into the content of a conjecture. If a single object admits both descriptions, and a common limit reduces each to a well-defined interacting theory, those two theories must be *equal*, as full theories at all couplings, with (16.18)–(16.19) telling us which side is computable where.

16.5 The Maldacena decoupling limit

The limit isolates, on both sides, the low-energy dynamics intrinsic to the branes, decoupled from gravity, massive string modes and the asymptotic region [58]:

$$\alpha' \to 0, \qquad \text{with} \quad u \equiv \frac{r}{\alpha'}, \;\; g_s, \;\; N \quad \text{held fixed.} \tag{16.20}$$

The quantity u is an energy: $u/2\pi$ is the mass (16.1) of a string from a probe brane at radius r to the stack, a Higgs scale. Holding u fixed holds *field-theory energies* fixed while $M_s = 1/\sqrt{\alpha'} \to \infty$, zooming onto $r \to 0$; since g_s and N are fixed, so are g_{YM} and λ.

- *Open-string description.* The low-energy world is the massless open-string modes on the branes ($\mathcal{N} = 4$ SYM), the massless closed-string modes in the bulk (ten-dimensional supergravity) and their mutual interactions. Every brane–bulk interaction and every gravitational self-interaction is weighted by positive powers of $\kappa_{10} \sim \sqrt{G_{10}} \sim g_s\alpha'^2 \to 0$; the Dirac–Born–Infeld corrections to Yang–Mills, such as $\alpha'^2 \operatorname{Tr} F^4$, carry positive powers of α' at fixed energy and vanish; massive string states decouple. What survives is

$$\text{(open-string side)} \qquad \mathcal{N} = 4 \ \mathrm{SU}(N)\ \mathrm{SYM} \ \oplus \ \text{free 10d supergravity in flat space,} \tag{16.21}$$

two non-interacting systems, the first interacting and nontrivial since g_{YM} stays put.

- *Gravity description.* Low energy as measured at infinity again splits in two. At $r \gg L$ spacetime is flat and low energy means long-wavelength supergravity waves. Near the brane the redshift matters: a mode of proper energy E_{proper} at radius r is measured at infinity to have

$$E_\infty = \sqrt{-g_{tt}}\, E_{\text{proper}} = H^{-1/4} E_{\text{proper}} \xrightarrow{r \ll L} \frac{r}{L}\, E_{\text{proper}}, \tag{16.22}$$

so an excitation of *any* proper energy, even a highly excited string, appears at infinity with arbitrarily small energy if it sits far enough down the throat: the throat contributes a full string theory in the near-horizon geometry. The two pieces decouple. From outside, a wave of frequency ω has absorption cross section $\sigma_{\text{abs}} \sim \omega^3 L^8$, vanishing as $\omega L \to 0$ because a wavelength $1/\omega \gg L$ cannot resolve the throat (the absorption computations of [73] were precursors of the correspondence); from inside, an excitation deep in the throat cannot climb out. Hence

$$\text{(closed-string side)} \quad \text{IIB strings in the near-horizon geometry} \ \oplus \ \text{free 10d supergravity.} \tag{16.23}$$

- *The near-horizon geometry is* $\mathrm{AdS}_5 \times S^5$ (Figure 16.2). In the limit at fixed u,

$$H = 1 + \frac{4\pi g_s N}{\alpha'^2 u^4} \longrightarrow \frac{L^4}{r^4} \qquad \text{(the 1 always drops out as } \alpha' \to 0\text{)}, \tag{16.24}$$

so the limit lands automatically in the near-horizon region, where (16.8) becomes

$$\mathrm{d}s^2 = \frac{r^2}{L^2}\, \eta_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu + \frac{L^2}{r^2}\, \mathrm{d}r^2 + L^2\, \mathrm{d}\Omega_5^2. \tag{16.25}$$

The sphere has constant radius L, which is why the geometry is regular, and with $z = L^2/r$

$$\frac{r^2}{L^2}\, \eta_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu + \frac{L^2}{r^2} \mathrm{d}r^2 = \frac{L^2}{z^2} \left(\eta_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu + \mathrm{d}z^2 \right), \tag{16.26}$$

AdS_5 in the Poincaré coordinates of Chapter 10:

$$\boxed{\text{near-horizon geometry} \ = \ \mathrm{AdS}_5 \times S^5, \quad \text{both factors of radius } L, \quad \int_{S^5} F_5 = N \text{ units.}} \tag{16.27}$$

The flux threads the sphere and, by self-duality, fills AdS_5, supporting the geometry and providing the negative cosmological constant of the five-dimensional description. The equality of the two radii follows from the power r^{-4} in H: the curvature scales are locked, so no regime has one factor weakly curved and the other not.

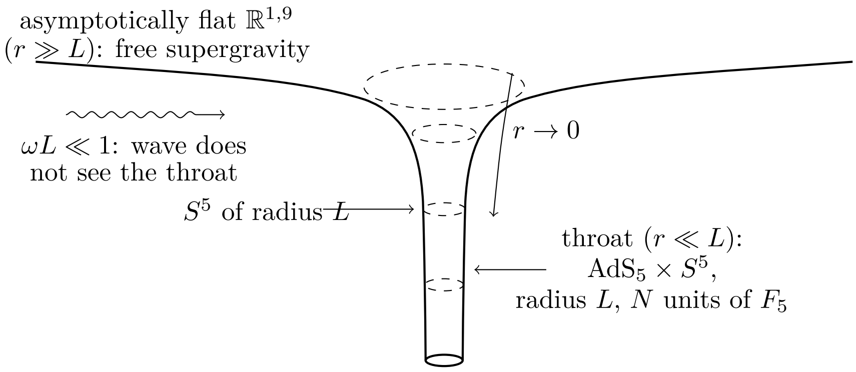


Figure 16.2: The geometry of the extremal black 3-brane. Far from the branes ($r \gg L$) spacetime is flat; for $r \lesssim L$ an infinite throat opens up, whose geometry approaches $\mathrm{AdS}_5 \times S^5$ with both radii equal to L. In the decoupling limit, low-energy physics splits into free supergravity waves in the flat region — too long-wavelength to be absorbed — and the full string theory living down the throat.

- *Two consistency remarks.* In the coordinate $u = r/\alpha'$,

$$\mathrm{d}s^2 = \alpha'\sqrt{\lambda}\left[\frac{u^2}{\lambda}\,\eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu + \frac{\mathrm{d}u^2}{u^2} + \mathrm{d}\Omega_5^2\right], \tag{16.28}$$

  with $\lambda = 4\pi g_s N$ fixed; the overall α' is what multiplies every worldsheet action, so string observables in the throat depend only on the bracket and the limit is finite as a string background. And the isometries of (16.27) are $\mathrm{SO}(2,4)\times\mathrm{SO}(6)$, the conformal group times the rotations of the transverse sphere, exactly the bosonic symmetries of $\mathcal{N} = 4$ SYM; $\mathrm{AdS}_5 \times S^5$ is moreover maximally supersymmetric with 32 Killing spinors, the enhancement of the 16 supercharges of the brane matching that of Poincaré to superconformal symmetry in the CFT (Chapter 17).

- *The conjecture.* The same limit of the same object has produced (16.21) and (16.23), each "interacting core $\oplus$ free bulk supergravity" with identical free factors. Equating the cores [58]:

The AdS/CFT correspondence. Four-dimensional $\mathcal{N} = 4$ super Yang–Mills theory with gauge group $\mathrm{SU}(N)$ is *equal*, as a quantum theory for all values of the parameters, to Type IIB superstring theory on $\mathrm{AdS}_5 \times S^5$, with both radii equal to L and N units of five-form flux through the S^5, with the identification of parameters

$$g_{\mathrm{YM}}^2 = 4\pi g_s, \qquad \frac{L^4}{\alpha'^2} = g_{\mathrm{YM}}^2 N = \lambda, \qquad \theta = 2\pi C_0.$$

"Equal" is meant in the strongest sense: every state, observable and correlation function of one theory has a counterpart in the other, and the dictionary of Chapters 12–13 is its working implementation, now anchored to a specific pair of theories. The decoupled $\mathrm{U}(1)$ of $\mathrm{U}(N)$ is not part of the correspondence; the difference between $\mathrm{SU}(N)$ and $\mathrm{U}(N)$ is invisible at leading order in N but detectable in the Kaluza–Klein spectrum (Chapter 17).

16.6 The parameter map and the shape of the duality

- *Newton's constant in five dimensions.* Since the S^5 has constant radius L, reduction is elementary, $\int \mathrm{d}^{10}x\sqrt{g_{10}}R_{10} \supset \mathrm{Vol}(S^5_L)\int \mathrm{d}^5x\sqrt{g_5}R_5$:

$$\frac{1}{16\pi G_5} = \frac{\mathrm{Vol}(S^5_L)}{16\pi G_{10}} = \frac{\pi^3 L^5}{16\pi G_{10}} \qquad \Longrightarrow \qquad G_5 = \frac{G_{10}}{\pi^3 L^5}. \tag{16.29}$$

The combination controlling every five-dimensional holographic observable, with $G_{10} = 8\pi^6 g_s^2 \alpha'^4$ from (16.16) and $L^8 = 16\pi^2 g_s^2 N^2 \alpha'^4$, is

$$\frac{L^3}{G_5} = \frac{\pi^3 L^8}{G_{10}} = \frac{\pi^3 \cdot 16\pi^2 g_s^2 N^2 \alpha'^4}{8\pi^6 g_s^2 \alpha'^4} \qquad \Longrightarrow \qquad \boxed{\frac{L^3}{G_5} = \frac{2N^2}{\pi}.} \tag{16.30}$$

All g_s and α' have cancelled: five-dimensional gravity in AdS units is a pure function of N, classical exactly when the gauge theory has many colours. Inserted into the holographic central charge of Chapter 14, $a = c = \pi L^3/8G_5 = N^2/4$.

- *The map.* The gauge theory has two dimensionless parameters, $\lambda = g^2_{\mathrm{YM}}N$ and N; the string side has g_s and L^2/α'. The dictionary pairs them as

$$\frac{L^2}{\alpha'} = \sqrt{\lambda}, \qquad g_s = \frac{g^2_{\mathrm{YM}}}{4\pi} = \frac{\lambda}{4\pi N}, \qquad \frac{L^3}{G_5} = \frac{2N^2}{\pi}. \tag{16.31}$$

- $1/N$ *counts quantum gravity effects.* String loops are weighted by $g_s \sim \lambda/N$ and gravitational loops in AdS_5 by $G_5/L^3 \sim 1/N^2$, the dimensionless ratio of Newton's constant to the curvature radius (the dimensionful G_5 alone cannot be an expansion parameter). The genus expansion of string theory is the $1/N$ expansion of Chapter 11 [55], realized concretely with actual Type IIB worldsheets: planar gauge theory = classical string theory.

- $1/\sqrt{\lambda}$ *counts stringy effects.* Corrections from massive string modes and higher-derivative terms are weighted by $\alpha'/L^2 = 1/\sqrt{\lambda}$: finite string-size effects are strong-coupling-suppressed in the gauge theory.

- *Classical supergravity* = planar gauge theory at strong coupling, $N \to \infty$ *and* $\lambda \gg 1$: the regime of Chapters 12–14 (Figure 16.3), and the weak form of the conjecture in the classification of Section 12.2. Integrability and localization now test the 't Hooft form in detail, and no discrepancy with the strong form has been found (Chapter 17); the modern consensus treats the strong form as the definition of what Type IIB string theory on $\mathrm{AdS}_5 \times S^5$ *is.*

The two expansion parameters of holography.

$$\frac{1}{N} \leftrightarrow g_s \ \text{(quantum gravity)}, \qquad \frac{1}{\sqrt{\lambda}} \leftrightarrow \frac{\alpha'}{L^2} \ \text{(stringy corrections)}.$$

Classical Type IIB supergravity on $\mathrm{AdS}_5 \times S^5$ computes the leading term of the double expansion: the *planar, strongly coupled* limit of $\mathcal{N} = 4$ SYM. Corrections in $1/\sqrt{\lambda}$ are α' (higher-derivative) corrections; corrections in $1/N$ are string-loop corrections.

The thermal side of this physics, the near-extremal black 3-brane, AdS black holes, the Hawking–Page transition and the factor $3/4$ in the entropy of strongly coupled $\mathcal{N} = 4$ SYM [73], is taken up in Chapter 18, after the spectral tests of the next chapter.

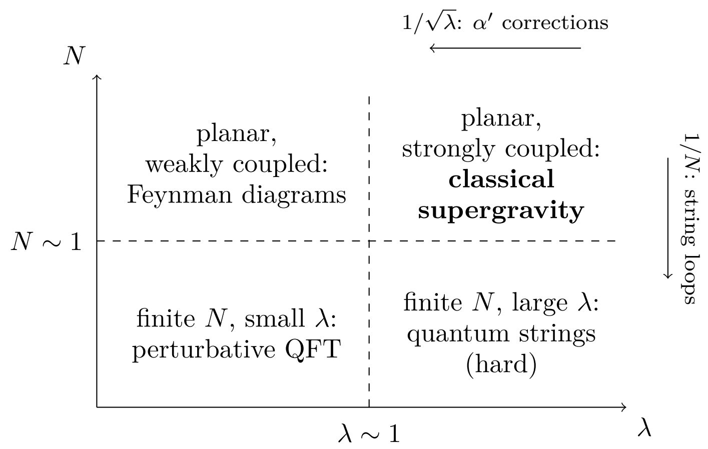


Figure 16.3: The parameter space of the duality. The gravity description is quantitatively simple in the upper-right corner: large N (classical strings, no quantum gravity) and large λ (weak curvature, no stringy corrections). Moving left turns on α' corrections; moving down turns on string loops.

Chapter 17

Matching the Spectrum: Tests of the Correspondence

Chapter 16 left us with a sharp claim, and this chapter collects the evidence. Kinematics first: conformal group, R-symmetry, thirty-two supercharges, and even the non-perturbative $\mathrm{SL}(2,\mathbb{Z})$ duality must coincide, and do. Then the spectrum: the complete Kaluza–Klein tower of Type IIB supergravity on S^5 is reproduced by the chiral primaries $\mathrm{Tr}\,\phi^{(i_1}\cdots\phi^{i_k)}$ with $\Delta = k$ *exactly*, and understanding why, multiplet shortening at unitarity-bound saturation against the $\lambda^{1/4}$ growth of everything unprotected, is the concrete realization of the "few light operators, large gap" structure required of any CFT with an Einstein-gravity dual. We close with tests beyond the spectrum, correlators and Wilson loops, and a look at the landscape beyond $\mathcal{N} = 4$.

17.1 Symmetries: the kinematical test

The most basic requirement on a duality is that the two sides have the same global symmetries, visible without solving either theory:

$\mathcal{N} = 4$ SYM in $d = 4$	Type IIB on $\mathrm{AdS}_5 \times S^5$
conformal group $\mathrm{SO}(2,4)$	isometry group of AdS_5
R-symmetry $\mathrm{SU}(4)_R \simeq \mathrm{SO}(6)$	isometry group of S^5
$16\ Q\ +\ 16\ S$ supercharges	32 Killing spinors
Montonen–Olive $\mathrm{SL}(2,\mathbb{Z})$	Type IIB S-duality $\mathrm{SL}(2,\mathbb{Z})$

- *The conformal group.* $\mathcal{N} = 4$ SYM is an exact CFT, with spacetime symmetry $\mathrm{SO}(2,4)$; the isometry group of AdS_5 is $\mathrm{SO}(2,4)$ acting on the boundary as conformal transformations (Chapter 10). This is the backbone of every AdS/CFT pair; specific to the present one is that the near-horizon limit *produced* an exact AdS factor, dual to a beta function that vanishes identically rather than at a fixed point.

- *The R-symmetry.* $\mathrm{SU}(4)_R \simeq \mathrm{SO}(6)$ is the rotation group of the transverse directions (Chapter 16), which survive in the near-horizon geometry as the five-sphere with isometry group $\mathrm{SO}(6)$. An internal symmetry on one side is an isometry on the other, and the $\mathrm{SO}(6)$ Killing vectors give $\mathrm{SO}(6)$ gauge fields in AdS_5 dual to the R-currents, as used in Chapter 14.

- *Thirty-two supercharges.* The D3-branes preserve 16 of the 32 supersymmetries of Type IIB, and the worldvolume theory has the 16 Poincaré supercharges Q^a_α, $\bar{Q}_{a\dot{\alpha}}$ ($4\times2\times2$ real). But $\mathrm{AdS}_5 \times S^5$ with self-dual flux is *maximally supersymmetric*, with 32 Killing spinors, the extra 16 existing only in the throat; on the field-theory side conformal invariance requires the 16 conformal supercharges S_a^α. Both sides realize $\mathfrak{su}(2,2|4)$, and all operators and bulk states fall into its representations labelled by Δ, the spins (j_1, j_2) and the $\mathrm{SU}(4)_R$ representation, which is what makes the spectral comparison well posed.

- *S-duality.* $\mathcal{N} = 4$ SYM is believed invariant under Montonen–Olive duality [74]: $\tau \to -1/\tau$ of $\tau = \frac{\theta}{2\pi} + \frac{4\pi \mathrm{i}}{g_{\mathrm{YM}}^2}$, exchanging weak and strong coupling and electric charges with monopoles, which with $\theta \to \theta + 2\pi$ generates $\mathrm{SL}(2,\mathbb{Z})$. Type IIB has its own $\mathrm{SL}(2,\mathbb{Z})$ acting on $\tau_{\mathrm{IIB}} = C_0 + \mathrm{i}e^{-\phi}$ and exchanging fundamental strings with D1-branes. Under $g_{\mathrm{YM}}^2 = 4\pi g_s$, $\theta = 2\pi C_0$ the two couplings are identified, $\tau = \tau_{\mathrm{IIB}}$, and the duality groups coincide: a deep unproven property of a gauge theory mapped to a better-established symmetry of string theory, acting at fixed N and inverting $g_s \to 1/g_s$.

17.2 The field theory side: operators and protected dimensions

- *Gauge-invariant operators.* The elementary fields are gauge-variant; the observables are correlators of gauge-invariant composites, the simplest being single traces

$$\mathcal{O}(x) = \mathrm{Tr}\left(\Phi_1(x)\, \Phi_2(x) \cdots \Phi_n(x)\right), \qquad \Phi_k \in \{\phi^i, \psi_a, F_{\mu\nu}, D_\mu \cdots\}, \tag{17.1}$$

and their multi-trace products (single traces are the single-particle objects at large N, Chapter 11). Although the elementary fields are not renormalized, composites are: bringing n fields to a point produces short-distance singularities and an anomalous dimension,

$$\Delta_{\mathcal{O}}(\lambda) = \Delta_{\text{classical}} + \gamma_{\mathcal{O}}(\lambda), \tag{17.2}$$

generically a nontrivial function of the coupling, unknown at large λ. The comparison is saved by operators whose dimensions are exactly coupling-independent.

- *Protection by shortening.* The mechanism is the supersymmetric version of what we met in Chapter 9: there unitarity bounded Δ from below, and at saturation a descendant had zero norm and had to vanish ($\Box\mathcal{O} = 0$ for a scalar at $\Delta = d-2$, $\partial^\mu J_\mu = 0$ for a vector at $\Delta = d-1$), leaving a multiplet smaller than generic. Here the supercharges play the role of P_μ.

  1. *The supercharges.* They are Q_α^a and $\bar{Q}_{a\dot{\alpha}}$, with $a = 1, \ldots, 4$ an $\mathrm{SU}(4)_R$ index and $\alpha, \dot{\alpha} \in \{+,-\}$ Weyl spinor indices: $4 \times 2 = 8$ of each kind, 16 in all. Each carries dimension $+\frac{1}{2}$ and Lorentz spin $(\frac{1}{2}, 0)$ for Q, $(0, \frac{1}{2})$ for $\bar{Q}$ (the conformal supercharges S carry $-\frac{1}{2}$, as K_μ carries -1).
  2. *The long multiplet.* Acting with the supercharges on a superconformal primary $|\mathcal{O}\rangle$ generates its supermultiplet. Supercharges anticommute, so each acts at most once, and the multiplet is spanned by every subset of the 16: 2^{16} conformal primaries, each with its own tower of P_μ descendants, with dimensions from Δ to $\Delta + 8$ [75]. The highest spin is reached by applying the four Q_+^a, $a = 1, \ldots, 4$, each raising j_1 by $\frac{1}{2}$, and the four $\bar{Q}_{a\dot{+}}$, each raising j_2: $(j_1, j_2) = (2, 2)$, spin 4.
  3. *The bound.* The norm of $Q|\mathcal{O}\rangle$ is computed, as in Chapter 9, from the anticommutator $\{Q, S\}$, which is a combination of Δ, the Lorentz spins and the $\mathrm{SU}(4)_R$ generators. Positivity of all these norms bounds Δ from below in terms of the spin and R-representation of the primary. For a scalar primary in the rank-k symmetric traceless representation of $\mathrm{SO}(6)$, Dynkin label $[0, k, 0]$,

$$\Delta \;\geq\; k. \tag{17.3}$$

4. *Saturation and shortening.* At $\Delta = k$ the states $Q|\mathcal{O}\rangle$ for 8 of the 16 supercharges, say Q^a_α with $a = 3, 4$ and $\bar{Q}_{a\dot{\alpha}}$ with $a = 1, 2$, have zero norm and must vanish:[1] these 8 annihilate the primary, which is *half-BPS*. The multiplet is spanned by the remaining 8 alone: 2^8 conformal primaries instead of 2^{16}, and with only two values of a acting the spin tower stops at $(1, 1)$, spin 2. Such a multiplet is called *short*; the spin-2 member of the $k = 2$ multiplet is the stress tensor.

- *Saturation implies protection*, by a continuity argument using no dynamics. The dimension of a short primary is locked to its R-charge, $\Delta = k$ exactly, and k is a *discrete* quantum number. Varying the exactly marginal λ, if Δ rose above k the null states would acquire positive norm and the multiplet would have to become long, but a long multiplet contains many more states, and states cannot materialize discontinuously. (The converse, several short multiplets recombining into one long multiplet that then lifts off, is the loophole to check; for the multiplets below no candidate partners exist.) Hence *a short multiplet has Δ fixed by its R-charge, exactly, for all λ*: the analogues of conserved currents, computable at $\lambda = 0$ and comparable with supergravity at $\lambda = \infty$, as the anomalies of Chapter 14.

- *The chiral primaries.* $\mathcal{N} = 4$ SYM has a natural family saturating (17.3),

$$\mathcal{O}_k^{i_1\cdots i_k} = \mathrm{Tr}\left(\phi^{(i_1}\phi^{i_2}\cdots\phi^{i_k)}\right) - \text{traces}, \qquad k = 2, 3, 4, \dots \tag{17.4}$$

symmetrized with all δ^{ij} traces removed, precisely the $[0, k, 0]$ with classical dimension k. These sit in short multiplets $\mathcal{A}_k$, the *only* single-trace short multiplets of the theory, so

$$\Delta_{\mathcal{O}_k} = k \qquad \text{for all } \lambda. \tag{17.5}$$

The case $k = 2$, the $\mathbf{20}'$,

$$\mathcal{O}_2^{ij} = \mathrm{Tr}\left(\phi^i\phi^j - \frac{\delta^{ij}}{6}\,\phi^l\phi^l\right), \qquad \Delta = 2, \tag{17.6}$$

is the bottom of the multiplet containing all conserved currents: the R-currents $J_\mu^{[ij]}$ ($\Delta = 3$), the supercurrents, the stress tensor ($\Delta = 4$), and two singlet scalars of dimension 4, the exactly marginal $\mathrm{Tr}(F^2 + \cdots)$ and $\mathrm{Tr}(F\tilde{F} + \cdots)$ whose couplings are $1/g^2_{\mathrm{YM}}$ and θ.

- *The Konishi operator*, the unprotected prototype, is (17.6) with the trace *not* removed:

$$K = \mathrm{Tr}\,\phi^i\phi^i, \qquad \Delta_K = 2 + \gamma_K(\lambda), \qquad \gamma_K = \frac{3\lambda}{4\pi^2} + O(\lambda^2) \tag{17.7}$$

[76]. As an SO(6) singlet it is subject only to the generic bound, nothing forces saturation, and it sits at the bottom of a long multiplet: two operators built from the same fields, distinguished by a trace, one protected and one not. At strong coupling the correspondence predicts (Section 17.4.1) that K is dual to a *massive string state*,

$$\Delta_K \;\simeq\; 2\lambda^{1/4}, \qquad \lambda \to \infty, \tag{17.8}$$

untestable in 1998 and confirmed today by integrability [77] (Section 17.5).

[1] On a scalar primary the anticommutator $\{S, Q\}$ gives $\|Q^a_\alpha|\mathcal{O}\rangle\|^2 = \frac{1}{2}(\Delta - 2r_a)$, no sum, with r_a the eigenvalue of the diagonal R-generator $R^a{}_a$. On the highest-weight state $\mathrm{Tr}\,Z^k$, $Z = \phi^1 + \mathrm{i}\phi^2$, of $[0, k, 0]$ the charges are $r_a = (\frac{k}{2}, \frac{k}{2}, -\frac{k}{2}, -\frac{k}{2})$, so four of the Q's have norm $\frac{1}{2}(\Delta - k)$ and four $\frac{1}{2}(\Delta + k)$, with the roles of $a = 1, 2$ and $a = 3, 4$ exchanged for the $\bar{Q}$'s. Positivity of the first set is the bound (17.3), and at $\Delta = k$ exactly those eight states are null; higher levels give nothing stronger [75].

17.3 The gravity side: Kaluza–Klein reduction on S^5

- *Harmonics on a circle, harmonics on a sphere.* Below the string scale, dimensions below $\lambda^{1/4}$, the dual of the operator spectrum is the spectrum of Type IIB *supergravity* on $\mathrm{AdS}_5 \times S^5$, organized in Kaluza–Klein towers. On $\mathbb{R}^{1,3} \times S^1$ of radius R a massless scalar $\Phi = \sum_n \phi_n(x) \mathrm{e}^{iny/R}$ obeys

$$0 = \Box_5 \Phi = \sum_n \mathrm{e}^{iny/R} \Big(\Box_4 - \frac{n^2}{R^2} \Big) \phi_n, \tag{17.9}$$

  a tower of masses $|n|/R$. On the unit $S^5 \subset \mathbb{R}^6$ the scalar harmonics are restrictions of harmonic polynomials,

$$Y^{(k)}(\hat{y}) = C_{i_1 \cdots i_k} \hat{y}^{i_1} \cdots \hat{y}^{i_k}, \qquad C \text{ symmetric traceless}, \qquad -\Box_{S^5} Y^{(k)} = k(k+4)\, Y^{(k)}, \tag{17.10}$$

  generalizing $Y_{\ell m}$ with $\ell(\ell+1)$. Two features: the degree-k harmonics transform in *the same representations* $[0,k,0]$ as the chiral primaries (17.4); and since the sphere radius is tied to the AdS radius, the KK splittings are of order $1/L$, the AdS curvature scale, so the tower never decouples and holography must account for all of it.

- *The mass towers.* The complete linearized spectrum was computed by Kim, Romans and van Nieuwenhuizen [78]: each ten-dimensional field is expanded in harmonics, fluctuation equations mix fields with the same quantum numbers, and diagonalization yields towers labelled by k. The tower relevant for the chiral primaries comes from the mixing of the sphere trace of the metric $h^\alpha{}_\alpha$ with the sphere components of the four-form: a scalar in the $[0,k,0]$ with

$$m_k^2 L^2 = k(k-4), \qquad k = 2, 3, 4, \dots \tag{17.11}$$

  i.e. $m^2 L^2 = -4, -3, 0, +5, \dots$ (Figure 17.1). At $k=2$ one finds besides it the massless graviton, fifteen SO(6) gauge fields (from $g_{\mu\alpha}$ mixed with the four-form), a massless complex scalar from the dilaton–axion and their fermionic partners: the maximal gauged supergravity multiplet in five dimensions. The entire spectrum assembles into *one short multiplet of* $\mathfrak{su}(2,2|4)$ *for each* $k \geq 2$, with (17.11) as lowest member and top spin two; no long multiplets appear. The negative m^2 cause no alarm: the $k=2$ member sits exactly *at* the Breitenlohner–Freedman bound $m^2 L^2 = -4$ of Chapter 10, as it must, being dual to the operator of lowest dimension, $\Delta = 2$.

- *The missing* $k=1$ *level* would form a lone ultrashort multiplet, the doubleton of Günaydin–Marcus [79], often called the singleton [80]; it is pure gauge in the bulk, its modes removable everywhere except on the boundary, and absent from the physical spectrum. This is the free U(1) centre-of-mass factor of the brane stack, $\mathrm{U}(N) = \mathrm{SU}(N) \times \mathrm{U}(1)$, which decouples: the theory dual to the $\mathrm{AdS}_5 \times S^5$ bulk is $\mathrm{SU}(N)$, as in Chapter 16.

17.4 The comparison

- *Dimensions from masses:* $\Delta = k$*, exactly.* A scalar of mass m in AdS_5 is dual to an operator with

$$\Delta(\Delta - 4) = m^2 L^2, \qquad \Delta = 2 + \sqrt{4 + m^2 L^2} \tag{17.12}$$

  (the larger root), and inserting (17.11),

$$\Delta_k = 2 + \sqrt{4 + k(k-4)} = 2 + \sqrt{(k-2)^2} = 2 + (k-2) = k. \tag{17.13}$$

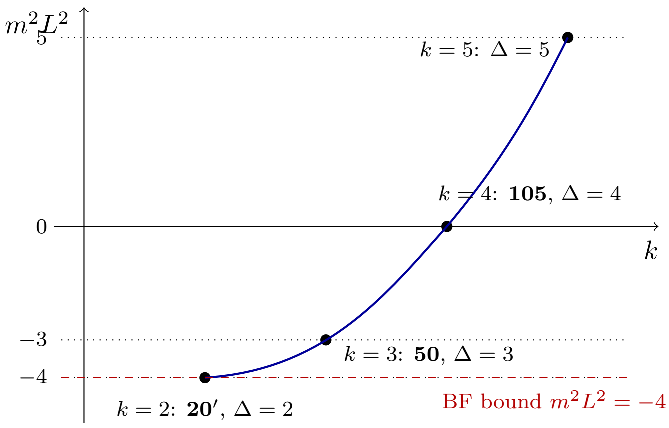


Figure 17.1: The lightest Kaluza–Klein scalar tower on $\mathrm{AdS}_5\times S^5$: $m^2L^2=k(k-4)$ in the $[0,k,0]$ of SO(6) [78]. The $k=2$ member saturates the Breitenlohner–Freedman bound; via $\Delta(\Delta-4)=m^2L^2$ the tower is dual to the chiral primaries with $\Delta=k$. Marginal and relevant operators ($\Delta\le 4$) correspond to $m^2\le 0$.

The discriminant is a perfect square for *every* k, anything but generic, and the result matches the protected dimension (17.5) of $\mathcal{O}_k$ in the same SO(6) representation. Both numbers are coupling-independent, $\Delta_{\mathcal{O}_k}$ by shortening and m_k because no correction can shift a protected multiplet: a genuine test, in the class of the anomaly matchings of Chapter 14.

Spectral test of AdS/CFT. The full Kaluza–Klein spectrum of Type IIB supergravity on $\mathrm{AdS}_5\times S^5$ consists of one short multiplet of $\mathfrak{su}(2,2|4)$ for each $k\ge 2$, built on a scalar in the $[0,k,0]$ of SO(6) with $m^2L^2=k(k-4)$. The single-trace short multiplets of $\mathcal{N}=4$ SU(N) SYM consist of one multiplet $\mathcal{A}_k$ for each $k\ge 2$, built on the chiral primary $\mathrm{Tr}\,\phi^{(i_1}\cdots\phi^{i_k)}$ with protected dimension $\Delta=k$. Under $\Delta(\Delta-4)=m^2L^2$ the two lists coincide, multiplet by multiplet, state by state.

- *The working dictionary* for the most important fields:

bulk field on AdS_5	SO(6) rep	m^2L^2	dual operator	Δ
graviton $g_{\mu\nu}$	**1**	0	stress tensor $T_{\mu\nu}$	4
gauge fields A_μ	**15**	0	R-currents $J_\mu^{[ij]}$	3
dilaton φ	**1**	0	$\mathrm{Tr}(F^2+\cdots)$	4
axion C_0	**1**	0	$\mathrm{Tr}(F\tilde F+\cdots)$	4
scalars $(h^\alpha{}_\alpha, C_4)$ tower	$[0,k,0]$	$k(k-4)$	$\mathcal{O}_k=\mathrm{Tr}\,\phi^{(i_1}\cdots\phi^{i_k)}$	k

Every line is an instance of the rule that bulk gauge symmetries correspond to boundary global symmetries: massless graviton and vectors map to the conserved stress tensor and currents with protected dimensions 4 and 3, and the massless dilaton–axion, whose constant values are (g_{YM},θ), to the marginal operators varying those couplings. Three refinements: the absent $k=1$ multiplet means $\mathrm{Tr}\,\phi^i$ has no bulk counterpart, evidence for SU(N); the multiplets match member by member, fermions and tensor modes included [78, 11]; and multi-trace operators such as $\mathcal{O}_2\mathcal{O}_2$ are multi-particle states, as in Chapter 11.

17.4.1 The gap: where did everything else go?

- *Massive string states.* The gauge theory contains vastly more operators than the $\mathcal{A}_k$: Konishi, operators with derivatives, high-spin towers $\mathrm{Tr}\,\phi\,\partial_{\mu_1}\cdots\partial_{\mu_s}\phi$, an exponentially growing density near their classical dimensions at weak coupling. They are the excited string states, with masses
$$m_n^2 = \frac{4n}{\alpha'}, \qquad n = 1, 2, \ldots \tag{17.14}$$
for which (17.12) gives at large mass
$$\Delta_n \simeq m_n L = 2\sqrt{n}\,\frac{L}{\sqrt{\alpha'}} = 2\sqrt{n}\,\lambda^{1/4}, \tag{17.15}$$
using $L^2/\alpha' = \sqrt{\lambda}$. At strong coupling all unprotected operators acquire dimensions of order $\lambda^{1/4}$ and drop out of the operator algebra; (17.8) is (17.15) with $n = 1$, Konishi being the bottom of the first excited string level.

- *The gap structure of Chapter 12*, realized concretely: a small set of light operators (the protected $\mathcal{A}_k$, top spin two, no light operators of spin > 2) separated by a parametrically large gap $\lambda^{1/4}$ from everything else, with large-N factorization. Supersymmetric protection pins a subset of dimensions while strong coupling inflates the rest. Conversely, as $\lambda \to 0$ the string states come down to dimensions of order one, the bulk becomes string-scale curved with a proliferation of light higher-spin fields, and the gravity description dissolves, as it must for a free gauge theory.

17.5 Further tests: correlators, Wilson loops, and the correspondence today

- *Three-point functions of chiral primaries.* Conformal invariance fixes the two- and three-point functions of the $\mathcal{O}_k$ up to the coefficients $C_{k_1k_2k_3}$ of unit-normalized operators (Chapter 9). On the gravity side these are contact Witten diagrams of Chapter 13 with the cubic couplings of the KK-reduced supergravity, a heroic computation by Lee, Minwalla, Rangamani and Seiberg [81], whose result is that for all k_1, k_2, k_3 the supergravity answer *coincides with the free-field value* from Wick contractions at $\lambda = 0$. This is now understood as a non-renormalization theorem following from superconformal Ward identities [82], verified by localization [83]: the correspondence *suggested* a theorem later proven in field theory. The first coupling-dependent data appear in four-point functions, which exchange unprotected operators such as Konishi.

- *Wilson loops.* Besides local operators, a gauge theory has observables attached to closed curves. In $\mathcal{N} = 4$ SYM the natural one is
$$W[C] \;=\; \frac{1}{N}\,\mathrm{Tr}\,\mathcal{P}\exp\oint_C \mathrm{d}s\,\Big(\mathrm{i}A_\mu\dot{x}^\mu + |\dot{x}|\,\theta^I\phi^I\Big), \qquad \theta^I\theta^I = 1, \tag{17.16}$$
the path-ordered exponential in the fundamental with the scalars coupled along a unit vector $\theta^I \in S^5$ so that the loop preserves some supersymmetry [84]: the phase of an infinitely heavy quark dragged around C. For a rectangle of width r and length $\mathcal{T} \gg r$, a quark–antiquark pair held at separation r, inserting a complete set of states as for $Z(\beta)$ in (7.1) gives
$$\langle W[C]\rangle \;\sim\; \mathrm{e}^{-\mathcal{T}E(r)} \qquad (\mathcal{T}\to\infty), \tag{17.17}$$
the static potential; in a CFT $E(r) = -f(\lambda)/r$, and at weak coupling one-gluon and one-scalar exchange give $f = \lambda/4\pi + O(\lambda^2)$ [85].

- *The holographic Wilson loop* (Figure 17.2). The heavy quark is the endpoint of a fundamental string: separating one D3-brane from the stack and sending it to the boundary leaves a W boson of mass (16.1), an open string stretched from the boundary into the bulk, and dragging the quark drags the endpoint around C [84, 86]. The worldsheet is weighted by the Nambu–Goto action (15.1) in the AdS_5 metric, and with $L^2/\alpha' = \sqrt{\lambda}$ by (16.17),
$$S_{\mathrm{NG}}[\Sigma] \;=\; \frac{\sqrt{\lambda}}{2\pi}\,\mathcal{A}[\Sigma], \tag{17.18}$$
with $\mathcal{A}$ the area in units of L^2. At large λ the string path integral is dominated by the minimal worldsheet bounded by C,
$$\langle W[C]\rangle \;\simeq\; \mathrm{e}^{-S_{\mathrm{NG}}[\Sigma_{\min}]} \qquad (N\to\infty,\ \lambda\gg 1). \tag{17.19}$$
The area diverges at $z\to 0$, where (10.11) blows up; the divergence is the action of a string from the boundary to the horizon, the mass of the infinitely heavy quark, and is subtracted. For the rectangle the minimal worldsheet is the U-shaped string reaching $z_* \propto r$ by the scaling symmetry (10.12), its regularized area an elliptic integral, and [84]
$$E(r) \;=\; -\,\frac{4\pi^2}{\Gamma(\frac14)^4}\,\frac{\sqrt{\lambda}}{r}, \tag{17.20}$$
a genuine strong-coupling prediction against the weak-coupling $\lambda/4\pi$.

- *The circular loop*, where the prediction can be tested. The sum of gluon and scalar propagators between two points of a circle is constant, so the ladder diagrams sum in closed form at large N [87, 88],
$$\langle W_{\mathrm{circle}}\rangle \;=\; \frac{2}{\sqrt{\lambda}}\,I_1(\sqrt{\lambda}) \;\xrightarrow{\lambda\gg 1}\; \sqrt{\frac{2}{\pi}}\,\frac{\mathrm{e}^{\sqrt{\lambda}}}{\lambda^{3/4}}, \tag{17.21}$$
and that all other diagrams cancel was later proven by localization, which reduces $\langle W_{\mathrm{circle}}\rangle$ to a Gaussian matrix integral [83]. The minimal surface spanning a circle has regularized area $\mathcal{A} = -2\pi$, so (17.19) gives $\log\langle W\rangle \to \sqrt{\lambda}$, in agreement; the prefactor $\lambda^{-3/4}$ has been reproduced from the one-loop string fluctuation determinant [89]. Among the sharpest confirmations in a non-protected observable.

- *Planar integrability.* The planar one-loop dilatation operator of $\mathcal{N}=4$ SYM on single-trace operators is an integrable spin-chain Hamiltonian [90], and Bethe-ansatz methods and their descendants compute $\Delta_{\mathcal{O}}(\lambda)$ of *unprotected* operators at any coupling; see [91] for a review. For Konishi the curve [77] interpolates between (17.7) and (17.8), subleading coefficients matched against semiclassical string energies: in the 't Hooft limit the spectral problem is solved in principle.

- *Localization* computes exact quantities at finite N and λ, sphere partition functions, the circular loop (17.21), protected correlators, confronted with string theory beyond supergravity including $1/N$ effects; all checks are consistent with the *strong* form of the conjecture.

- *Black-hole microstates.* The Strominger–Vafa counting [19] predates Maldacena's paper and belongs to the same circle of ideas: degeneracies of a D1–D5 bound state, counted at weak coupling by two-dimensional CFT methods, reproduce $A/4G$ of the corresponding extremal black hole, the same open/closed double vision applied to entropy. In modern language it is an $\mathrm{AdS}_3/\mathrm{CFT}_2$ correspondence, and its $\mathrm{AdS}_5\times S^5$ analogue, counting

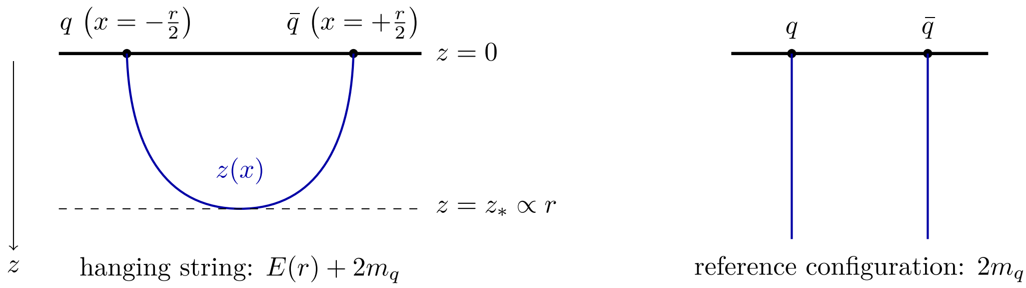


Figure 17.2: The holographic rectangular Wilson loop, in the Poincaré coordinates of (10.11) with the boundary at $z=0$ and one spatial direction x shown. Left: the fundamental string ends on the quark and antiquark worldlines at the boundary and sags into the bulk down to the turning point $z_*\propto r$; its regularized area gives $E(r)$ in (17.20). Right: the reference configuration of two straight strings falling into the bulk, whose divergent action is the self-energy $2m_q$ of the infinitely heavy quarks.

black-hole microstates with the superconformal index, has recently reached the same success: holography provides exactly the microscopic account of black-hole entropy whose absence haunted Part I, and Part III builds on this.

- *Beyond $\mathcal{N}=4$.* The near-horizon logic is a machine: any brane configuration whose worldvolume theory decouples yields a dual pair. D3-branes at singular points of curved transverse spaces give CFTs with less supersymmetry and $\mathrm{AdS}_5\times M_5$ duals, orbifolds S^5/Γ giving quiver gauge theories [92, 93] and the conifold the Klebanov–Witten theory on $\mathrm{AdS}_5\times T^{1,1}$ [94]; deformations by relevant operators or temperature move the duality off conformality, with geometries that cap off in the interior describing confinement and mass gaps, the prototype being Klebanov–Strassler [95]. These developments, §3.4 and §4 of [2], lie beyond this course; $\mathcal{N}=4$ SYM is not an isolated miracle but the cleanest member of a large family.

With this, Part II is complete: a working, quantitatively tested duality between a four-dimensional quantum field theory and a theory of quantum gravity. The deepest questions of Part I, what microstates the Bekenstein–Hawking entropy counts, whether evaporation is unitary, what happens to information behind a horizon, can now be posed where quantum gravity has a complete non-perturbative definition. Part III returns to black holes with holography in hand: thermal AdS black holes and the Hawking–Page transition (Chapter 18), then entropy, entanglement and the Page curve, and finally the Ryu–Takayanagi formula, where the geometry of spacetime emerges from patterns of quantum information.

Part III

Black Holes Revisited: Holography and Quantum Information

Chapter 18

Black Holes in Anti-de Sitter Space and the Hawking–Page Transition

An asymptotically flat black hole has negative specific heat and no canonical ensemble (Chapters 5 and 7); this lecture shows that a negative cosmological constant repairs everything. We construct AdS–Schwarzschild, find two branches of black holes—the large one thermodynamically *stable*—and derive the first-order Hawking–Page transition [96] by comparing Euclidean actions. Through the dictionary of Part II the transition becomes the confinement/deconfinement transition of the dual gauge theory on $S^3 \times S^1$, and the planar limit yields the entropy of strongly coupled $\mathcal{N} = 4$ SYM with the famous factor 3/4. Finally the eternal AdS black hole is identified as the thermofield double of two CFTs [49], giving precise meaning to the Euclidean constructions of Chapters 6 and 7.

18.1 Anti-de Sitter space as a box, and the AdS–Schwarzschild solution

- *The problem.* The flat-space black hole has negative specific heat, (7.26), and its canonical partition function diverges, (5.17), because the density of states $\mathrm{e}^{S(M)} = \mathrm{e}^{4\pi G M^2}$ outgrows every Boltzmann factor. Section 5.5 announced the cure: a box.
- *AdS is a covariant box of size L*, for two reasons established earlier. A light ray reaches the conformal boundary in the finite time (10.15), so with the reflecting boundary conditions of Chapter 12 radiation emitted by a black hole returns in a time of order L. And by Tolman's law (6.15) a state of temperature T with respect to global time has local temperature $T/\sqrt{1 + r^2/L^2} \to TL/r$, so a thermal gas in AdS has energy density falling like $r^{-(d+1)}$ and finite total energy. This chapter puts a black hole in the box and studies the gravitational canonical ensemble at temperature T with AdS boundary conditions.
- *The solution.* The unique spherically symmetric solution of the vacuum Einstein equations with cosmological constant $\Lambda = -d(d-1)/2L^2$, (10.2), is the AdS–Schwarzschild metric (Birkhoff's theorem extends to $\Lambda \neq 0$):

$$\boxed{\mathrm{d}s^2 = -f(r)\,\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{f(r)} + r^2\,\mathrm{d}\Omega_{d-1}^2, \qquad f(r) = 1 + \frac{r^2}{L^2} - \frac{\mu}{r^{d-2}}. \qquad (18.1)}$$

- *Limits and mass.* For $\mu = 0$ this is global AdS; for $L \to \infty$ it is the $(d+1)$-dimensional Schwarzschild solution. The integration constant μ is related to the ADM mass by

$$\mu \;=\; \frac{16\pi G\,M}{(d-1)\,\Omega_{d-1}}, \qquad \Omega_{d-1} \equiv \mathrm{Vol}(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)}, \tag{18.2}$$

and for $\mu > 0$ the function f has a single positive zero $r = r_h$, the horizon radius, in terms of which

$$\mu \;=\; r_h^{d-2}\Big(1+\frac{r_h^2}{L^2}\Big), \qquad M \;=\; \frac{(d-1)\,\Omega_{d-1}}{16\pi G}\, r_h^{d-2}\Big(1+\frac{r_h^2}{L^2}\Big). \tag{18.3}$$

The mass increases monotonically with r_h. Our main example is $d = 4$, where $\Omega_3 = 2\pi^2$, $\mu = 8GM/3\pi$, and G is the five-dimensional Newton constant; we keep d general since nothing depends on it.

- *Small and large.* A black hole with $r_h \ll L$ does not feel the box, since r^2/L^2 is negligible in f near the horizon; one with $r_h \gg L$ has no flat-space analogue.

18.2 Temperature and the two branches

- *The cigar.* Continuing $t \to -\mathrm{i}\tau$ as in Chapter 7, the Euclidean section of (18.1) is defined for $r \geq r_h$ and is a cigar with tip at $r = r_h$. Smoothness of the tip fixes the period of τ as in (7.7), with surface gravity $\kappa = f'(r_h)/2$ from (1.9): $\beta = 2\pi/\kappa = 4\pi/f'(r_h)$. Since the conformal boundary is $S^1_\beta \times S^{d-1}$, the black hole is in equilibrium at temperature $T = 1/\beta$ with respect to the boundary time, and with $f'(r_h) = d\,r_h/L^2 + (d-2)/r_h$ from (18.3),

$$\boxed{T(r_h) \;=\; \frac{1}{4\pi}\left(\frac{d\,r_h}{L^2} + \frac{d-2}{r_h}\right) \;=\; \frac{d\,r_h^2 + (d-2)L^2}{4\pi L^2\, r_h}.} \tag{18.4}$$

- *Limits.* For $r_h \ll L$ (and $d = 3$) this is the Hawking temperature $1/4\pi r_h$ of (7.8); for $r_h \gg L$ it is linear in the size,

$$T \;\simeq\; \frac{d}{4\pi}\,\frac{r_h}{L^2} \qquad (r_h \gg L). \tag{18.5}$$

- *Two branches.* Because $T(r_h)$ diverges at both ends, it has a minimum, at $\mathrm{d}T/\mathrm{d}r_h = 0$:

$$r_{\min} = L\sqrt{\frac{d-2}{d}}, \qquad T_{\min} = \frac{\sqrt{d(d-2)}}{2\pi L} \qquad \Big(r_{\min} = \tfrac{L}{\sqrt{2}},\ T_{\min} = \tfrac{\sqrt{2}}{\pi L}\ \text{ for } d = 4\Big). \tag{18.6}$$

Hence (Figure 18.1): for $T < T_{\min}$ there is no black hole, and the only Euclidean geometry with boundary $S^1_\beta \times S^{d-1}$ is **thermal AdS**, i.e. (18.1) with $\mu = 0$ and $\tau \sim \tau + \beta$; for $T > T_{\min}$ there are two, a ***small* black hole** with $r_h < r_{\min}$ and a ***large* one** with $r_h > r_{\min}$. Their stability is settled in Section 18.3 once the entropy is known.

- *Topology.* The two fillings of the same boundary differ in topology. The black hole is $\mathbb{R}^2 \times S^{d-1}$, the disc of the cigar times the sphere, and the thermal circle is contractible; thermal AdS is $S^1 \times B^d$, and the thermal circle is not. This distinction is the key to Sections 18.3 and 18.5. (It also fixes the fermion boundary condition: on the black hole the spin structure forces antiperiodicity around the contractible circle, the thermal boundary condition of Section 7.1; on thermal AdS both choices are allowed and thermal physics selects the antiperiodic one.)

18.3 The Euclidean action

Following Chapter 7, we estimate the partition function with boundary $S^1_\beta \times S^{d-1}$ by saddle points,

$$Z(\beta) \;\simeq\; \sum_{\text{saddles}} \mathrm{e}^{-I_E[g]}, \qquad I_E \;=\; -\frac{1}{16\pi G}\int \mathrm{d}^{d+1}x\,\sqrt{g}\,(R-2\Lambda) \;-\; \frac{1}{8\pi G}\oint \mathrm{d}^d x\,\sqrt{\gamma}\,K, \tag{18.7}$$

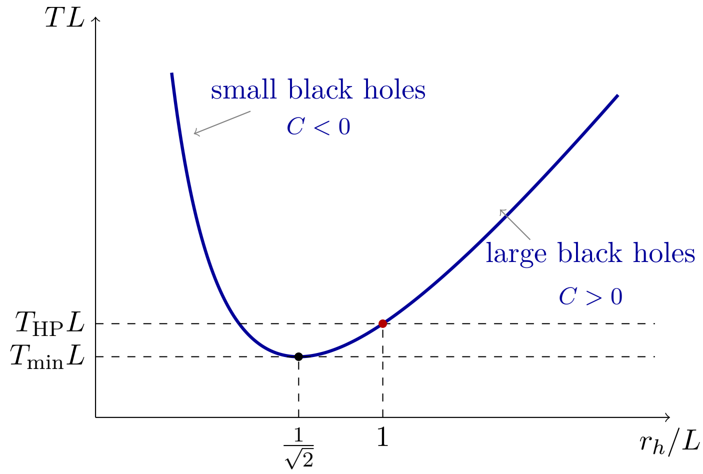


Figure 18.1: Temperature of the AdS–Schwarzschild black hole as a function of horizon radius, for $d = 4$ (equation (18.4)). Below $T_{\min} = \sqrt{2}/\pi L$ no black hole exists. Above it there are two branches; the small branch has negative specific heat, the large branch positive. The red dot marks $r_h = L$, where the Hawking–Page transition of Section 18.3 occurs, at $T_{\mathrm{HP}} = 3/2\pi L$.

the action (7.10) with the cosmological term added, and the boundary term evaluated on a cutoff surface $r = R \to \infty$. The saddles are thermal AdS and, for $T > T_{\min}$, the two black holes.

18.3.1 The subtracted action ΔI

1. *On shell the bulk integrand is a constant.* The vacuum Einstein equations with cosmological constant in $d+1$ dimensions, $R_{\mu\nu} - \frac{1}{2}R\, g_{\mu\nu} + \Lambda g_{\mu\nu} = 0$, have trace $R - \frac{d+1}{2}R + (d+1)\Lambda = 0$, i.e. $R = 2(d+1)\Lambda/(d-1)$; with $\Lambda = -d(d-1)/2L^2$ this is $R = -d(d+1)/L^2$, and therefore
$$R - 2\Lambda \;=\; -\frac{d(d+1)}{L^2} + \frac{d(d-1)}{L^2} \;=\; -\frac{2d}{L^2}. \tag{18.8}$$
Inserting (18.8) into the bulk term of (18.7), the bulk action is proportional to the volume of the Euclidean geometry,
$$I_E^{\text{bulk}} \;=\; -\frac{1}{16\pi G}\Big(-\frac{2d}{L^2}\Big)\int \mathrm{d}^{d+1}x\,\sqrt{g} \;=\; \frac{d}{8\pi G L^2}\,\mathrm{Vol}(\mathcal{M}). \tag{18.9}$$

2. *Matching at the cutoff.* Each volume is infinite; the meaningful quantity is the difference of the actions of the black hole and of thermal AdS at the same boundary data, as in Section 7.4. Cutting both geometries at $r = R$, the S^{d-1} factors agree automatically, and the thermal circles must have the same proper length,
$$\beta'\sqrt{f_0(R)} \;=\; \beta\sqrt{f(R)}, \qquad f_0(r) \equiv 1 + \frac{r^2}{L^2}, \qquad f(r) = 1 + \frac{r^2}{L^2} - \frac{\mu}{r^{d-2}}, \tag{18.10}$$
where β is the black hole period and β' the period assigned to thermal AdS at the cutoff. For large R,
$$\beta' \;=\; \beta\left(1 - \frac{\mu L^2}{2R^d} + O(R^{-d-2})\right). \tag{18.11}$$
The mismatch is subleading but multiplies a volume of order R^d, so it survives the limit; it is where the mass enters. In the difference the boundary terms cancel as $R \to \infty$.[1]

[1]The counterterms of holographic renormalization, Chapter 13, give the same result without a subtraction,

3. *The regulated volumes* follow from the Euclidean metric (18.1), whose determinant is $\sqrt{g} = r^{d-1}\sqrt{g_\Omega}$ because the factors f from $g_{\tau\tau}$ and $1/f$ from g_{rr} cancel:

$$\mathrm{Vol} \;=\; \int_0^\beta \mathrm{d}\tau \int \mathrm{d}r\, r^{d-1} \int \mathrm{d}\Omega_{d-1} \;=\; \beta\,\Omega_{d-1} \int \mathrm{d}r\, r^{d-1}. \tag{18.12}$$

For the black hole the Euclidean section exists only for $r \geq r_h$, the tip of the cigar, so r runs from r_h to the cutoff R with the period β of (18.4); for thermal AdS, $\mu = 0$ and there is no horizon, so r runs from 0 to R, with the period β' fixed by the matching (18.10). Hence

$$\begin{aligned} \mathrm{Vol}_{\mathrm{BH}} \;&=\; \frac{\beta\,\Omega_{d-1}}{d}\,(R^d - r_h^d), \\ \mathrm{Vol}_{\mathrm{AdS}} \;&=\; \frac{\beta'\,\Omega_{d-1}}{d}\,R^d \;=\; \frac{\beta\,\Omega_{d-1}}{d}\left(R^d - \frac{\mu L^2}{2}\right) + O(R^{-2}), \end{aligned} \tag{18.13}$$

where the last form uses (18.11).

4. *The result.* By (18.9), inserting μ from (18.3) and β from (18.4),

$$\boxed{\begin{aligned} \Delta I \;\equiv\; \lim_{R\to\infty}\,(I_{\mathrm{BH}} - I_{\mathrm{AdS}}) \;&=\; \frac{\beta\,\Omega_{d-1}}{8\pi G L^2}\left(\frac{\mu L^2}{2} - r_h^d\right) \;=\; \frac{\Omega_{d-1}\,\beta}{16\pi G}\,\frac{r_h^{d-2}\,(L^2 - r_h^2)}{L^2} \\ &=\; \frac{\Omega_{d-1}}{4G}\,\frac{r_h^{d-1}(L^2 - r_h^2)}{d\,r_h^2 + (d-2)L^2}. \end{aligned}} \tag{18.14}$$

This is the result of Hawking and Page [96] for $d = 3$ and of Witten [97] in general dimension; for $d = 4$, $\Delta I = \pi\beta\, r_h^2(L^2 - r_h^2)/8GL^2$.

18.3.2 Energy, entropy and specific heat

With $\log Z = -\Delta I$, energies being measured relative to thermal AdS, the thermodynamic identities (7.4) read

$$E = \frac{\partial\,\Delta I}{\partial\beta}, \qquad S = \beta\,\frac{\partial\,\Delta I}{\partial\beta} - \Delta I, \tag{18.15}$$

with derivatives along the family of black holes, through $r_h(\beta)$.

1. *Energy.* Write $D \equiv d\,r_h^2 + (d-2)L^2$, so that $\beta = 4\pi L^2 r_h/D$ by (18.4). Differentiating β and (18.14) with respect to r_h,

$$\frac{\mathrm{d}\beta}{\mathrm{d}r_h} = 4\pi L^2\,\frac{(d-2)L^2 - d\,r_h^2}{D^2}, \qquad \frac{\mathrm{d}\,\Delta I}{\mathrm{d}r_h} = \frac{\Omega_{d-1}(d-1)}{4G}\,\frac{r_h^{d-2}(L^2 + r_h^2)\,[(d-2)L^2 - d\,r_h^2]}{D^2}, \tag{18.16}$$

and the ratio is

$$E \;=\; \frac{\mathrm{d}\,\Delta I/\mathrm{d}r_h}{\mathrm{d}\beta/\mathrm{d}r_h} \;=\; \frac{(d-1)\,\Omega_{d-1}}{16\pi G}\,r_h^{d-2}\Big(1 + \frac{r_h^2}{L^2}\Big) \;=\; M, \tag{18.17}$$

the ADM mass (18.3). The factor $[(d-2)L^2 - d\,r_h^2]$, which vanishes at $r_{\min}$, cancels between numerator and denominator, so the result holds on both branches.

up to a β-linear Casimir term: they assign to global AdS$_5$ the energy $E_0 = 3\pi L^2/32G$, the Casimir energy $3(N^2-1)/16L$ of $\mathcal{N} = 4$ SYM on S^3. This term is scheme dependent and does not affect the phase structure, which involves only differences.

2. *Entropy.* From (18.15), (18.17) and (18.14),

$$\begin{aligned} S \;=\; \beta M - \Delta I \;&=\; \frac{\Omega_{d-1}\,\beta\, r_h^{d-2}}{16\pi G L^2}\Big[(d-1)(L^2+r_h^2)-(L^2-r_h^2)\Big] \\ &=\; \frac{\Omega_{d-1}\,\beta\, r_h^{d-2}}{16\pi G L^2}\,D \;=\; \frac{\Omega_{d-1}\, r_h^{d-1}}{4G} \;=\; \frac{A}{4G}, \end{aligned} \tag{18.18}$$

using $\beta D = 4\pi L^2 r_h$ in the last step. The area law of Chapter 7 holds in the presence of a cosmological constant, as the conical-deficit argument of Section 7.6, local at the horizon, already implied.

3. *Specific heat.* Since S and T are both known as functions of r_h,

$$C \;=\; T\,\frac{\mathrm{d}S/\mathrm{d}r_h}{\mathrm{d}T/\mathrm{d}r_h} \;=\; \frac{(d-1)\,\Omega_{d-1}\, r_h^{d-1}}{4G}\;\frac{d\, r_h^2+(d-2)L^2}{d\, r_h^2-(d-2)L^2}, \tag{18.19}$$

whose sign is that of $r_h - r_{\min}$ by (18.6): $C<0$ on the small branch, which is locally a flat-space black hole and inherits (7.26), and $C>0$ on the large branch. A large black hole that fluctuates to a higher temperature radiates, loses energy, and cools back; with the radiation reflected within a time of order L, it is a genuinely stable equilibrium. For $r_h \gg L$, (18.5) and (18.18) give $S \propto T^{d-1}$, the entropy of a thermal system in $d-1$ spatial dimensions, a first hint of the dual description.

18.4 The Hawking–Page transition

- *The crossover.* The sign of (18.14) changes at $r_h = L$ in any dimension: black holes with $r_h < L$ have $\Delta I > 0$ and are subdominant to thermal AdS, those with $r_h > L$ have $\Delta I < 0$ and dominate. The crossover temperature is, by (18.4),

$$T_{\mathrm{HP}} \;=\; T(r_h{=}L) \;=\; \frac{d-1}{2\pi L} \qquad \Big(=\frac{3}{2\pi L}\;\text{ for } d=4\Big), \tag{18.20}$$

and since $L > r_{\min}$, $T_{\mathrm{HP}} > T_{\min}$ and the transition takes place on the large, stable branch.

- *Free energies as functions of T.* Since (18.4) is a quadratic equation for r_h whose discriminant is $T^2 - T_{\min}^2$ by (18.6),

$$r_\pm(T) \;=\; \frac{2\pi L^2}{d}\Big[T \pm \sqrt{T^2 - T_{\min}^2}\,\Big], \qquad F_\pm(T) \;\equiv\; \frac{\Delta I}{\beta} \;=\; \frac{\Omega_{d-1}}{16\pi G}\,\frac{r_\pm^{d-2}\,(L^2 - r_\pm^2)}{L^2}, \tag{18.21}$$

with r_+ the large and r_- the small black hole. Three features can be read off:

  1. at $T = T_{\min}$ the square root vanishes and the branches meet at $r_\pm = 2\pi L^2 T_{\min}/d = r_{\min}$, a cusp;
  2. $F_+ = 0$ requires $r_+ = L$, i.e. $\sqrt{T^2 - T_{\min}^2} = 1/2\pi L$, which gives $T_{\mathrm{HP}}^2 = T_{\min}^2 + 1/4\pi^2 L^2 = (d-1)^2/4\pi^2L^2$, reproducing (18.20);
  3. for $T \gg T_{\min}$, $r_+ \simeq 4\pi L^2 T/d$ and $F_+ \simeq -\frac{\Omega_{d-1}}{16\pi G}\big(\frac{4\pi}{d}\big)^d L^{2d-2}\,T^d$, the free energy of a conformal gas in d dimensions, which Section 18.5 will identify with the $\mathcal{N}=4$ plasma.

  Since $\partial F/\partial T = -S < 0$, both branches decrease with T.

- *The ensemble* (Figure 18.2, which plots (18.21) for $d=4$ together with thermal AdS, whose free energy is $O(G^0)$, from the graviton gas, hence zero at this order):

1. $T < T_{\min}$: thermal AdS only.
2. $T_{\min} < T < T_{\rm HP}$: thermal AdS is the global minimum of F, the large black hole a metastable local minimum, and the small black hole, with $C < 0$, the unstable saddle between them: the critical droplet, in nucleation language.
3. $T > T_{\rm HP}$: the large black hole is the global minimum, with $F = O(1/G)$, and thermal AdS is metastable.

- *First order.* At $T_{\rm HP}$ the free energies of thermal AdS and of the large black hole cross with different slopes, $\partial F/\partial T = 0$ for the former and $-S_{\rm BH}$ for the latter: the entropy $S = -\partial F/\partial T$ of the dominant phase jumps from $O(1)$ to $\Omega_{d-1}L^{d-1}/4G = O(1/G)$, with latent heat $T_{\rm HP}\,\Delta S = M(r_h{=}L)$. This resolves, within AdS, the puzzle of Chapter 5: **the canonical ensemble exists at all temperatures, and the negative-specific-heat black holes appear only as unstable saddles controlling nucleation, never as equilibrium states**. Gravity in a box has conventional thermodynamics because, as we now show, it *is* the thermodynamics of an ordinary quantum system, whose order parameter is a Polyakov loop.

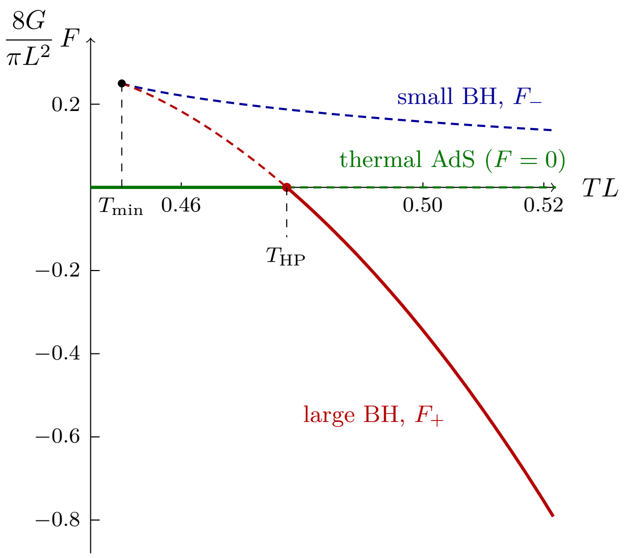


Figure 18.2: The free energies $F_\pm(T)$ of (18.21) for $d = 4$, in units of $\pi L^2/8G$, relative to thermal AdS ($F = 0$ at leading order in G). The two black hole branches emerge in a cusp at $T_{\min}L = \sqrt{2}/\pi \simeq 0.450$, where $F = \frac{1}{4}$ in these units. Both decrease with T. The small branch is never dominant. The large branch crosses $F = 0$ at $T_{\rm HP}L = 3/2\pi \simeq 0.477$ and dominates above it. Solid lines mark the dominant phase at each temperature; the kink at $T_{\rm HP}$ is the first-order transition.

18.5 The dual interpretation: confinement and deconfinement

- *The dictionary.* Everything so far was pure gravity, and Hawking and Page's paper predates AdS/CFT by fifteen years. But Part II hands us a dictionary: quantum gravity with $\mathrm{AdS}_5 \times S^5$ boundary conditions *is* $\mathcal{N} = 4$ super Yang–Mills with gauge group $SU(N)$, with

$$\frac{L^3}{G_5} = \frac{2N^2}{\pi}, \tag{18.22}$$

as derived in (16.30). The Euclidean path integral with boundary $S^1_\beta \times S^3$ computes the *thermal partition function of the CFT on a spatial* S^3 (of radius L, in our conformal frame) at temperature T:

$$Z_{\text{CFT}}[S^1_\beta \times S^3] \;=\; \sum_{\text{fillings}} \mathrm{e}^{-I_E} + \dots \;=\; \mathrm{e}^{-I_{\text{AdS}}} + \mathrm{e}^{-I_{\text{BH}}} + \dots, \tag{18.23}$$

where the sum runs over the smooth bulk geometries found above [97]. The gravitational phase transition is therefore a genuine thermodynamic phase transition *of the gauge theory*, in the limit $N \to \infty$, $\lambda = g^2_{\text{YM}} N \gg 1$. (At finite N both saddles contribute and the free energy is analytic; a sharp transition requires the thermodynamic limit, here played by $N \to \infty$, since on a compact S^3 the volume cannot diverge.)

18.5.1 Order 1 versus order N^2

Translate the free energies using (18.22).

- *Thermal AdS phase* ($T < T_{\text{HP}}$): the classical actions cancel in the subtraction, and F is of order $G^0 \sim N^0$. In the gauge theory this is a phase where the free energy is $O(1)$: only color-singlet states contribute. On S^3, where a Gauss-law constraint forces all states to be singlets anyway, the low-temperature spectrum consists of a discrete tower of "glueball-like" singlet excitations (dual to the graviton and other supergravity modes in AdS) with gap set by $1/L$; their thermal gas has $F = O(1)$. This is the finite-volume analogue of a *confined* phase.

- *Black hole phase* ($T > T_{\text{HP}}$): $F = \Delta I/\beta \sim -L^{d-1}/G \sim -N^2$. The free energy is that of $\sim N^2$ liberated gluon (and superpartner) degrees of freedom: a *deconfined* plasma.

- *The identification.* The Hawking–Page transition is the holographic image of the *confinement/deconfinement transition* of the $SU(N)$ gauge theory at large N [97, 2]: the formation of a black hole horizon in five dimensions and the deconfinement of gluons in four are the *same phenomenon* in two descriptions. The black hole entropy $A/4G \sim N^2(TL)^3$ is, microscopically, the entropy of the $\mathcal{N} = 4$ plasma: the promised answer, for this class of black holes, to the question left open in Chapter 7. The Euclidean gravitational entropy *is* a counting of states, the states of the dual gauge theory.

18.5.2 The Polyakov loop diagnostic

- *The order parameter.* In a thermal gauge theory the *Polyakov loop* is the trace of the holonomy of the gauge field around the Euclidean time circle,

$$P(\vec{x}) \;=\; \frac{1}{N}\, \mathrm{Tr}\, \mathcal{P} \exp\Big(\mathrm{i} \oint_0^\beta \mathrm{d}\tau\, A_\tau(\tau, \vec{x})\Big). \tag{18.24}$$

and $\langle P \rangle \sim \mathrm{e}^{-\beta F_q}$ measures the free energy F_q of a single static external quark. Gauge transformations periodic up to a centre element $\mathrm{e}^{2\pi \mathrm{i} k/N} \in \mathbb{Z}_N$ leave all adjoint fields invariant but multiply P by the centre element, so the theory on S^1_β has a global $\mathbb{Z}_N$ symmetry under which P is charged:

$$\langle P \rangle = 0 \;\Leftrightarrow\; \mathbb{Z}_N \text{ unbroken} \;\Leftrightarrow\; F_q = \infty \;\Leftrightarrow\; \text{confined}; \qquad \langle P \rangle \neq 0 \;\Leftrightarrow\; \text{deconfined}. \tag{18.25}$$

- *Holographically*, a static quark is the endpoint of a fundamental string, and a Wilson loop is computed by the worldsheet ending on the loop, (17.19). For the Polyakov loop the loop is purely temporal: the quark sits at fixed $\vec{x}$ and C is its worldline over one period of

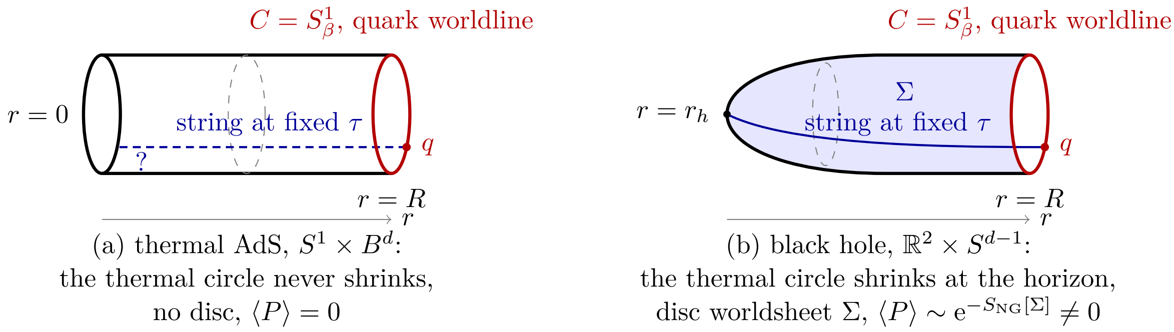


Figure 18.3: The Polyakov loop in the two Euclidean geometries, drawn in the (τ, r) plane with the S^{d-1} suppressed, as in Figure 7.1. The static quark q sits at the cutoff $r = R$ at a fixed point of S^{d-1} and does not move; its Euclidean worldline is the thermal circle S^1_β (red), closed by the periodicity of τ, and this is the loop C in (17.19). At fixed τ the string attached to q is a radial line into the bulk (blue), and sweeping τ around the circle generates the worldsheet. (a) In thermal AdS the radial lines reach $r = 0$, where only the S^{d-1} shrinks and there is nothing for an open string to end on (denoted by ?): the worldsheet would need a second boundary, no disc ending on C exists, and $\langle P\rangle = 0$ at leading order. (b) In the black hole the radial lines meet at the tip $r = r_h$, where the thermal circle shrinks, and close into the disc worldsheet Σ, the cigar itself at a fixed point of S^{d-1}; its Nambu–Goto action is finite after the subtraction of Section 17.5, hence $\langle P\rangle \neq 0$.

Euclidean time, closed by the periodicity of τ, the red circle in Figure 18.3. The leading contribution is a disc worldsheet with that circle as boundary, and here the topological distinction between the two fillings becomes physics. In *thermal AdS* the boundary circle is non-contractible in the bulk; no disc exists, and $\langle P\rangle = 0$ at this order: confinement. In the *black hole* the thermal circle closes off at the horizon; a smooth disc worldsheet hangs from the boundary to the tip, and $\langle P\rangle \sim \mathrm{e}^{-S_{\mathrm{NG}}[\Sigma]} \neq 0$: deconfinement. "Is the Euclidean time circle contractible?" translates into "is the centre symmetry broken?". The horizon is the bulk incarnation of deconfinement.

Remark 18.1. Since $\mathcal{N} = 4$ SYM is conformal, on flat $\mathbb{R}^3$ it has no scale and no transition; the transition found here exists only because the S^3 curvature scale $1/L$ competes with T. Remarkably, the deconfinement transition on S^3 is visible even at *weak* coupling, where the free energy of the singlet-constrained gas exhibits a Hagedorn-like transition at $TL \sim O(1)$ [98, 99]; the Hawking–Page transition is its strong-coupling continuation. Witten [97] further showed that compactifying $\mathcal{N} = 4$ SYM on a circle with antiperiodic fermions, and taking the circle small, gives a gravity dual of *non-supersymmetric* Yang–Mills theory in three dimensions. The same pair of fillings appears, now distinguished by which circle contracts in the bulk: the geometry in which the supersymmetry-breaking circle contracts is the confining phase, with an area law for spatial Wilson loops and a mass gap, while the one in which the thermal circle contracts is the deconfined phase. That story belongs to a string theory course, and we set it aside.

18.6 Planar black holes and the entropy of strongly coupled $\mathcal{N} = 4$ SYM

18.6.1 The black brane and its temperature

- *Decompactifying the sphere.* To describe the CFT at temperature T on flat $\mathbb{R}^{d-1}$, take the limit $r_h/L \to \infty$ of the large black hole: zooming in on a small patch of the horizon, the metric (18.1) with the "1" in f dropped becomes the *planar AdS black hole*, or black brane,

$$\boxed{\mathrm{d}s^2 \;=\; \frac{r^2}{L^2}\Big(-h(r)\,\mathrm{d}t^2 + \mathrm{d}\vec{x}^{\,2}\Big) + \frac{L^2\,\mathrm{d}r^2}{r^2\,h(r)}, \qquad h(r) = 1 - \frac{r_0^d}{r^d}, \qquad (18.26)}$$

- *Temperature.* Here $\vec{x} \in \mathbb{R}^{d-1}$ and the horizon is at $r = r_0$. (In the coordinate $z = L^2/r$ this is the Euclidean black three-brane in the convention of [2], with $z_0 = L^2/r_0$; for $d = 4$ it is the near-horizon limit of the finite-temperature black 3-brane.) Smoothness of the Euclidean section at $r = r_0$ fixes the temperature as before: with $f(r) = r^2h/L^2 = g_{rr}^{-1}$,

$$T \;=\; \frac{f'(r_0)}{4\pi} \;=\; \frac{r_0^2\,h'(r_0)}{4\pi L^2} \;=\; \frac{d\,r_0}{4\pi L^2} \qquad \Big(= \frac{r_0}{\pi L^2} = \frac{1}{\pi z_0} \quad \text{for } d = 4\Big), \qquad (18.27)$$

which is precisely the asymptotic law (18.5).

- *No transition.* The boundary theory on $\mathbb{R}^{d-1}$ has no scale other than T: the dilatation $(t, \vec{x}) \to \lambda(t, \vec{x})$, $r \to r/\lambda$ maps the brane with r_0 to the brane with r_0/λ, rescaling the temperature. All temperatures are equivalent, so *on flat space the black brane dominates for every* $T > 0$ and the plasma is always deconfined (thermal Poincaré AdS has strictly larger action; equivalently $T_{\mathrm{HP}} \sim 1/L_{S^3} \to 0$ as the sphere grows). The two branches also disappear: $T(r_0)$ is monotonic, the specific heat is always positive, and $T_{\min} \to 0$.

18.6.2 The entropy and the $3/4$ factor

- *Entropy.* The horizon of (18.26) is the surface $r = r_0$, t fixed, with induced metric $(r_0^2/L^2)\,\mathrm{d}\vec{x}^{\,2}$. Regulating $\mathbb{R}^{d-1}$ by a coordinate volume $V = \int \mathrm{d}^{d-1}x$, the physical volume in the CFT, the horizon area and entropy are

$$A \;=\; \Big(\frac{r_0}{L}\Big)^{d-1} V, \qquad S \;=\; \frac{A}{4G} \;=\; \frac{r_0^{d-1}}{4G\,L^{d-1}}\,V \;=\; \frac{1}{4G}\Big(\frac{4\pi}{d}\Big)^{d-1} L^{d-1}\,T^{d-1}\,V, \qquad (18.28)$$

using (18.27). The scaling $S \propto VT^{d-1}$ is fixed by extensivity and conformal invariance; the coefficient is the prediction. For $d = 4$, with $G_5 = \pi L^3/2N^2$ from (18.22) and $r_0 = \pi L^2 T$:

$$\boxed{S_{\mathrm{sugra}} \;=\; \frac{(\pi L^2 T)^3\,V}{4G_5\,L^3} \;=\; \frac{\pi^3 L^3 T^3}{4G_5}\,V \;=\; \frac{\pi^2}{2}\,N^2\,V\,T^3. \qquad (18.29)}$$

- *Consistency.* Correspondingly $F = -\pi^2N^2VT^4/8$ and $E = 3\pi^2N^2VT^4/8$, satisfying $E = 3TS/4$ and $F = -E/3$ as conformal invariance demands; evaluating the subtracted Euclidean action on (18.26) as in Section 18.3.1 gives this F [2].

- *The free value.* Equation (18.29) is a parameter-free prediction for the entropy density of $\mathcal{N} = 4$ SYM at $N \to \infty$ and $\lambda \to \infty$. At the opposite extreme, $\lambda = 0$, the free $\mathcal{N} = 4$ multiplet has per generator $n_B = 8$ bosonic degrees of freedom (two gauge-boson polarizations plus six scalars) and $n_F = 8$ fermionic ones (four Weyl fermions), so the blackbody entropy is

$$S_{\text{free}} \;=\; \frac{2\pi^2}{45}\Big(n_B+\tfrac{7}{8}n_F\Big)(N^2-1)\,V\,T^3 \;\simeq\; \frac{2\pi^2}{45}\,(8+7)\,N^2\,V\,T^3 \;=\; \frac{2\pi^2}{3}\,N^2\,V\,T^3. \quad (18.30)$$

Hence

$$\frac{S_{\text{sugra}}}{S_{\text{free}}} \;=\; \frac{\pi^2/2}{2\pi^2/3} \;=\; \frac{3}{4}. \quad (18.31)$$

- *The famous* 3/4, first found by Gubser, Klebanov and Peet [73] *before* AdS/CFT was formulated, as a property of the black 3-brane, when it was unclear whether the discrepancy was a failure of the D-brane description. There is no contradiction: the two numbers are the endpoints of a single smooth function $S(\lambda) = \frac{2\pi^2}{3}N^2VT^3 f(\lambda)$, $f(0) = 1$ and $f(\infty) = 3/4$. The entropy is a thermal quantity, not protected by supersymmetry, and does depend on the coupling: two-loop perturbation theory gives $f = 1 - \frac{3}{2\pi^2}\lambda + \dots$ [100], the leading α' correction to supergravity $f = \frac{3}{4}(1 + \frac{15}{8}\zeta(3)\lambda^{-3/2} + \dots)$ [101], the two approaching each other from opposite ends. That an interacting gauge theory at infinite coupling has an entropy only 25% below the free value is itself a lesson: strongly coupled $\mathcal{N} = 4$ plasma is thermodynamically close to an ideal gas, a theme that resurfaced in the holographic modelling of the quark–gluon plasma.

18.7 The eternal AdS black hole as a thermofield double

- *Two copies.* Section 7.8 identified the Hartle–Hawking state of the eternal black hole with the thermofield double (7.37) of its two exteriors, with the caveat that in flat space the two "copies" had no independent definition. In AdS they do: the maximally extended AdS–Schwarzschild geometry has two asymptotic boundaries, causally disconnected and joined through the Einstein–Rosen bridge behind the horizons, each supporting its own copy of the dual CFT on S^{d-1}. Maldacena's proposal [49]:

The eternal two-sided AdS black hole at inverse temperature β is dual to *two non-interacting copies* of the CFT in the thermofield double state

$$|\mathrm{TFD}\rangle \;=\; \frac{1}{\sqrt{\mathrm{Z}(\beta)}}\sum_{\mathrm{n}} \mathrm{e}^{-\beta \mathrm{E_n}/2}\,|\mathrm{n}\rangle_{\mathrm{L}} \otimes |\mathrm{n}\rangle_{\mathrm{R}} \;\in\; \mathcal{H}_{\mathrm{L}} \otimes \mathcal{H}_{\mathrm{R}}, \quad (18.32)$$

i.e. (6.19) with H the CFT Hamiltonian on S^{d-1}. Neither copy is in a thermal state; the pair is in a pure state whose reduced density matrix on either copy is $\mathrm{e}^{-\beta H}/Z$.

Three checks, each already established in this course:

1. The half-cigar path integral (7.36) prepares the Hartle–Hawking state on the two-sided slice; on the CFT side the same half-period Euclidean path integral is the operator $\mathrm{e}^{-\beta H/2}$, whose associated state is (18.32) by (6.18)–(6.19).

2. Tracing out one copy gives the thermal state, (7.35): each exterior is the equilibrium black hole of Section 18.3.

3. Two-sided correlators $\langle \mathrm{TFD}|\mathcal{O}_{\mathrm{L}}\,\mathcal{O}_{\mathrm{R}}|\mathrm{TFD}\rangle$ are thermal correlators at imaginary-time separation $\beta/2$, (7.38); in the bulk they are computed by propagation through the Einstein–Rosen bridge, with quantitative agreement in solvable limits [49].

- *Two consequences for Part III.* Two decoupled CFTs evolve unitarily, so the eternal AdS black hole embeds black hole thermodynamics in a manifestly unitary quantum system; this sharpens, rather than resolves, the question of how unitarity coexists with Hawking's calculation. And the horizon entropy $A/4G$ is, from the two-sided viewpoint, the entropy of the reduced state of one CFT in (18.32): an entanglement entropy, in the sense defined in Chapter 19, interpreted in Chapter 20 and computed holographically in Chapter 21.

Chapter 19

Two Notions of Entropy

Chapter 18 called $A/4G$ an *entanglement* entropy; Part I used it as a *thermodynamic* entropy. These are genuinely different concepts, and this lecture develops both properly. We introduce density matrices and the von Neumann entropy, prove the basic inequalities from positivity of relative entropy, and confront the central tension: the fine-grained entropy of a closed system is exactly constant, while thermodynamic entropy grows—the latter is a coarse-grained quantity. After Rényi entropies and the replica idea (needed in Chapter 21) and a preview of the area law of entanglement in QFT (Chapter 20), we ask what the fine-grained entropy of Hawking radiation does as a black hole evaporates: Page's random-subsystem analysis gives the *Page curve*, and comparing it with Hawking's calculation turns the information problem into a crisp statement about two curves that part ways at the Page time.

19.1 The tea-cup revisited

- *Thermodynamic entropy.* In Chapter 2 "entropy" meant the entropy of nineteenth-century thermodynamics: a function of a few macroscopic variables (energy, volume, charge), defined for systems in or near equilibrium, and subject to the Second Law. A hot cup of tea in a cold room has some entropy S_{tea}; as tea and room equilibrate, the total thermodynamic entropy increases; and if the tea is dropped into a black hole, Bekenstein taught us to add $A/4G$ to the ledger so that the Generalized Second Law holds. Everything in Part I used entropy in this sense.

- *Fine-grained entropy.* If we know the exact quantum state of a system, not just a handful of macroscopic averages, we can ask how much genuine statistical uncertainty the state itself contains. The answer is the *von Neumann entropy* of its density matrix, also called the *fine-grained* or *microscopic* entropy, the quantum version of the Gibbs–Shannon entropy of a probability distribution. It does not obey a second law: for a closed system evolving unitarily it is *exactly constant*. The tea in the cold room, viewed as a closed quantum system, has a fine-grained entropy that never changes, even as thermometers everywhere report irreversible equilibration.

- *Both are physical.* The thermodynamic entropy is what heat engines, and horizons in Part I, know about; the fine-grained entropy is what *information* knows about, since its constancy under unitary evolution tracks exactly the question "can the initial state be recovered?". The questions to answer: which of the two is $A/4G$? and which did Hawking's calculation of Chapter 4 compute for the radiation?

19.2 Density matrices

19.2.1 Definition and basic structure

Let $\mathcal{H}$ be a Hilbert space, which in this lecture we take finite-dimensional, $\dim \mathcal{H} = d < \infty$, except where stated otherwise; this keeps the linear algebra honest, and the infinite-dimensional and field-theoretic subtleties are flagged as they arise.

Definition 19.1. A *density matrix* (or density operator, or state) on $\mathcal{H}$ is a linear operator $\rho : \mathcal{H} \to \mathcal{H}$ that is (i) self-adjoint, $\rho^\dagger = \rho$; (ii) positive semi-definite, $\langle\psi|\rho|\psi\rangle \geq 0$ for all ψ; (iii) normalized, $\operatorname{Tr}\rho = 1$. The expectation value of an observable $\mathcal{O}$ in the state ρ is $\langle \mathcal{O}\rangle_\rho = \operatorname{Tr}(\rho\,\mathcal{O})$.

Density matrices arise whenever a quantum system must be described with less than complete information. If the system is known to be in the pure state $|\psi_i\rangle$ with probability p_i, all expectation values are captured by

$$\rho \;=\; \sum_i p_i\,|\psi_i\rangle\langle\psi_i|, \qquad p_i \geq 0, \quad \sum_i p_i = 1, \tag{19.1}$$

which satisfies (i)–(iii). Conversely, by the spectral theorem any density matrix can be written as

$$\rho \;=\; \sum_{i=1}^{d} p_i\,|i\rangle\langle i|, \qquad \langle i|j\rangle = \delta_{ij}, \quad p_i \geq 0, \quad \sum_i p_i = 1, \tag{19.2}$$

with *orthonormal* eigenvectors and eigenvalues forming a probability distribution. Note the asymmetry between the two statements: the ensemble (19.1) producing a given ρ is highly non-unique (the $|\psi_i\rangle$ need not be orthogonal), whereas the spectral form (19.2) is canonical; no measurement can distinguish two ensembles with the same density matrix.

Definition 19.2. A state ρ is *pure* if $\rho = |\psi\rangle\langle\psi|$ for some unit vector ψ, i.e. if ρ is a rank-one projector. Otherwise it is *mixed*.

Proposition 19.1 (Purity criterion)**.** *For any density matrix,* $\operatorname{Tr}\rho^2 \leq 1$*, with equality if and only if* ρ *is pure.*

Proof. In the spectral basis, $\operatorname{Tr}\rho^2 = \sum_i p_i^2 \leq \sum_i p_i = 1$ since $p_i^2 \leq p_i$ for $0 \leq p_i \leq 1$. Equality requires $p_i^2 = p_i$ for every i, i.e. each $p_i \in \{0,1\}$; normalization then forces exactly one eigenvalue to equal 1, which is the statement that ρ is a rank-one projector. □

The quantity $\operatorname{Tr}\rho^2$ is called the *purity*; its failure to reach 1 is the first of a family of measures of mixedness (the Rényi entropies of Section 19.8). Under unitary time evolution $\rho \to U\rho\,U^\dagger$ (equivalently $\mathrm{i}\,\partial_t\rho = [H,\rho]$), the spectrum $\{p_i\}$ is exactly preserved: evolution reshuffles the eigenvectors but never the probabilities. Every quantity built from the spectrum alone—purity, all the entropies below—is therefore a constant of the motion for a closed system. Remember this; it is the engine of Section 19.7.

19.2.2 Reduced density matrices

The second, and for us most important, source of density matrices is *entanglement*. Suppose the Hilbert space factorizes,

$$\mathcal{H} \;=\; \mathcal{H}_A \otimes \mathcal{H}_B, \tag{19.3}$$

where A is the subsystem we can access (the exterior of a horizon, the Hawking radiation, one copy of the thermofield double) and B is the rest. Even if the total system is in a pure state $|\Psi\rangle$, an observer with access only to operators of the form $\mathcal{O}_A \otimes \mathbf{1}_B$ sees the state

$$\rho_A \;=\; \operatorname{Tr}_B|\Psi\rangle\langle\Psi|, \qquad \operatorname{Tr}(\rho_A\,\mathcal{O}_A) \;=\; \langle\Psi|\mathcal{O}_A \otimes \mathbf{1}_B|\Psi\rangle \quad \text{for all } \mathcal{O}_A, \tag{19.4}$$

where the *partial trace* Tr_B is defined in components by $(\rho_A)_{aa'} = \sum_b \Psi_{ab}\overline{\Psi_{a'b}}$ for $|\Psi\rangle = \sum_{a,b} \Psi_{ab}|a\rangle_A \otimes |b\rangle_B$. The defining property on the right of (19.4) determines ρ_A uniquely; the same definition applies when the total state is itself a density matrix ρ_{AB}, with $\rho_A = \mathrm{Tr}_B\,\rho_{AB}$.

The structure of a pure bipartite state is completely captured by a simple and extremely useful normal form.

Proposition 19.2 (Schmidt decomposition)**.** *Any unit vector* $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ *can be written as*

$$|\Psi\rangle \;=\; \sum_k \sqrt{p_k}\,|k\rangle_A \otimes |k\rangle_B, \qquad p_k > 0, \quad \sum_k p_k = 1, \tag{19.5}$$

where $\{|k\rangle_A\}$ *and* $\{|k\rangle_B\}$ *are orthonormal sets in* $\mathcal{H}_A$ *and* $\mathcal{H}_B$ *and the sum has at most* $\min(\dim \mathcal{H}_A, \dim \mathcal{H}_B)$ *terms. Consequently*

$$\rho_A = \sum_k p_k |k\rangle_A\langle k|_A, \qquad \rho_B = \sum_k p_k |k\rangle_B\langle k|_B : \tag{19.6}$$

the two reduced density matrices of a pure state have the same nonzero spectrum.

Proof. Regard the components Ψ_{ab} as a $\dim \mathcal{H}_A \times \dim \mathcal{H}_B$ matrix and apply the singular value decomposition, $\Psi = U\Sigma V^\dagger$ with U, V unitary and Σ diagonal with non-negative entries $\sqrt{p_k}$. Defining $|k\rangle_A$ as the columns of U and $|k\rangle_B$ as the complex conjugates of the columns of V gives (19.5); normalization of Ψ gives $\sum_k p_k = 1$, and (19.6) follows by direct computation of the partial traces, using orthonormality. □

If more than one p_k is nonzero, ρ_A is mixed even though the global state is pure: the mixedness is entirely due to entanglement between A and B, not to ignorance about the global state. This is precisely the situation of Chapter 6, where the Minkowski vacuum—a pure state—looked thermal to a Rindler observer: the reduced density matrix of the right wedge, $\rho_R = \mathrm{e}^{-2\pi K}/Z$, is mixed because the vacuum entangles the two wedges. It is also the situation of the thermofield double, to which we return in Section 19.5.

19.3 Von Neumann entropy

Definition 19.3. The *von Neumann entropy* of a density matrix ρ with eigenvalues $\{p_i\}$ is

$$S(\rho) \;=\; -\,\mathrm{Tr}\,\rho\log\rho \;=\; -\sum_i p_i \log p_i, \tag{19.7}$$

with the convention $0\log 0 = 0$ (justified by continuity of $x\log x$ at $x = 0^+$). When $\rho = \rho_A$ is the reduced density matrix of a subsystem, $S(\rho_A) \equiv S_A$ is called the *entanglement entropy* of A.

This is the Shannon entropy of the spectral probability distribution: $S(\rho)$ measures the information (in nats, since we use natural logarithms) that we lack about the microstate, or equivalently the number e^S of states over which ρ is effectively spread.

Proposition 19.3 (Basic properties)**.** *For any density matrix* ρ *on a* d*-dimensional Hilbert space:*

(i) $S(\rho) \geq 0$*, with equality if and only if* ρ *is pure.*

(ii) $S(U\rho U^\dagger) = S(\rho)$ *for any unitary* U*. In particular the von Neumann entropy of a closed system is constant in time.*

(iii) $S(\rho) \leq \log d$*, with equality if and only if* $\rho = \mathbf{1}/d$ *is maximally mixed.*

(iv) If the global state of AB *is* pure*, then* $S_A = S_B$*.*

Proof. (i) Each term $-p_i \log p_i \geq 0$ for $p_i \in [0,1]$, vanishing only for $p_i \in \{0,1\}$; so $S = 0$ forces the spectrum $\{1, 0, \ldots, 0\}$, i.e. a pure state, and conversely. (ii) The spectrum is unitarily invariant and S depends only on the spectrum. (iv) is immediate from Proposition 19.2: the two reduced density matrices have the same nonzero eigenvalues, and zero eigenvalues do not contribute to (19.7). We postpone (iii) to Corollary 19.4, where it will follow from Klein's inequality in one line. □

- *Property (iv)*: for a pure global state the entanglement entropy is a property of the *split*, not of either side separately. A tiny subsystem entangled with a huge one, and the huge one itself, have exactly equal entropies.

- *Property (ii)* is the statement advertised in Section 19.1: no unitary evolution, however complicated, changes $S(\rho)$ of the whole system. It does *not* apply to a subsystem: ρ_A evolves non-unitarily, A exchanging entanglement with B, and S_A can and does change with time. This loophole is where all the physics of the Page curve lives.

- *Infinite dimensions.* The bound (iii) becomes empty and $S(\rho)$ can diverge even for normalizable states; in quantum field theory the entropy of a subregion *always* diverges (Section 19.9).

19.4 Relative entropy and the fundamental inequalities

Almost every inequality in quantum information theory is most efficiently derived from a single master quantity.

Definition 19.4. The *relative entropy* of ρ with respect to σ is

$$S(\rho \,\|\, \sigma) \;=\; \mathrm{Tr}\,(\rho \log \rho) - \mathrm{Tr}\,(\rho \log \sigma), \tag{19.8}$$

defined whenever the support of ρ lies in the support of σ (and set to $+\infty$ otherwise).

Relative entropy is not a distance—it is asymmetric in its arguments—but it behaves like one: it measures how hard it is to mistake the true state ρ for the hypothesis σ (after N measurements on copies of ρ, the probability of confusion decays as $\mathrm{e}^{-NS(\rho\|\sigma)}$). We will need only the following theorem.

Theorem 19.1 (Klein's inequality)**.** *$S(\rho \,\|\, \sigma) \geq 0$, with equality if and only if $\rho = \sigma$.*

Proof. Diagonalize both operators: $\rho = \sum_i p_i |i\rangle\langle i|$ and $\sigma = \sum_a q_a |\hat{a}\rangle\langle \hat{a}|$, with each basis orthonormal. Then

$$S(\rho \,\|\, \sigma) \;=\; \sum_i p_i \log p_i - \sum_{i,a} p_i \, P_{ia} \log q_a, \qquad P_{ia} \equiv |\langle i|\hat{a}\rangle|^2. \tag{19.9}$$

The matrix P_{ia} is *doubly stochastic*: $P_{ia} \geq 0$, $\sum_a P_{ia} = 1$ (completeness of $\{\hat{a}\}$) and $\sum_i P_{ia} = 1$ (completeness of $\{i\}$). Since log is concave, Jensen's inequality applied to the probability distribution $P_{i\cdot}$ gives, for each fixed i,

$$\sum_a P_{ia} \log q_a \;\leq\; \log r_i, \qquad r_i \equiv \sum_a P_{ia}\, q_a \;>\; 0, \tag{19.10}$$

so that

$$S(\rho \,\|\, \sigma) \;\geq\; \sum_i p_i \,\log \frac{p_i}{r_i}. \tag{19.11}$$

The right-hand side is a *classical* relative entropy between the distributions $\{p_i\}$ and $\{r_i\}$ (note $\sum_i r_i = \sum_a q_a \sum_i P_{ia} = \sum_a q_a = 1$). Using the elementary inequality $\log x \geq 1 - 1/x$ for $x > 0$ (with equality only at $x = 1$), applied to $x = p_i/r_i$ for the terms with $p_i > 0$,

$$\sum_i p_i \log \frac{p_i}{r_i} \;\geq\; \sum_i p_i\Big(1 - \frac{r_i}{p_i}\Big) \;=\; \sum_i p_i - \sum_i r_i \;=\; 1 - 1 \;=\; 0. \tag{19.12}$$

For the equality case: equality in (19.12) requires $p_i = r_i$ for all i, and equality in (19.10) requires (strict concavity of log) that for each i the values q_a be constant over all a with $P_{ia} \neq 0$. Together these conditions say that each eigenvector of ρ lies in an eigenspace of σ with eigenvalue $r_i = p_i$; hence ρ and σ are simultaneously diagonalizable with equal spectra on their common support, i.e. $\rho = \sigma$. □

Everything else in this section is a corollary.

Proposition 19.4 (Maximal entropy)**.** *$S(\rho) \leq \log d$, with equality iff $\rho = \mathbf{1}/d$.*

Proof. Take $\sigma = \mathbf{1}/d$ in Theorem 19.1: $0 \leq S(\rho \,\|\, \mathbf{1}/d) = \operatorname{Tr} \rho \log \rho - \operatorname{Tr} \rho\, (-\log d)\, \mathbf{1} = -S(\rho) + \log d$, with equality iff $\rho = \mathbf{1}/d$. □

Theorem 19.2 (Subadditivity)**.** *For any (pure or mixed) state ρ_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$, with reduced states ρ_A, ρ_B,*

$$S_{AB} \;\leq\; S_A + S_B, \tag{19.13}$$

with equality if and only if $\rho_{AB} = \rho_A \otimes \rho_B$.

Proof. Compare ρ_{AB} with the product of its own marginals, $\sigma = \rho_A \otimes \rho_B$. Since $\log(\rho_A \otimes \rho_B) = \log \rho_A \otimes \mathbf{1} + \mathbf{1} \otimes \log \rho_B$,

$$0 \;\leq\; S(\rho_{AB} \,\|\, \rho_A \otimes \rho_B) \;=\; -S_{AB} - \operatorname{Tr}\big(\rho_{AB} \log \rho_A \otimes \mathbf{1}\big) - \operatorname{Tr}\big(\rho_{AB}\, \mathbf{1} \otimes \log \rho_B\big). \tag{19.14}$$

By the defining property of the partial trace, $\operatorname{Tr}(\rho_{AB} \log \rho_A \otimes \mathbf{1}) = \operatorname{Tr}_A(\rho_A \log \rho_A) = -S_A$, and similarly for B. Hence $0 \leq -S_{AB} + S_A + S_B$, which is (19.13); equality holds iff $\rho_{AB} = \rho_A \otimes \rho_B$ by the equality case of Theorem 19.1. □

The combination appearing in the proof is important enough to have a name: the *mutual information*

$$I(A{:}B) \;\equiv\; S_A + S_B - S_{AB} \;=\; S(\rho_{AB} \,\|\, \rho_A \otimes \rho_B) \;\geq\; 0 \tag{19.15}$$

measures the total correlation—classical plus quantum—between A and B, vanishing exactly for uncorrelated systems.

Theorem 19.3 (Concavity)**.** *For any states ρ_i and probabilities p_i,*

$$S\Big(\sum_i p_i \rho_i\Big) \;\geq\; \sum_i p_i\, S(\rho_i). \tag{19.16}$$

Proof. We use an ancilla trick that will recur. Introduce an auxiliary system B with orthonormal states $|i\rangle_B$, one for each term, and consider on $\mathcal{H}_A \otimes \mathcal{H}_B$ the block-diagonal state

$$\rho_{AB} \;=\; \sum_i p_i\, \rho_i \otimes |i\rangle_B \langle i|_B. \tag{19.17}$$

Its eigenvalues are $\{p_i \lambda_k^{(i)}\}$, where $\lambda_k^{(i)}$ are the eigenvalues of ρ_i; hence

$$S_{AB} = -\sum_{i,k} p_i \lambda_k^{(i)} \log(p_i \lambda_k^{(i)}) = -\sum_i p_i \log p_i + \sum_i p_i\, S(\rho_i), \tag{19.18}$$

using $\sum_k \lambda_k^{(i)} = 1$ and $\sum_i p_i = 1$. The marginals are $\rho_A = \sum_i p_i\rho_i$ and $\rho_B = \mathrm{diag}(p_i)$, with $S_B = -\sum_i p_i \log p_i$. Subadditivity (19.13) then reads

$$-\sum_i p_i \log p_i + \sum_i p_i S(\rho_i) \ \le\ S\Big(\sum_i p_i\rho_i\Big) - \sum_i p_i \log p_i, \tag{19.19}$$

which is (19.16). □

Remark 19.1. Concavity says that *mixing increases entropy*: forgetting which ρ_i was prepared can only add uncertainty. There is a matching upper bound, $S(\sum_i p_i\rho_i) \le \sum_i p_i S(\rho_i) - \sum_i p_i \log p_i$, with equality when the ρ_i have orthogonal supports (manifest from (19.18) in that case); we will not need its general proof. This is a first quantitative glimpse of *coarse graining*: replacing a state by a mixture that encodes only partial knowledge raises the entropy, and by a controlled amount—at most the entropy of the mixing distribution.

19.5 Purification, the thermofield double, and Araki–Lieb

Property (iv) of Proposition 19.3 has a converse of great conceptual importance: mixedness can *always* be blamed on an unobserved partner.

Theorem 19.4 (Purification)**.** *Let ρ_A be any density matrix on $\mathcal{H}_A$. Then there exists a Hilbert space $\mathcal{H}_B$ with* $\dim \mathcal{H}_B = \mathrm{rank}\,\rho_A$ *and a* pure *state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ such that $\rho_A = \mathrm{Tr}_B\, |\Psi\rangle\langle\Psi|$. The purification is unique up to unitaries (more generally isometries, if* $\dim \mathcal{H}_B$ *is enlarged) acting on B alone.*

Proof. Existence is a one-line construction: write $\rho_A = \sum_k p_k |k\rangle_A\langle k|_A$ in spectral form, let $\{|k\rangle_B\}$ be an orthonormal basis of an auxiliary space of dimension equal to the number of nonzero p_k, and set

$$|\Psi\rangle \ =\ \sum_k \sqrt{p_k}\, |k\rangle_A \otimes |k\rangle_B. \tag{19.20}$$

Tracing out B manifestly returns ρ_A. For uniqueness: any purification, written in its Schmidt form (19.5), must have Schmidt coefficients $\sqrt{p_k}$ and A-vectors equal to the eigenvectors $|k\rangle_A$ (both are fixed by ρ_A); the only remaining freedom is the choice of the orthonormal set $\{|k\rangle_B\}$, and any two such choices are related by a unitary (or isometry) of $\mathcal{H}_B$. □

Example 19.1 (The thermofield double)**.** The canonical purification of the Gibbs state $\rho_\beta = \mathrm{e}^{-\beta H}/Z(\beta)$ is exactly the thermofield double state of Chapter 6,

$$|\Psi_{\mathrm{TFD}}\rangle \ =\ \frac{1}{\sqrt{Z(\beta)}} \sum_n \mathrm{e}^{-\beta E_n/2}\, |n\rangle_A \otimes |n\rangle_B, \tag{19.21}$$

i.e. the construction (19.20) applied to $p_n = \mathrm{e}^{-\beta E_n}/Z$. The entanglement entropy of either side equals the thermal entropy of one copy:

$$S_A \ =\ S_B \ =\ S(\rho_\beta) \ =\ \beta\langle H\rangle_\beta + \log Z(\beta), \tag{19.22}$$

the ordinary canonical entropy. This one formula ties together three lectures: the Rindler decomposition of the Minkowski vacuum (Chapter 6), the Hartle–Hawking state on the two-sided black hole (Chapter 7), and the eternal AdS black hole dual to two entangled CFTs (Chapter 18, [49]) are all instances of a thermal state presented through its purification. The statement at the end of Chapter 18—that the horizon area of the eternal black hole is the entanglement entropy between the two CFTs—now has a precise left-hand side: it is $S_A = S(\rho_\beta)$ in the state (19.21), and the claim is that this equals $A/4G$. Chapter 20 examines that claim.

The purification theorem is also a proof device: it lets us convert statements about mixed states into statements about pure states on a larger system, where the powerful symmetry $S_A = S_B$ is available. Here is the classic application.

Theorem 19.5 (Araki–Lieb triangle inequality)**.** *For any state ρ_{AB},*

$$|\, S_A - S_B \,| \;\leq\; S_{AB} \;\leq\; S_A + S_B. \tag{19.23}$$

Proof. The upper bound is Theorem 19.2. For the lower bound, purify: by Theorem 19.4 there is a system C and a pure state $|\Psi\rangle_{ABC}$ with $\mathrm{Tr}_C\,|\Psi\rangle\langle\Psi| = \rho_{AB}$. Purity of the global state gives, by Proposition 19.3(iv) applied to the splits $AB|C$ and $A|BC$,

$$S_{AB} = S_C, \qquad S_{BC} = S_A. \tag{19.24}$$

Now apply subadditivity to the pair (B, C): $S_{BC} \leq S_B + S_C$, i.e. $S_A \leq S_B + S_{AB}$, giving $S_A - S_B \leq S_{AB}$. Exchanging the roles of A and B (purify and use the pair (A, C)) gives $S_B - S_A \leq S_{AB}$. □

Remark 19.2. The lower bound replaces the classical intuition "the whole is at least as uncertain as any part," which is *false* quantum mechanically: a pure entangled state has $S_{AB} = 0$ with $S_A = S_B > 0$, saturating (19.23). What survives is that the whole cannot be less uncertain than the *imbalance* between the parts. We will use (19.23) in Section 19.10 in exactly this saturated regime: if black hole plus radiation is pure, the radiation entropy must precisely track the black hole's.

Finally, the deepest of the entropy inequalities, which we state without proof.

Theorem 19.6 (Strong subadditivity; Lieb–Ruskai)**.** *For any state ρ_{ABC} on a tripartite system,*

$$S_{ABC} + S_B \;\leq\; S_{AB} + S_{BC}. \tag{19.25}$$

Equivalently, mutual information is monotonic under discarding a subsystem, $I(A{:}B) \leq I(A{:}BC)$; equivalently again, relative entropy is monotonic under partial trace, $S(\rho_A \| \sigma_A) \leq S(\rho_{AB} \| \sigma_{AB})$.

- *Why it is hard.* Unlike everything proved so far, strong subadditivity (SSA) has no few-line proof (the original is [102]; a textbook account is [13]); the classical analogue is an easy exercise in Shannon theory, and the difficulty stems from the non-commutativity of the three reduced states. Its meaning is simple: *information cannot increase when a system is discarded.*
- *Where it appears.* It underlies the monotonicity of relative entropy, the engine of Casini's proof of the Bekenstein bound [103]; it constrains the entropies in the evaporation problem of Section 19.10; and its holographic counterpart becomes an almost trivial statement about minimal surfaces, striking evidence for the Ryu–Takayanagi formula (Chapter 21).

19.6 Thermal states maximize the entropy

Why is the Gibbs state $\rho_\beta = \mathrm{e}^{-\beta H}/Z$ the state of thermal equilibrium? The information-theoretic answer, due in essence to Gibbs and sharpened by Jaynes, is that it is the *least biased* state compatible with a known mean energy—and relative entropy turns this slogan into a theorem with a one-line proof.

Theorem 19.7 (Maximum-entropy property of the Gibbs state)**.** *Fix a Hamiltonian H and an energy E, and let β be chosen so that $\mathrm{Tr}(\rho_\beta H) = E$. Then among all density matrices ρ with $\mathrm{Tr}(\rho H) = E$, the Gibbs state ρ_β is the unique maximum of the von Neumann entropy.*

Proof. Since $\log \rho_\beta = -\beta H - \log Z(\beta)\,\mathbf{1}$, for *any* state ρ

$$S(\rho \,\|\, \rho_\beta) \;=\; -S(\rho) - \mathrm{Tr}\,(\rho \log \rho_\beta) \;=\; -S(\rho) + \beta\,\mathrm{Tr}(\rho H) + \log Z(\beta). \tag{19.26}$$

Evaluating the same identity at $\rho = \rho_\beta$ (where the left side vanishes) gives $S(\rho_\beta) = \beta E + \log Z$. Subtracting the two, for any ρ with $\mathrm{Tr}(\rho H) = E$, yields the boxed identity (19.27) below; positivity of the relative entropy (Theorem 19.1) and its equality case finish the proof. □

$$S(\rho) \;=\; S(\rho_\beta) \;-\; S(\rho \,\|\, \rho_\beta) \;\le\; S(\rho_\beta), \tag{19.27}$$

valid for any ρ with $\mathrm{Tr}(\rho H) = E$, with equality if and only if $\rho = \rho_\beta$: the relative entropy to the Gibbs state measures exactly the entropy deficit of a non-equilibrium state of the same energy.

Rearranging (19.26) without fixing the energy gives an equally useful statement about the free energy $F(\rho) \equiv \mathrm{Tr}(\rho H) - T\,S(\rho)$:

$$F(\rho) \;=\; F(\rho_\beta) + T\,S(\rho \,\|\, \rho_\beta) \;\ge\; F(\rho_\beta), \qquad F(\rho_\beta) = -T \log Z, \tag{19.28}$$

so the Gibbs state is also the unique global minimum of the free energy at fixed temperature. Equations (19.27) and (19.28) are the precise sense in which equilibrium thermodynamics is a variational consequence of quantum information theory.

19.7 Fine-grained versus coarse-grained entropy

- *The puzzle.* Take the tea-cup of Section 19.1 seriously: tea plus room plus everything they will ever interact with form a closed system in some state $\rho(t)$, evolving unitarily. By Proposition 19.3(ii),

$$S(\rho(t)) \;=\; S(\rho(0)) \qquad \text{for all } t: \tag{19.29}$$

  the fine-grained entropy never increases nor decreases. If the universe started in a pure state, its fine-grained entropy is zero now and forever. And yet the tea cools, and every thermodynamics textbook correctly asserts that entropy has increased. What increased?

- *Coarse graining.* The thermodynamic entropy of Chapter 2 is not $S(\rho)$. An experimenter does not know ρ; they measure a handful of macroscopic observables $\mathcal{A}_1, \dots, \mathcal{A}_n$, energies of tea and room, volumes, particle numbers, and assign the largest fine-grained entropy consistent with what they know:

Definition 19.5. Given a state ρ and a list of macroscopic observables $\{\mathcal{A}_j\}$, the *coarse-grained (thermodynamic) entropy* is

$$S_{\text{coarse}}(\rho) \;=\; \max_{\sigma} \Big\{ S(\sigma) \;:\; \mathrm{Tr}(\sigma \mathcal{A}_j) = \mathrm{Tr}(\rho\,\mathcal{A}_j) \;\; \forall j \Big\}. \tag{19.30}$$

- *Coarse graining loses information.* Since ρ itself satisfies the constraints,

$$S_{\text{coarse}}(\rho) \;\ge\; S(\rho): \tag{19.31}$$

  When the observables are the energies of weakly coupled subsystems, the maximizer in (19.30) is a product of Gibbs states by Theorem 19.7 (subsystem by subsystem, with subadditivity), and S_{coarse} is exactly the thermodynamic entropy of Chapter 2, the sum of the equilibrium entropies of tea and room at their current energies. For a system *in* global equilibrium, $\rho = \rho_\beta$, the two notions coincide; ordinary statistical mechanics is the special case in which coarse and fine agree.

- *What grows.* Out of equilibrium it is S_{coarse} that grows. Unitary evolution preserves $S(\rho)$ exactly but transfers information from the macroscopic observables into ever more complicated multi-body correlations that no feasible measurement tracks, the position of one tea molecule relative to an air molecule that scattered off the cup an hour ago. As the information migrates into these inaccessible correlations, the maximization in (19.30) is over an ever larger set and S_{coarse} rises toward its equilibrium value. That, and only that, is the Second Law.

> *Two notions of entropy.* The *fine-grained* (von Neumann, microscopic) entropy $S(\rho)$ is unitarily invariant: constant in time for a closed system, zero forever if the state is pure. It is the entropy that knows about information and unitarity. The *coarse-grained* (thermodynamic) entropy $S_{\text{coarse}} \geq S(\rho)$ retains only macroscopic data; it is the entropy that obeys the Second Law and the one measured by calorimetry. The two coincide in equilibrium and differ out of it. The Second Law is a statement about coarse-grained entropy; the information problem is a statement about fine-grained entropy.

- *Which is $A/4G$?* Part I argues for "thermodynamic": $A/4G$ enters the first law alongside T_H, obeys a Second Law (area theorem, GSL), and the Euclidean path integral of Chapter 7 computes it the way $\log Z$ computes any equilibrium entropy; Chapter 18 identified it as the coarse-grained entropy log(# of dual CFT states at energy M) of a unitary quantum system. Whether $A/4G$ *also* admits a fine-grained reading, as the entanglement entropy across the horizon, is the subject of Chapter 20. The distinction is not pedantry: the information problem is the statement that Hawking's calculation assigns the radiation a *fine*-grained entropy that eventually exceeds the black hole's *coarse*-grained one.

19.8 Rényi entropies and the replica idea

The von Neumann entropy sits inside a one-parameter family that is often easier to compute and carries strictly more information.

Definition 19.6. For $n > 0$, $n \neq 1$, the *Rényi entropy* of order n is

$$S_n(\rho) \;=\; \frac{1}{1-n}\,\log \operatorname{Tr}\rho^n \;=\; \frac{1}{1-n}\,\log \sum_i p_i^n. \tag{19.32}$$

For integer $n \geq 2$, $\operatorname{Tr}\rho^n$ is a polynomial in the state—for $n = 2$ it is the purity of Proposition 19.1—and this is what makes the family computable. The main properties:

Proposition 19.5. *(i) $S_n(\rho) \geq 0$, and S_n is invariant under $\rho \to U\rho U^\dagger$. (ii) $S_n(\rho)$ is non-increasing in n. (iii) The limits of the family recover, respectively, the von Neumann entropy, the logarithm of the rank, and the largest eigenvalue:*

$$\lim_{n\to 1} S_n(\rho) = S(\rho), \qquad \lim_{n\to 0} S_n(\rho) = \log \operatorname{rank}\rho, \qquad \lim_{n\to\infty} S_n(\rho) = -\log p_{\max}. \tag{19.33}$$

Proof of the $n \to 1$ limit. Write $\operatorname{Tr}\rho^n = \sum_i p_i^n$ and expand about $n = 1$: $p_i^n = p_i\,\mathrm{e}^{(n-1)\log p_i} = p_i + (n-1)\,p_i \log p_i + O((n-1)^2)$, so $\operatorname{Tr}\rho^n = 1 - (n-1)S(\rho) + O((n-1)^2)$ and

$$S_n = \frac{\log\left[1 - (n-1)S(\rho) + \cdots\right]}{1-n} \xrightarrow{\;n\to 1\;} S(\rho). \tag{19.34}$$

The other statements follow from elementary manipulations of $\sum_i p_i^n$ (for (ii), one checks $\mathrm{d}S_n/\mathrm{d}n \leq 0$ using Jensen's inequality on the escort distribution $p_i^n/\sum_j p_j^n$); we omit the details. $\square$

- *The replica method.* In any interesting field-theory or gravitational problem the spectrum of ρ_A is inaccessible, so (19.7) cannot be evaluated directly. But for integer n,

$$\operatorname{Tr}\rho_A^n \;=\; (\text{partition function on an } n\text{-sheeted branched cover}), \tag{19.35}$$

because ρ_A is computed by a Euclidean path integral with a cut along A (for Rindler space, Chapter 6, the path integral on a wedge of angle θ built $e^{-\theta K}$), and multiplying n such matrices and taking the trace glues n copies of the geometry cyclically along the cut. One computes (19.35) for every integer $n \geq 2$, continues analytically in n (the spectrum is uniquely determined by its integer moments in finite dimensions; in general the continuation requires care), and extracts the von Neumann entropy from (19.34), most conveniently as

$$S(\rho_A) \;=\; -\,\partial_n\big[\log\operatorname{Tr}\rho_A^n\big]_{n=1} \;=\; \Big(1 - n\,\partial_n\Big)\log\operatorname{Tr}\rho_A^n\,\Big|_{n=1}. \tag{19.36}$$

- *Where it leads.* Developed systematically in two-dimensional CFT by Calabrese and Cardy [104], the trick becomes powerful in gravity: the branched cover can be filled in by a smooth bulk geometry, and evaluating the gravitational action on it and continuing $n \to 1$ is how Lewkowycz and Maldacena derived the Ryu–Takayanagi formula [45] (Chapter 21). The logic to remember: *integer moments of ρ_A are partition functions; entropy is their derivative at $n = 1$.*

19.9 Entanglement entropy in quantum field theory

Before turning to black holes we must confess what happens to the above in quantum field theory. Let A be a region of space, the interior of a sphere or a half-space, and ρ_A the reduced state of the vacuum.

In quantum field theory the entanglement entropy of any spatial region is *ultraviolet divergent*. With a short-distance cutoff ϵ, the leading divergence in d spacetime dimensions is proportional to the *area* of the entangling surface,

$$S_A \;=\; c_0\,\frac{\operatorname{Area}(\partial A)}{\epsilon^{\,d-2}} \;+\; (\text{subleading}), \tag{19.37}$$

with c_0 a non-universal constant depending on the cutoff scheme and field content. Fine-grained entropy in QFT is cutoff-dependent; only suitable differences and combinations (mutual information, relative entropy, and—crucially for Chapter 20—the *generalized* entropy) are finite.

- *Origin of the divergence.* The vacuum entangles field modes on the two sides of ∂A at all wavelengths: for each patch of the entangling surface and each scale λ, the modes of wavelength $\sim \lambda$ straddling the surface contribute $O(1)$ entanglement per patch of area λ^{d-2}. Summing over scales from the infrared down to the cutoff,

$$S_A \;\sim\; \int_\epsilon \frac{\mathrm{d}\lambda}{\lambda}\,\frac{\operatorname{Area}(\partial A)}{\lambda^{\,d-2}} \;\sim\; \frac{\operatorname{Area}(\partial A)}{\epsilon^{\,d-2}}, \tag{19.38}$$

which is (19.37): the entropy is dominated by the shortest distances and scales with the *area*, not the volume, because vacuum entanglement is local. The area law was found numerically for a free scalar by Srednicki [105], and earlier, with black holes in mind, by Bombelli, Koul, Lee and Sorkin [106].

- *Two exact cases.* For the half-space Chapter 6 gives $\rho_A = e^{-2\pi K}/Z$ exactly, and turning this into controlled computations of S_A, and into the identification of the cutoff-*independent* combination, is the business of Chapter 20. In two-dimensional CFT the counterpart of the area law is the formula for an interval of length ℓ, $S = \frac{c}{3}\log(\ell/\epsilon)$: the "area" of the entangling surface is a point, the divergence is logarithmic, and its coefficient is *universal*, the central charge. Chapter 21 derives it twice, by the replica method in the CFT and holographically as the length of a geodesic [104].

- *A familiar pattern.* An entropy proportional to the area of a horizon-like surface, diverging as the cutoff is removed: if the effective cutoff of quantum gravity is the Planck length, $\epsilon \sim \ell_P$, the vacuum entanglement across a horizon is of order $A/\ell_P^2 \sim A/G$, parametrically the Bekenstein–Hawking entropy. How the divergence is absorbed into the renormalization of Newton's constant, so that the *generalized* entropy $S_{\text{gen}} = A/4G + S_{\text{out}}$ is cutoff-independent, is the main business of Chapter 20.

- *A caution.* In QFT the operator ρ_A, and even the factorization (19.3), do not strictly exist (the algebra of a region is of von Neumann type III, admitting no trace). The cutoff-regularized language suffices for this course, but the divergence in (19.37) is the smoke of a deeper structural fire; see [1].

19.10 The Page curve

We now assemble the tools into the sharpest available statement of the black hole information problem. The key input is a beautiful result of Page about a deceptively elementary question: *how entangled is a typical pure state?*

19.10.1 Page's theorem: the entropy of a random subsystem

Let $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ with $\dim \mathcal{H}_A = m \le \dim \mathcal{H}_B = n$, and draw a pure state $|\psi\rangle \in \mathcal{H}$ *at random*, uniformly with respect to the unitarily invariant (Haar) measure on the unit sphere[1], concretely $|\psi\rangle = U|\psi_0\rangle$ with U a Haar-random unitary and ψ_0 arbitrary. What is the expected entanglement entropy of A?

Theorem 19.8 (Page [107])**.** *For $1 \le m \le n$, the average over the Haar measure of the entanglement entropy of subsystem A is exactly*

$$\langle S_A \rangle \;=\; \sum_{k=n+1}^{mn} \frac{1}{k} \;-\; \frac{m-1}{2n} \;=\; \log m \;-\; \frac{m}{2n} \;+\; O\Big(\frac{1}{mn}, \frac{m}{n^2}\Big). \tag{19.39}$$

Page conjectured the closed form in (19.39) and proved its asymptotics; the exact formula was proved by Foong and Kanno [108] and, more simply, by Sen [109]. We do not reproduce the exact computation. Instead we prove the two bounds

$$\log m - \frac{m}{n} \;\le\; \langle S_A \rangle \;\le\; \log m, \tag{19.40}$$

which capture the content of the theorem up to the factor $\frac{1}{2}$ in the deficit, by an argument that uses only the second Rényi entropy.

Proof of (19.40). ⋆ *Self-study.* The upper bound holds state by state, $S_A \le \log m$ by Proposition 19.4, hence on average. The lower bound takes four steps.

[1]The Haar measure on $U(d)$ is the unique probability measure invariant under $U \to VU$ and $U \to UV$; the induced measure on $|\psi\rangle = U|\psi_0\rangle$ is the uniform measure on the unit sphere $S^{2d-1} \subset \mathbb{C}^d$, equivalently $\psi = z/\|z\|$ with the z_i independent complex Gaussians. Its moments follow by symmetry and normalization, e.g. $\langle |\psi_i|^2 \rangle = 1/d$ and $\langle |\psi_i|^2 |\psi_j|^2 \rangle = (1+\delta_{ij})/d(d+1)$, the component form of (19.43).

1. *Reduce to the purity.* By Proposition 19.5(ii) the Rényi entropies are non-increasing in their index, so $S_A \geq S_2(\rho_A) = -\log \operatorname{Tr} \rho_A^2$. Since $-\log$ is convex, Jensen's inequality gives $\langle -\log \operatorname{Tr} \rho_A^2 \rangle \geq -\log \langle \operatorname{Tr} \rho_A^2 \rangle$, and therefore

$$\langle S_A \rangle \;\geq\; -\log \big\langle \operatorname{Tr} \rho_A^2 \big\rangle. \tag{19.41}$$

It remains to compute the average purity.

2. *Write the purity as a trace over two copies.* Let $\mathbb{F}_A$ be the operator on $\mathcal{H}_A \otimes \mathcal{H}_A$ that swaps the two factors, $\mathbb{F}_A |i\rangle \otimes |j\rangle = |j\rangle \otimes |i\rangle$. For any operator X on $\mathcal{H}_A$, $\operatorname{Tr}\left[(X \otimes X)\,\mathbb{F}_A\right] = \sum_{ij} X_{ij} X_{ji} = \operatorname{Tr} X^2$. Applying this to $X = \rho_A = \operatorname{Tr}_B \rho$ and noting that $\rho_A \otimes \rho_A = \operatorname{Tr}_{BB}(\rho \otimes \rho)$, where Tr_{BB} traces over both copies of $\mathcal{H}_B$,

$$\operatorname{Tr} \rho_A^2 \;=\; \operatorname{Tr}\Big[(\rho \otimes \rho)\,(\mathbb{F}_A \otimes \mathbf{1}_{BB})\Big]. \tag{19.42}$$

This is the *swap trick*: it makes the purity linear in $\rho \otimes \rho$, so only the average of $\rho \otimes \rho$ is needed.

3. *Average $\rho \otimes \rho$ over the Haar measure.* Write $d = mn$ and let $\mathbb{F}$ swap the two copies of the full space $\mathcal{H}$. Since $|\psi\rangle \otimes |\psi\rangle$ is symmetric under $\mathbb{F}$, so is its average, and the average commutes with every $U \otimes U$ because the Haar measure is invariant. The symmetric subspace of $\mathcal{H} \otimes \mathcal{H}$ is an irreducible representation of $U \otimes U$, so by Schur's lemma the average is proportional to the projector $\frac{1}{2}(\mathbf{1} + \mathbb{F})$ onto it. The constant follows from taking the trace, $\operatorname{Tr} \frac{1}{2}(\mathbf{1} + \mathbb{F}) = \frac{1}{2} d(d+1)$ while $\operatorname{Tr}(\rho \otimes \rho) = 1$:

$$\langle \rho \otimes \rho \rangle \;=\; \frac{\mathbf{1} + \mathbb{F}}{d(d+1)}. \tag{19.43}$$

4. *Evaluate.* Insert (19.43) into (19.42). The two terms are $\operatorname{Tr}\left[\mathbb{F}_A \otimes \mathbf{1}_{BB}\right] = \operatorname{Tr} \mathbb{F}_A \cdot \operatorname{Tr} \mathbf{1}_{BB} = m \cdot n^2$ and, using $\mathbb{F} = \mathbb{F}_A \otimes \mathbb{F}_B$ and $\mathbb{F}_A^2 = \mathbf{1}$, $\operatorname{Tr}\left[\mathbb{F}\,(\mathbb{F}_A \otimes \mathbf{1}_{BB})\right] = \operatorname{Tr} \mathbf{1}_{AA} \cdot \operatorname{Tr} \mathbb{F}_B = m^2 \cdot n$. Hence

$$\big\langle \operatorname{Tr} \rho_A^2 \big\rangle \;=\; \frac{mn^2 + m^2 n}{mn(mn+1)} \;=\; \frac{m+n}{mn+1}, \tag{19.44}$$

and (19.41) gives

$$\langle S_A \rangle \;\geq\; \log \frac{mn+1}{m+n} \;>\; \log \frac{m}{1 + m/n} \;\geq\; \log m - \frac{m}{n}, \tag{19.45}$$

using $\log(1+x) \leq x$ in the last step.

□

Page's exact formula (19.39) improves the deficit in (19.45) from m/n to $m/2n$. And the statement is not only about the average: on a sphere of large dimension a smooth function such as S_A takes values close to its mean on all but an exponentially small fraction of the sphere (concentration of measure), so the probability that a random state has S_A below $\log m - m/2n - \delta$ is exponentially small in mn [110]. *Typical* states, not just the average, have $S_A \simeq \log m - m/2n$.

Page's lesson. Let AB be in a random pure state, with $\dim \mathcal{H}_A = m \leq \dim \mathcal{H}_B = n$. Then A is almost exactly maximally mixed: its entropy falls short of the maximum $\log m$ by only $m/2n$ on average, and by appreciably more only with probability exponentially small in mn. The smaller subsystem therefore carries essentially no information about the global state; the information, and the purity of the whole, reside in the correlations between A and B and within B.

19.10.2 A random-state model of black hole evaporation

Following Page [111], let a black hole form by collapse from a *pure* state and evaporate over the lifetime $t_{\rm evap} \sim G^2 M_0^3$ of (4.28). Suppose that the global state stays pure at all times; this is the unitarity hypothesis to be tested against Hawking's calculation. At retarded time t, split the Hilbert space in the coarse sense of Section 19.9 into the hole and the radiation emitted so far,

$$\mathcal{H} \supset \mathcal{H}_{\rm BH}(t) \otimes \mathcal{H}_{\rm rad}(t), \qquad \dim \mathcal{H}_{\rm BH} \simeq e^{S_{\rm BH}(t)}, \quad \dim \mathcal{H}_{\rm rad} \simeq e^{S^{\rm th}_{\rm rad}(t)}. \tag{19.46}$$

What is known and what is asked. Two entropies are thermodynamic, known functions of time fixed by Part I: $S_{\rm BH}(t) = A(t)/4G$ for the hole and $S^{\rm th}_{\rm rad}(t)$ for the radiation emitted so far. The question of this section is the value of a third, the fine-grained (von Neumann) entropy of the radiation, $S_{\rm rad}(t) \equiv S(\rho_{\rm rad}(t))$. Two candidate answers will be compared: Hawking's, $S_{\rm rad} = S^{\rm th}_{\rm rad}$, and the unitary one, the Page curve. Notation: $S(\rho)$ with a density matrix as argument is always von Neumann; $S_{\rm BH}$ and $S^{\rm th}_{\rm rad}$ are always thermodynamic.

The three entropies in detail:

1. $S_{\rm BH}(t) = A(t)/4G$, the thermodynamic entropy of the hole. With $S_{\rm BH} \propto M^2$ and M^3 decreasing linearly in time, (4.28),

$$S_{\rm BH}(t) = S_{\rm BH}(0)\,(1 - t/t_{\rm evap})^{2/3}, \tag{19.47}$$

   decreasing monotonically to zero.

2. $S^{\rm th}_{\rm rad}(t)$, the thermodynamic entropy of the radiation emitted up to time t. Quasi-stationary emission produces it at a fixed rate relative to what the hole loses, $dS^{\rm th}_{\rm rad} = \beta\,|dS_{\rm BH}|$, so

$$S^{\rm th}_{\rm rad}(t) = \beta\,[S_{\rm BH}(0) - S_{\rm BH}(t)], \tag{19.48}$$

   increasing monotonically to $\beta\, S_{\rm BH}(0)$. For black-body photons $\beta = \frac{4}{3}$, the factor of Chapter 2; for the actual grey-body emission of photons and gravitons $\beta \approx 1.48$ [112].

3. $S_{\rm rad}(t)$, the fine-grained entropy of the radiation, the unknown. Since the global state is pure, Proposition 19.3(iv) gives

$$S_{\rm rad}(t) \;=\; S(\rho_{\rm hole}(t)). \tag{19.49}$$

The Page curve. Two inequalities hold without any further assumption.

1. $S_{\rm rad} \le S^{\rm th}_{\rm rad}$: coarse-graining only increases entropy, (19.31).

2. $S_{\rm rad} = S(\rho_{\rm hole}) \le \log\dim\mathcal{H}_{\rm BH} \simeq S_{\rm BH}$, by (19.49) and Proposition 19.3(iii), provided the hole has only $e^{S_{\rm BH}}$ states.

Hence $S_{\rm rad} \le \min(S^{\rm th}_{\rm rad}, S_{\rm BH})$. If, moreover, evaporation is a complicated but ordinary unitary process, so that the global state is for entropy purposes a random state in (19.46), Page's theorem (19.39) says that the smaller factor is maximally mixed up to $O(1)$, and the bound is saturated:

The Page curve. Under unitarity and the state counting $\dim\mathcal{H}_{\rm BH} = e^{S_{\rm BH}}$, the fine-grained entropy of the radiation is

$$S_{\rm rad}(t) \;\simeq\; \min\Big(S^{\rm th}_{\rm rad}(t)\,,\; S_{\rm BH}(t)\Big) \tag{19.50}$$

up to $O(1)$: it rises with Hawking's thermal growth while the radiation is the smaller factor, and falls with the horizon area once it is the larger.

The Page time. The turnover is at

$$S_{\mathrm{rad}}^{\mathrm{th}}(t_{\mathrm{Page}}) \;=\; S_{\mathrm{BH}}(t_{\mathrm{Page}}). \tag{19.51}$$

Inserting (19.47) and (19.48),

$$\frac{S_{\mathrm{BH}}(t_{\mathrm{Page}})}{S_{\mathrm{BH}}(0)} = \frac{\beta}{1+\beta}, \qquad \frac{M(t_{\mathrm{Page}})}{M_0} = \left(\frac{\beta}{1+\beta}\right)^{1/2}, \qquad \frac{t_{\mathrm{Page}}}{t_{\mathrm{evap}}} = 1 - \left(\frac{\beta}{1+\beta}\right)^{3/2}, \tag{19.52}$$

which for $\beta \approx 1.48$ [112] gives

$$t_{\mathrm{Page}} \approx 0.54\, t_{\mathrm{evap}}, \qquad M(t_{\mathrm{Page}}) \approx 0.77\, M_0, \qquad S_{\mathrm{BH}}(t_{\mathrm{Page}}) \approx 0.60\, S_{\mathrm{BH}}(0) \tag{19.53}$$

(and $t_{\mathrm{Page}} \approx 0.57\, t_{\mathrm{evap}}$ for $\beta = \frac{4}{3}$). The curve is plotted in Figure 19.1. Two features matter for what follows.

- Before t_{Page} the radiation is exponentially close to maximally mixed by Page's theorem, so it contains essentially no information about the initial state; information emerges only after t_{Page}, in correlations between early and late quanta, and the last of it only as the hole disappears [111].
- At t_{Page} the hole still has three quarters of its mass and negligible curvature at the horizon: whatever enforces the descent cannot be a Planck-scale effect near the singularity or at the end of evaporation.

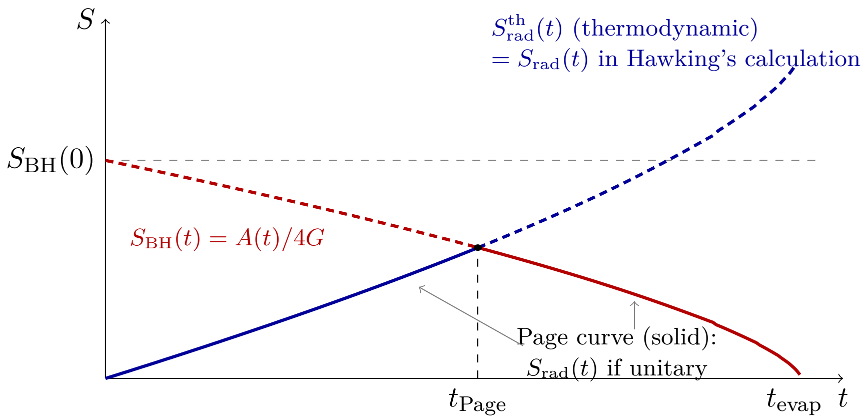


Figure 19.1: The two known thermodynamic entropies: $S_{\mathrm{rad}}^{\mathrm{th}}$ of the emitted radiation (blue), rising to about $1.5\, S_{\mathrm{BH}}(0)$, and S_{BH} of the shrinking hole (red); Schwarzschild rates, schematic units. Each curve is drawn solid where it is the smaller of the two and dashed where it is the larger. The unknown is the fine-grained entropy $S_{\mathrm{rad}}(t)$. Hawking's answer is the entire blue curve, solid and dashed, for all t. The unitary answer, if the hole has $\mathrm{e}^{S_{\mathrm{BH}}}$ states, is the minimum of the two, i.e. the solid parts: the Page curve, rising until $t_{\mathrm{Page}} \approx 0.54\, t_{\mathrm{evap}}$ and then falling to zero.

19.10.3 The information problem in entropy language

The black hole information problem is a conflict between two computations of one quantity, the fine-grained entropy $S_{\mathrm{rad}}(t)$ of the Hawking radiation.

- *Hawking's answer* [26]. In the semiclassical calculation of Chapter 4 each outgoing mode is emitted in the two-mode entangled state (4.26) with a partner behind the horizon; tracing out the partners, the radiation is thermal mode by mode, with no correlations between successive quanta. Then $S_{\mathrm{rad}}(t) = S^{\mathrm{th}}_{\mathrm{rad}}(t)$, rising monotonically for the whole lifetime, and if the hole evaporates completely the final state is mixed: a pure state has evolved into a thermal density matrix, which no unitary evolution can do (Proposition 19.3(ii)).

- *The unitary answer.* If gravity is an ordinary quantum theory, as Chapter 18 showed for AdS black holes, the global state stays pure; if moreover the hole has $\mathrm{e}^{A/4G}$ states, then $S_{\mathrm{rad}} \leq S_{\mathrm{BH}}$ and the Page curve (19.50) follows.

The information problem, sharply. For $t > t_{\mathrm{Page}}$, Hawking's calculation gives $S_{\mathrm{rad}}(t) = S^{\mathrm{th}}_{\mathrm{rad}}(t) > S_{\mathrm{BH}}(t)$: the fine-grained entropy of the radiation exceeds the number of states of the black hole it is entangled with. This is impossible for a unitary quantum system with $\mathrm{e}^{A/4G}$ states. The conflict is of order $S_{\mathrm{BH}} \sim 1/G$, not a small correction, and it arises at t_{Page}, when the geometry is weakly curved everywhere and semiclassical gravity should be reliable. One of three things must give: unitarity fails in gravity (Hawking's conclusion); the black hole has far more than $\mathrm{e}^{A/4G}$ states; or the semiclassical computation of the *fine-grained* entropy of the radiation receives $O(1/G)$ corrections at the Page time even though all local physics stays semiclassical.

Nothing is wrong with the two curves of Figure 19.1: both are thermodynamic entropies, and $S^{\mathrm{th}}_{\mathrm{rad}}$ should rise, by the generalized second law of Section 2.3. The disagreement is only about S_{rad}, a fine-grained entropy: Hawking's calculation puts it on the entire blue curve, while unitarity puts it on the Page curve, the solid blue and solid red pieces, and for $t > t_{\mathrm{Page}}$ the two are incompatible. This is why the distinction of Section 19.7 is the heart of the matter. The eternal black hole has the same tension in another form: semiclassical two-sided correlators in the thermofield double of Chapter 18 decay forever, whereas in a unitary theory with a discrete spectrum they must eventually fluctuate back up [49].

A recent proposal. Since 2019 a concrete proposal for the third alternative has been developed [113, 114], reviewed in [115]: the fine-grained entropy of the radiation is to be computed not from Hawking's density matrix but from the quantum extremal surface prescription of Chapter 21, equation (21.42), extended to allow a disconnected *island* inside the black hole to count as part of the radiation system. In this prescription the configuration without an island reproduces Hawking's rising curve, an island covering most of the interior takes over after t_{Page} and tracks $A(t)/4G$, and the minimum of the two is (19.50).[2]

Whatever its ultimate status, the island proposal presupposes that an area divided by $4G$ can appear inside a fine-grained entropy. In what sense $A/4G$ is an entanglement entropy is the question of the next chapter, where the divergence (19.37) of the entanglement across a horizon combines with the renormalization of Newton's constant into the cutoff-independent generalized entropy $S_{\mathrm{gen}} = A/4G + S_{\mathrm{out}}$.

[2] The derivation applies the replica method of Section 19.8 to the gravitational path integral, where new saddles, replica wormholes, dominate after the Page time. Three caveats: the controlled computations are mostly in two-dimensional dilaton gravity coupled to a bath, the higher-dimensional case resting on additional assumptions; the prescription reproduces the Page curve without exhibiting the microstates or saying how the information is encoded in the radiation; and its status is debated, so the reader should regard islands as the current proposal, not the final word.

Chapter 20

Black Hole Entropy as Entanglement Entropy

In what sense is $A/4G$ an entanglement entropy? We state the proposal of Bombelli–Koul–Lee–Sorkin and Srednicki, then compute the entanglement of a free scalar across a planar cut—replica trick, cone, heat kernel—obtaining $S = A/48\pi\epsilon^2$ in four dimensions. The divergent coefficient depends on the regularization, but the *same* heat-kernel coefficient renormalizes Newton's constant: the generalized entropy $S_{\text{gen}} = A/4G + S_{\text{out}}$ is cutoff-independent, with entanglement entropy revealed as the matter loop correction to $A/4G$. We then give Casini's relative-entropy proof of the Bekenstein bound, turning the box-lowering heuristic of Chapter 2 into a theorem of quantum field theory, and show that the gravitational path integral computes S_{gen} automatically.

20.1 The proposal: horizon entropy from vacuum entanglement

- *The reduced state of the exterior.* Consider a quantum field on a black hole background in the Hartle–Hawking state of Chapter 7, and a Cauchy slice of the extended geometry through the bifurcation surface, divided by the horizon into an exterior region and its complement. An exterior observer describes physics by

$$\rho_{\text{out}} \;=\; \text{Tr}_{\text{in}}\, |\text{HH}\rangle\langle\text{HH}|, \tag{20.1}$$

  whose von Neumann entropy $S_{\text{out}} = S(\rho_{\text{out}})$ is an entanglement entropy in the strict sense of Chapter 19: the global state is pure, and the mixedness is due entirely to correlations across the horizon.

- *Two facts* suggest connecting S_{out} with the Bekenstein–Hawking entropy. S_{out} is divergent with an *area-law* leading behaviour, $S_{\text{out}} \sim \text{Area}\,/\epsilon^{d-2}$ (19.37): the divergence comes from short-wavelength entangled pairs straddling the horizon, so it is local on the horizon and proportional to its area. And if the effective cutoff of quantum gravity is the Planck length, $\epsilon \sim \ell_P = \sqrt{G}$ in four dimensions,

$$S_{\text{out}} \;\sim\; \frac{A}{\ell_P^2} \;\sim\; \frac{A}{G}, \tag{20.2}$$

  parametrically the Bekenstein–Hawking entropy. The observation was made with black holes in mind by Bombelli, Koul, Lee and Sorkin [106], and independently by Srednicki [105], who traced the reduced density matrix of a free scalar outside an imaginary sphere in flat space, with no black hole at all, and found numerically $S \approx 0.30\,(R/\epsilon)^2$ for a sphere of radius R with lattice cutoff ϵ: an area law from vacuum entanglement alone.

The proposal, in its boldest form, is that this is not a coincidence:

> **Entanglement interpretation of black hole entropy.** The Bekenstein–Hawking entropy $A/4G$ is the entanglement entropy of the quantum fields across the horizon, cut off at the Planck scale—or, in the refined form we will reach in Section 20.4, the divergent entanglement entropy of matter across the horizon is the one-loop *renormalization* of $A/4G$, and the physically meaningful, cutoff-independent quantity is the generalized entropy $S_{\text{gen}} = A/4G + S_{\text{out}}$.

To assess this we need the entanglement entropy of a horizon cut as an actual formula rather than a scaling estimate, starting from the one horizon we control completely.

20.2 The Rindler wedge revisited

Every horizon is, sufficiently close up, a Rindler horizon: the near-horizon geometry of a non-extremal black hole is Rindler space times the horizon sphere (Chapter 1), and the Hartle–Hawking state restricted to the near-horizon region is the Minkowski vacuum restricted to a wedge (Chapter 7). The computation to do is therefore the entanglement entropy of the Minkowski vacuum across the plane $x = 0$, and Chapter 6 has done the hard part: the reduced density matrix of the right wedge is the boost-thermal state (6.9),

$$\rho_r \;=\; \frac{\mathrm{e}^{-2\pi K}}{Z}, \qquad K \;=\; \int_{x>0} \mathrm{d}x\, \mathrm{d}^{d-2}y \;\, x\, T_{00}(x, \vec{y}), \qquad Z = \mathrm{Tr}\; \mathrm{e}^{-2\pi K}, \tag{20.3}$$

with K the boost generator (6.8) restricted to the wedge (we drop the subscript r on K from now on). Its von Neumann entropy has the thermal form

$$S(\rho_r) \;=\; -\,\mathrm{Tr}\,\rho_r \log \rho_r \;=\; 2\pi\, \langle K \rangle_{\rho_r} \;+\; \log Z, \tag{20.4}$$

i.e. $S = \beta\langle H\rangle + \log Z$ at the modular temperature $1/2\pi$, the entanglement analogue of (19.22). The formula is exact for any theory in any dimension, but both terms are ultraviolet divergent, and the divergence sits at the entangling surface.

1. The weight x in K vanishes at $x = 0$, so modes of arbitrarily short wavelength near the entangling surface contribute at modular temperature $1/2\pi$; by the Tolman law of Chapter 6 this is a local temperature $T(\xi) = 1/2\pi\xi$ at proper distance ξ from the horizon, which diverges as $\xi \to 0$.

2. Treating each slice at proper distance ξ as a thermal gas at temperature $T(\xi)$, with the flat-space entropy density $s = \frac{2\pi^2}{45}\, g\, T^3$ of g massless bosonic degrees of freedom in $d = 4$, and cutting off at $\xi = \epsilon$,

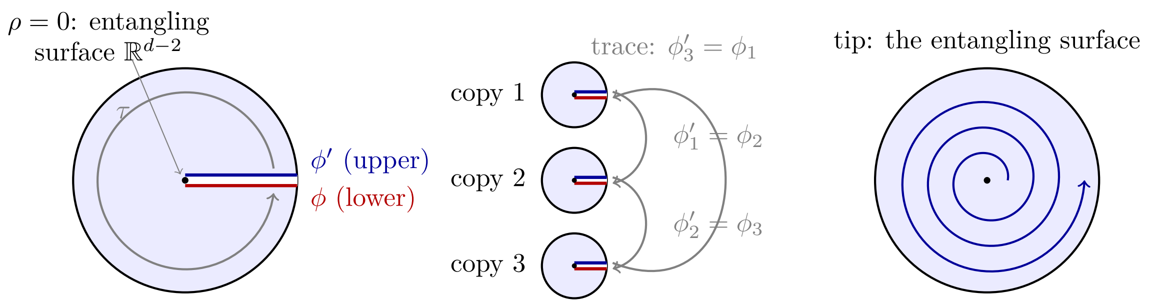
$$S \;=\; A \int_\epsilon^\infty \mathrm{d}\xi\; \frac{2\pi^2}{45}\, g \left(\frac{1}{2\pi\xi}\right)^3 \;=\; \frac{g}{360\pi}\, \frac{A}{\epsilon^2}. \tag{20.5}$$

The estimate places the divergence at the horizon with the area as coefficient, but the coefficient depends on the cutoff prescription, and the estimate says nothing about how the entropy enters the gravitational action. Both questions need the replica method of Section 19.8.

20.3 A sample computation: a free scalar across a planar cut

We now compute, from first principles and keeping every factor, the entanglement entropy of a free massless scalar field in d-dimensional Minkowski space, reduced to the half-space $x > 0$. Following (19.36), the strategy is

$$S \;=\; (1 - n\,\partial_n)\, \log Z_n \Big|_{n=1}, \tag{20.6}$$



(a) one copy, $\langle\phi'|\rho_r|\phi\rangle$: the plane cut along $x > 0$, angle τ from 0 to 2π

(b) $\mathrm{Tr}\,\rho_r^3$: the upper edge of each copy is glued to the lower edge of the next, and the last to the first

(c) the cone $C_{2\pi n}$: a single surface on which a loop around the tip closes only after an angle $2\pi n$

Figure 20.1: The replica construction for the Rindler cut. (a) One copy: the matrix element $\langle\phi'|\rho_r|\phi\rangle$ is the path integral on the plane cut along $x > 0$, with the field fixed to ϕ' on the upper edge of the cut and to ϕ on the lower edge; the entangling surface is the origin $\rho = 0$ (times $\mathbb{R}^{d-2}$). (b) The product ρ_r^n (here $n = 3$) identifies the upper edge of each copy with the lower edge of the next, and the trace identifies the last with the first. (c) The result is a single surface, the cone $C_{2\pi n}$: flat everywhere except at the tip, around which a loop closes only after an angle $2\pi n$. For $n \geq 2$ this is an angular *excess*, so the surface cannot be drawn as an ordinary cone in three dimensions; the spiral shows the three turns.

where Z_n is the Euclidean path integral on the n-fold branched cover of $\mathbb{R}^d$ over the entangling surface; we must construct Z_n, compute it, and continue in n. We compute Z_n with the heat kernel because the same object, expanded in curvature, will renormalize $1/G$ in Section 20.4, and the two coefficients must be compared in one regularization.

20.3.1 The replica geometry: a cone

Recall the path-integral representation of ρ_r from Chapter 6 (Figure 6.3): working in polar coordinates (ρ, τ) on the Euclidean (x, t_E) plane, with the flat metric

$$\mathrm{d}s^2 \;=\; \mathrm{d}\rho^2 + \rho^2\,\mathrm{d}\tau^2 + \mathrm{d}\vec{y}^{\,2}, \qquad \rho \geq 0, \tag{20.7}$$

the matrix element $\langle\phi'|\rho_r|\phi\rangle$ is the path integral over the full angular range $\tau \in [0, 2\pi]$ with the fields on the two edges of the cut fixed to ϕ and ϕ'. Multiplying n copies of ρ_r glues the upper edge of each wedge to the lower edge of the next, and taking the trace glues the last to the first. The result is the path integral on a *cone*: the geometry (20.7) with the angular coordinate now of period $2\pi n$,

$$\mathrm{Tr}\,\rho_r^n \;=\; \frac{Z_n}{Z_1^n}, \qquad Z_n \;\equiv\; Z\big[\,C_{2\pi n} \times \mathbb{R}^{d-2}\,\big], \tag{20.8}$$

where C_δ denotes the two-dimensional cone of total opening angle δ (flat everywhere except the tip, where the geometry is singular for $\delta \neq 2\pi$), and $\mathbb{R}^{d-2}$ is the entangling surface itself, the transverse plane spanned by $\vec{y}$. The denominator Z_1^n normalizes $\mathrm{Tr}\,\rho_r = 1$. Since $\log Z_1$ enters (20.6) only through the combination $(1 - n\partial_n)\, n \log Z_1 = 0$, we may work directly with (20.6) as written. Figure 20.1 summarizes the construction.

For integer $n \geq 2$ the cone has an angular excess; the continuation to $n \to 1$ uses that C_δ makes sense for any $\delta > 0$. The replica trick has thus turned the entanglement entropy into the response of a partition function to a conical deformation of the geometry, the same object from which the black hole entropy was extracted in Section 7.6; Section 20.4 makes this parallel quantitative.

20.3.2 The partition function via the heat kernel

For a free scalar, the partition function is a Gaussian integral,

$$\log Z_n \;=\; -\tfrac{1}{2}\,\log\det\big(-\nabla^2\big) \;=\; \tfrac{1}{2}\int_{\epsilon^2}^{\infty}\frac{\mathrm{d}t}{t}\;\mathrm{Tr}\;\mathrm{e}^{t\nabla^2}, \tag{20.9}$$

where in the second step we used the proper-time representation $\log\lambda = -\int_0^\infty(\mathrm{d}t/t)\,\mathrm{e}^{-\lambda t}$ (up to a λ-independent constant), and regulated the ultraviolet by cutting the proper-time integral off at $t=\epsilon^2$; the parameter ϵ has dimensions of length and plays the role of a short-distance cutoff. (A mass would multiply the integrand by e^{-m^2t}, which does not affect the divergent terms, which come from $t\lesssim\epsilon^2\ll 1/m^2$; we set $m=0$.)

The virtue of the heat kernel $\mathrm{Tr}\,\mathrm{e}^{t\nabla^2}$ is that it factorizes on a product geometry. On $C_{2\pi n}\times\mathbb{R}^{d-2}$,

$$\mathrm{Tr}\;\mathrm{e}^{t\nabla^2} \;=\; \Theta(t;2\pi n)\;\times\;\frac{A}{(4\pi t)^{(d-2)/2}}, \qquad \Theta(t;\delta)\;\equiv\;\mathrm{Tr}_{C_\delta}\;\mathrm{e}^{t\nabla_2^2}, \tag{20.10}$$

where the transverse factor is the flat-space heat kernel on $\mathbb{R}^{d-2}$ and its infrared-regularized volume A is the area of the entangling surface. Everything that follows is proportional to A: the geometry is translation invariant along the entangling surface, so the free energy, and hence the entropy, is extensive in A. This is the origin of the area law.

20.3.3 The heat kernel on a cone

It remains to compute $\Theta(t;\delta)$. We first do this exactly for the angles $\delta=2\pi/N$ with integer N, where the cone is the quotient $\mathbb{R}^2/\mathbb{Z}_N$ and the method of images applies, and then state the unique continuation in δ.

On the quotient $\mathbb{R}^2/\mathbb{Z}_N$, with $\mathbb{Z}_N$ acting by rotations R_k through $2\pi k/N$ about the origin, the heat kernel is the image sum

$$K_C(x,x';t) \;=\; \sum_{k=0}^{N-1} K_{\mathbb{R}^2}(x,R_kx';t), \qquad K_{\mathbb{R}^2}(x,x';t) = \frac{1}{4\pi t}\,\mathrm{e}^{-|x-x'|^2/4t}, \tag{20.11}$$

which manifestly satisfies the heat equation and the required periodicity. On the diagonal, $|x-R_kx| = 2\rho\sin(\pi k/N)$ for x at radius ρ, so

$$\Theta\Big(t;\tfrac{2\pi}{N}\Big) \;=\; \int_C \mathrm{d}^2x\left[\frac{1}{4\pi t}+\frac{1}{4\pi t}\sum_{k=1}^{N-1}\mathrm{e}^{-\rho^2\sin^2(\pi k/N)/t}\right]. \tag{20.12}$$

The $k=0$ term gives the universal bulk contribution $V/4\pi t$, with V the (infrared-regularized) area of the cone. For the image terms, the integral over the cone is $1/N$ of the integral over the plane, and the Gaussian radial integral is elementary:

$$\frac{1}{N}\int_0^\infty 2\pi\rho\,\mathrm{d}\rho\;\frac{\mathrm{e}^{-\rho^2\sin^2(\pi k/N)/t}}{4\pi t} \;=\; \frac{1}{4N\sin^2(\pi k/N)}, \tag{20.13}$$

independent of t: the cone is scale invariant, so its tip can only contribute a pure number. Using the classical trigonometric identity $\sum_{k=1}^{N-1}\sin^{-2}(\pi k/N) = (N^2-1)/3$, we obtain

$$\Theta\Big(t;\tfrac{2\pi}{N}\Big) \;=\; \frac{V}{4\pi t}\;+\;\frac{1}{12}\Big(N-\frac{1}{N}\Big). \tag{20.14}$$

The tip term is a function of the opening angle alone, and rewriting $N=2\pi/\delta$ puts it in a form that continues naturally to all δ:

$$\Theta(t;\delta) \;=\; \frac{V(\delta)}{4\pi t}\;+\;c(\delta)+O(\mathrm{e}^{-\#/t}), \qquad c(\delta)\;=\;\frac{1}{12}\Big(\frac{2\pi}{\delta}-\frac{\delta}{2\pi}\Big). \tag{20.15}$$

The continuation is fixed by the Sommerfeld contour representation of the heat kernel on C_δ, valid for all $\delta > 0$ [116], which reproduces (20.14) at $\delta = 2\pi/N$; the exponentially small terms depend on the infrared regularization and will not matter. Two checks: $c(2\pi) = 0$ (the plane has no tip), and $c(\delta) > 0$ for a deficit ($\delta < 2\pi$), as positivity of the image terms requires. For our replica cone, $\delta = 2\pi n$:

$$c(2\pi n) \;=\; \frac{1}{12}\Big(\frac{1}{n} - n\Big). \tag{20.16}$$

20.3.4 Assembling the entropy

Insert (20.10) and (20.15) into (20.9). The bulk area of the cone is linear in the opening angle, $V(2\pi n) = n\,V_1$ with V_1 the regularized area at $n = 1$, so

$$\log Z_n \;=\; \frac{1}{2}\int_{\epsilon^2}^{\infty}\frac{\mathrm{d}t}{t}\left[n\,\frac{V_1}{4\pi t} \;+\; \frac{1}{12}\Big(\frac{1}{n} - n\Big)\right]\frac{A}{(4\pi t)^{(d-2)/2}}. \tag{20.17}$$

Now apply the replica formula (20.6). The operator $(1 - n\partial_n)$ annihilates anything linear in n: the entire extensive part of the free energy drops out, and with it all dependence on the infrared regulator V_1: the entropy is a property of the entangling surface alone. The tip term $c(2\pi n)$ satisfies $c(2\pi) = 0$ and

$$(1 - n\,\partial_n)\,\frac{1}{12}\Big(\frac{1}{n} - n\Big)\Big|_{n=1} \;=\; -\,\frac{1}{12}\Big(-\,\frac{1}{n^2} - 1\Big)\Big|_{n=1} \;=\; \frac{1}{6}, \tag{20.18}$$

so the entropy is

$$S \;=\; \frac{1}{6}\cdot\frac{1}{2}\int_{\epsilon^2}^{\infty}\frac{\mathrm{d}t}{t}\,\frac{A}{(4\pi t)^{(d-2)/2}} \;=\; \frac{A}{6\,(d-2)\,(4\pi)^{(d-2)/2}}\,\frac{1}{\epsilon^{d-2}} \qquad (d > 2). \tag{20.19}$$

Entanglement entropy of a free scalar across a planar cut. In $d = 4$,

$$S \;=\; \frac{A}{48\pi\,\epsilon^2}, \tag{20.20}$$

with A the area of the entangling surface and ϵ the proper-time ultraviolet cutoff. The entropy is an area law; the coefficient is finite and computable once the regularization is fixed, but diverges as $\epsilon \to 0$: the vacuum entanglement across any surface in quantum field theory is ultraviolet dominated.

Remark 20.1 (The two-dimensional check)**.** In $d = 2$ the transverse factor is absent ($A \to 1$: the entangling surface is a single point) and the proper-time integral in (20.19) becomes logarithmic. Cutting it off at an infrared scale $t \sim \ell^2$,

$$S \;=\; \frac{1}{6}\cdot\frac{1}{2}\int_{\epsilon^2}^{\ell^2}\frac{\mathrm{d}t}{t} \;=\; \frac{1}{6}\,\log\frac{\ell}{\epsilon}, \tag{20.21}$$

which is the Calabrese–Cardy result $\frac{c}{6}\log(\ell/\epsilon)$ [104] for a region with a single entangling point and $c = 1$.[1]

[1] The agreement is exact rather than coincidental: the tip coefficient (20.16) is $c(2\pi n) = -2h_n$, with h_n the twist weight (21.10) at $c = 1$, so (20.21) is the Calabrese–Cardy computation for a free boson and one endpoint, done with the heat kernel instead of the conformal map. What the cone cannot give is general c, the second endpoint of an interval (which doubles the coefficient to $\frac{c}{3}$), and the finite-size and finite-temperature forms; these come from the twist-field computation of Chapter 21.

Remark 20.2 (Scheme dependence)**.** The coefficient $1/48\pi$ in (20.20) is tied to the proper-time regularization. The thermal-gas estimate (20.5), which cuts off the Rindler wedge at proper distance ϵ from the horizon, gave $1/360\pi$ per bosonic degree of freedom; Srednicki's lattice gave yet another number. There is no contradiction: the leading entanglement entropy is not universal, since different regulators define different cutoffs. What is universal is the area law with a coefficient of order one per field in cutoff units, and, as we now show, the combination in which this coefficient enters physical quantities.

20.4 The renormalization of Newton's constant and S_{gen}

- *Why the identification cannot be literal.* Set $\epsilon \sim \ell_P$ in (20.20) and compare with $A/4G$. The coefficient in (20.20) depends on the regularization (Remark 20.2), while $A/4G$ does not; and the comparison treats G as a fixed number, which it is not. The Newton constant in $A/4G$ is the coefficient of $\int \sqrt{g}\, R$ in the effective action, and integrating out the matter field shifts it: the same one-loop determinant that produced (20.20) contains an $\int \sqrt{g}\, R$ term whose coefficient diverges as $\epsilon \to 0$. The two divergences are one and the same.

20.4.1 One loop of matter renormalizes $1/G$

1. *Seeley–DeWitt.* Integrate the scalar out in a slowly varying background metric $g_{\mu\nu}$. The one-loop effective action is again (20.9), and the short-time expansion of the curved-space heat kernel is
$$\operatorname{Tr} e^{t\nabla^2} = \frac{1}{(4\pi t)^{d/2}} \int \sqrt{g} \left[1 + \frac{t}{6} R + O(t^2) \right] \tag{20.22}$$
for a minimally coupled scalar (with a coupling $\xi R\phi^2$ the coefficient of tR is $\frac{1}{6} - \xi$).

2. *The induced Einstein–Hilbert term.* Inserting the term linear in R into (20.9) in $d = 4$,
$$\log Z_{\text{1-loop}} \supset \frac{1}{2} \int_{\epsilon^2}^{\infty} \frac{\mathrm{d}t}{t} \frac{1}{(4\pi t)^2} \frac{t}{6} \int \sqrt{g}\, R = \frac{1}{192\pi^2 \epsilon^2} \int \sqrt{g}\, R. \tag{20.23}$$
and a term proportional to $\int \sqrt{g}\, R$ in $-\log Z$ is an Einstein–Hilbert action: with $I = -\frac{1}{16\pi G} \int \sqrt{g}\, R + \cdots$ from (7.10) and $\log Z = -I_{\text{eff}}$, the matter loop shifts the coefficient,
$$\frac{1}{16\pi G_{\text{eff}}} = \frac{1}{16\pi G_{\text{bare}}} + \frac{1}{192\pi^2 \epsilon^2}, \qquad \text{i.e.} \qquad \frac{1}{4G_{\text{eff}}} = \frac{1}{4G_{\text{bare}}} + \frac{1}{48\pi\epsilon^2}. \tag{20.24}$$

3. *Same coefficient, and why.* The coefficient $1/48\pi\epsilon^2$ is exactly that of the entanglement entropy (20.20) [117, 118], and not by accident: both are responses of the same one-loop $\log Z$ to a conical deformation. The entropy was read off from $\log Z$ on a cone through $(1 - n\partial_n)$; the gravitational entropy of Section 7.6 is read off from the response of the effective action to a conical deficit at the horizon, whose tip curvature (7.30) feeds the $\int \sqrt{g}\, R$ term of (20.23). One computation, two readings: matter entanglement across the surface, or a shift of $A/4G$.

20.4.2 The generalized entropy is the invariant object

Cutoff independence of the generalized entropy. Define the generalized entropy of a horizon cut as in Chapter 2,

$$S_{\text{gen}} = \frac{A}{4G(\epsilon)} + S_{\text{out}}(\epsilon), \tag{20.25}$$

with $G(\epsilon)$ the bare Newton constant at cutoff ϵ and $S_{\text{out}}(\epsilon)$ the matter entanglement entropy at the same cutoff. The dressed constant G_{eff}, the one a Cavendish experiment measures with all matter loops included, is ϵ-independent, so by (20.24)

$$\frac{1}{4G(\epsilon)} = \frac{1}{4G_{\text{eff}}} - \frac{1}{48\pi\epsilon^2}. \tag{20.26}$$

The two terms of (20.25) depend on ϵ in opposite ways. From (20.26), $A/4G(\epsilon) = A/4G_{\text{eff}} - A/48\pi\epsilon^2$; from (20.20), $S_{\text{out}}(\epsilon) = A/48\pi\epsilon^2 + \text{finite}$. Hence

$$\begin{aligned} \epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}\frac{A}{4G(\epsilon)} &= \epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Big(-\frac{A}{48\pi\epsilon^2}\Big) = +\frac{2A}{48\pi\epsilon^2}, \\ \epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}S_{\text{out}}(\epsilon) &= \epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Big(\frac{A}{48\pi\epsilon^2}\Big) = -\frac{2A}{48\pi\epsilon^2}, \end{aligned} \tag{20.27}$$

and the two cancel:

$$\epsilon\frac{\mathrm{d}}{\mathrm{d}\epsilon}S_{\text{gen}} = 0. \tag{20.28}$$

Lowering the cutoff moves the entanglement of the newly integrated-out modes out of S_{out} and into $A/4G$ with exactly compensating coefficient. Neither term is separately physical; the sum is.

- *Refinements.* The cancellation (20.28) was verified only for the leading divergence of a minimally coupled scalar. Two refinements are known to work but are more intricate [119]: the subleading logarithmic divergences in $d = 4$ match the renormalization of curvature-squared couplings, whose entropy contributions go beyond the area (Wald-type terms); and for non-minimally coupled scalars and gauge fields both S_{out} and the $\int\sqrt{g}\,R$ coefficient contain contact terms on the entangling surface, and the matching requires edge modes.

- *What is and is not explained.* The robust statement is that the entanglement interpretation holds in renormalized form: *the matter entanglement across the horizon accounts for the one-loop part of $A/4G$.* It does not explain the tree-level part $A/4G_{\text{bare}}$: it shows how matter loops shift that coefficient, not which states it counts. Supplying the states is the task of a theory of quantum gravity, as the dual CFT does for AdS black holes in Chapter 18 and the D-brane counting of Strominger and Vafa [19] for supersymmetric black holes. The quantity (20.25) is the one Bekenstein introduced on thermodynamic grounds; that it is also the ultraviolet-finite version of a fine-grained entanglement entropy is the conceptual heart of this chapter.

20.5 The Bekenstein bound ⋆

20.5.1 The bound and its ambiguities

Once black holes carry entropy, the generalized second law of Chapter 2 constrains ordinary matter. Lower a box of energy E and size R slowly on a string toward the horizon of a large black hole and release it (the Geroch process). By the near-horizon metric (1.11) the time-translation Killing vector has norm $\kappa\rho$ at proper distance ρ, so the energy of the box measured at infinity is E times the redshift $\kappa\rho$, the difference being extracted as work at the top of the string. Released with its centre at $\rho \sim R$, the box adds $\delta M \simeq \kappa R E$ to the hole, whose entropy grows by $\delta S_{\mathrm{BH}} = \delta M/T_H = 2\pi E R$, while the exterior loses S. The generalized second law therefore demands the Bekenstein bound [120],

$$S \;\leq\; 2\pi E R. \tag{20.29}$$

Newton's constant has cancelled, since $\kappa R E$ and $T_H = \kappa/2\pi$ each carry one power of κ, so this is a claim about ordinary quantum matter. As stated, however, each symbol hides an ambiguity:

1. which S: the entanglement entropy of the region the system occupies is ϵ-divergent, (20.20), and violates any finite bound, so some subtraction is implied but unspecified;
2. which E: the energy in a sharply defined region fluctuates, and vacuum energy contributes;
3. which R: a quantum state has no sharp edge.

The derivation is also suspect: a box lowered through the black hole's thermal atmosphere feels a buoyant force, which changes the energy bookkeeping and, in the analysis of Unruh and Wald [121], rescues the generalized second law without any bound on S/E. Casini's observation [103] is that relative entropy and its positivity (Theorem 19.1) supply exactly the subtractions needed, and that the bound then becomes a theorem.

20.5.2 Casini's proof

Let $\mathcal{R}$ be the right Rindler wedge $x > 0$ of Minkowski space, the idealization of the region next to the horizon. The proof takes four steps.

1. *Two states.* Compare the vacuum Ω with an arbitrary global state ψ, reduced to the wedge:
$$\sigma \;=\; \mathrm{Tr}_{\bar{\mathcal{R}}}\, |\Omega\rangle\langle\Omega| \;=\; \frac{\mathrm{e}^{-2\pi K}}{Z}, \qquad \rho \;=\; \mathrm{Tr}_{\bar{\mathcal{R}}}\, |\psi\rangle\langle\psi|, \tag{20.30}$$
where the first equality is (6.9).

2. *Vacuum subtraction.* Define
$$\Delta S \;\equiv\; S(\rho) - S(\sigma), \qquad \Delta K \;\equiv\; 2\pi\Big(\mathrm{Tr}(\rho\, K) - \mathrm{Tr}(\sigma\, K)\Big). \tag{20.31}$$
Both are ultraviolet finite: the divergences of S and of $\langle K\rangle$ come from modes at the entangling surface, and a state that differs from the vacuum only in the interior of the wedge has the same short-distance structure there, so the divergences cancel in the differences. ΔS is the entropy the excitation adds to the wedge; ΔK, built from the boost generator (6.8), is its energy weighted by the distance from the horizon.

3. *Relative entropy.* With $\log\sigma = -2\pi K - \log Z$,
$$S(\rho\,\|\,\sigma) \;=\; \mathrm{Tr}\,\rho\log\rho \;-\; \mathrm{Tr}\,\rho\log\sigma \;=\; -S(\rho) \;+\; 2\pi\,\mathrm{Tr}(\rho K) \;+\; \log Z. \tag{20.32}$$
At $\rho = \sigma$ the left side vanishes, giving $0 = -S(\sigma) + 2\pi\,\mathrm{Tr}(\sigma K) + \log Z$; subtracting removes $\log Z$,
$$S(\rho\,\|\,\sigma) \;=\; \Delta K \;-\; \Delta S. \tag{20.33}$$

4. *Positivity.* Relative entropy is non-negative, Theorem 19.1.

Bekenstein bound, Casini's form [103]. For any state reduced to the Rindler wedge,

$$\Delta S \;\le\; \Delta K, \tag{20.34}$$

with ΔS the vacuum-subtracted entanglement entropy of the wedge and $\Delta K = 2\pi\,\Delta\int_{x>0} \mathrm{d}x\,\mathrm{d}^{d-2}y\; x\,\langle T_{00}\rangle$ the vacuum-subtracted modular energy. Both sides are finite, gravity appears nowhere, and equality holds if and only if the state is indistinguishable from the vacuum inside the wedge.

20.5.3 Recovering $2\pi ER$

Let ψ describe an object of energy E above the vacuum, localized on the $t = 0$ slice at distances $x \in (0, \ell)$ from the edge of the wedge with ℓ of order R, the configuration of the box at the moment of release. Then $x \le \ell$ on the support of the excitation, and for a positive energy density

$$\Delta K \;=\; 2\pi \int x\,\Delta\langle T_{00}\rangle\,\mathrm{d}x\,\mathrm{d}^{d-2}y \;\le\; 2\pi\,\ell \int \Delta\langle T_{00}\rangle\,\mathrm{d}x\,\mathrm{d}^{d-2}y \;=\; 2\pi\,\ell\,E \;\sim\; 2\pi R\,E, \tag{20.35}$$

so (20.34) gives $\Delta S \le 2\pi RE$. Each ambiguity of Section 20.5.1 has been resolved:

1. S is the vacuum-subtracted entanglement entropy, finite;
2. E is the boost energy, weighted by position so that neither the vacuum nor distant matter contributes;
3. R is the distance to the entangling surface, a property of the wedge rather than of the system, which is what the lowering experiment was probing;
4. buoyancy never arises, since no black hole and no dynamical process appears: the bound is a kinematic statement about states in flat space.

Remark 20.3 (First law of entanglement)**.** For an infinitesimal perturbation of the vacuum the relative entropy (20.33) vanishes to second order, since it is non-negative and vanishes at $\rho = \sigma$, so $\delta S = \delta\langle 2\pi K\rangle$: the first law of entanglement [122], of which (20.34) is the integrated form. Relative entropy to the vacuum and its monotonicity under restriction (Theorem 19.6) also underlie Wall's proof of the generalized second law for causal horizons [123], and the consistency of holographic entanglement entropy in Chapter 21.

20.6 Rényi entropies and the generalized entropy of a black hole

20.6.1 Gravitational replicas: smooth cones

For the thermal state $\rho = \mathrm{e}^{-\beta H}/Z(\beta)$ of quantum gravity with AdS boundary conditions (Chapter 18),

$$\operatorname{Tr}\rho^n \;=\; \frac{Z(n\beta)}{Z(\beta)^n}, \qquad S_n \;=\; \frac{1}{1-n}\Big[\log Z(n\beta) - n\log Z(\beta)\Big], \tag{20.36}$$

and the gravitational path integral computes each factor as $\log Z(\beta) \simeq -I(\beta)$, the on-shell action of the cigar with boundary circle of length β. The geometry computing $Z(n\beta)$ is not a singular cone but the smooth cigar whose horizon radius has adjusted to the period $n\beta$: in gravity the geometry is dynamical, and the equations of motion remove the would-be conical singularity by selecting a different black hole. Hence

$$S \;=\; \lim_{n\to 1} S_n \;=\; (\beta\,\partial_\beta - 1)\, I(\beta), \tag{20.37}$$

the thermodynamic entropy (7.4), which gives $A/4G$. As a check, for the planar black brane of Chapter 18, $\log Z = c\,V\beta^{-(d-1)}$ with $c > 0$, and (20.36) gives

$$S_n \;=\; \frac{cV}{\beta^{d-1}}\,\frac{n^{-(d-1)} - n}{1-n} \;\xrightarrow{n\to 1}\; \frac{d\,cV}{\beta^{d-1}}, \tag{20.38}$$

positive and decreasing in n, as Proposition 19.5 requires. (For flat-space Schwarzschild, $I = \beta^2/16\pi G$ gives $S_n = n\beta^2/16\pi G$, increasing in n: one more symptom that the flat-space canonical ensemble does not exist, Chapter 5.)

Remark 20.4 (Fixed-geometry replicas and refined entropies)**.** Computing $\operatorname{Tr}\rho^n$ instead on the original geometry with a conical excess at the horizon, as in field theory, is off shell in gravity, and the action of that configuration computes not S_n but the refined entropies $\tilde{S}_n = n^2\partial_n\big(\frac{n-1}{n}S_n\big)$, which equal $A(n)/4G$ with $A(n)$ the horizon area of the conical saddle [124]. This is the natural language for the derivation of the Ryu–Takayanagi formula in Chapter 21, where the boundary conditions forbid the geometry from smoothing itself out.

20.6.2 One loop: the path integral computes S_{gen}

At one loop the partition function is the classical action plus the fluctuation determinant of the matter (and graviton) fields on the cigar X_β:

$$\log Z(\beta) \;=\; -\,I_{\text{grav}}(\beta) \;+\; \log Z_{\text{matter}}\big[X_\beta\big] \;+\; \cdots\,. \tag{20.39}$$

Apply the replica derivative $(\beta\partial_\beta - 1)$ of (20.37) to both terms. The first gives $A/4G$. For the second, changing β acts in two places:

1. away from the tip, it changes the geometry smoothly and yields the thermal entropy of the matter gas in the black hole atmosphere;
2. at the tip, where the cigar is the flat cone $C_{2\pi n} \times$ (horizon), it is the conical deformation of Section 20.3 and yields the entanglement entropy of the matter across the horizon, (20.20).

Together these are S_{out}, the von Neumann entropy (20.1) of the fields outside the horizon, which in the Hartle–Hawking state is one continuous object from the entangling surface out to the atmosphere (Section 20.2). Hence

$$S \;=\; (\beta\partial_\beta - 1)\Big[I_{\text{grav}} - \log Z_{\text{matter}}\Big] \;=\; \frac{A}{4G} \;+\; S_{\text{out}} \;=\; S_{\text{gen}}, \tag{20.40}$$

with the cutoff dependence of the two terms cancelling as in Section 20.4.2. The gravitational path integral at one loop computes the generalized entropy without S_{gen} being put in by hand, and the same holds order by order, with Wald-type higher-derivative corrections joining the area term.

20.7 Synthesis: the meaningful fine-grained entropy in gravity

- *Two meanings of $A/4G$.* Chapter 19 asked which of the two notions of entropy $A/4G$ is. The verdict of Parts I and II, confirmed in Chapter 18 where $A/4G$ counted the dual CFT states at a given energy, stands: it is a thermodynamic, coarse-grained entropy. This chapter adds a second meaning: the entanglement entropy S_{out} of the fields outside the horizon is a fine-grained entropy, ultraviolet divergent, and its divergence is absorbed by the renormalization of G, so that only $S_{\text{gen}} = A/4G + S_{\text{out}}$ is finite and regularization independent, (20.28). Neither term means anything by itself, one being a bare coupling and the other a cutoff-dependent divergence.

The lesson of this chapter. In a gravitational theory, the fine-grained entropy associated with a surface Σ bounding a region is the generalized entropy

$$S_{\text{gen}}(\Sigma) \;=\; \frac{A(\Sigma)}{4G} \;+\; S_{\text{out}}, \tag{20.41}$$

with G the renormalized Newton constant and S_{out} the renormalized von Neumann entropy of the fields outside Σ. The two terms are not independent: lowering the cutoff moves entropy from S_{out} into $A/4G$, (20.27), and only the sum is cutoff independent, (20.28). The generalized second law of Chapter 2, the island proposal of Section 19.10.3, and the Ryu–Takayanagi formula of Chapter 21 are all statements about S_{gen}.

Three closing comments.

1. The generalized second law is a second law for a quantity that is part geometry, part entanglement: when matter crosses the horizon, S_{out} drops and $A/4G$ rises by more, and the cutoff independence of (20.41) is what makes this bookkeeping physical.

2. The Bekenstein bound became precise only for vacuum-subtracted quantities, the same kind of subtraction that renders S_{gen} finite: only relative, renormalized information quantities are physical.

3. So far the entangling surface has been a horizon, imposed by causality. The final chapter liberates it: for an arbitrary boundary region in AdS/CFT, the bulk surface that computes its fine-grained entropy is found by minimizing $A/4G$, and then S_{gen}, over admissible surfaces. That is the Ryu–Takayanagi formula.

Chapter 21

The Ryu–Takayanagi Formula and Closing Reflections

This final lecture assembles everything into a single formula: the Ryu–Takayanagi prescription [125], which computes the entanglement entropy of a boundary region as the area of a minimal bulk surface over $4G$. We motivate it from S_{gen}, test it in $\text{AdS}_3/\text{CFT}_2$, where a geodesic length reproduces the Calabrese–Cardy entropy and BTZ recovers the thermofield-double entropy, work out strips and spheres in higher dimensions, verify strong subadditivity by surgery on surfaces, and present the Lewkowycz–Maldacena derivation [45], in which the conical geometries of Chapters 7 and 20 make their final appearance. A short closing section records what the course has established and what it has not.

21.1 Heuristic motivation: liberating the surface

- *So far* the entangling surface was dictated by causality, a black hole horizon or the edge of a Rindler wedge, to which Chapter 20 attached the generalized entropy $S_{\text{gen}}(\Sigma) = A(\Sigma)/4G + S_{\text{out}}(\Sigma)$, (20.41).
- *Now* let A be a spatial region on the boundary of AdS, in a state such as the vacuum. The CFT is an ordinary quantum system, so S_A is defined as in Chapter 19, and the dictionary must compute it in the bulk. Any bulk codimension-two surface γ that together with A bounds a bulk region, we say γ is *homologous* to A, separates the part of the bulk accessible from A from the rest, and Chapter 20 assigns to that division the entropy $S_{\text{gen}}(\gamma) \simeq A(\gamma)/4G$ at leading order in G. Each such γ is a consistent accounting surface for the information in A; the most economical description uses the cheapest one (Figure 21.1).

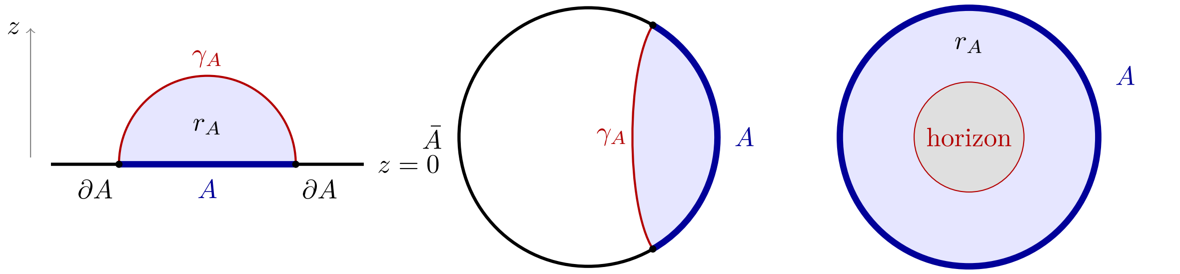
The Ryu–Takayanagi formula [125]. In a static state of a holographic CFT, the entanglement entropy of a boundary spatial region A is

$$S_A \;=\; \min_{\gamma_A} \frac{\text{Area}(\gamma_A)}{4G}, \tag{21.1}$$

where the minimum is over bulk codimension-two surfaces γ_A, lying in the bulk time-reflection-symmetric slice, that are anchored on the entangling surface, $\partial\gamma_A = \partial A$, and are *homologous* to A: $\gamma_A \cup A$ bounds a bulk region. G is the bulk Newton constant.

Three consistency checks:

1. if A is the entire boundary and the state is pure (global AdS, Figure 21.1(b)), the empty surface is homologous to A and $S = 0$, as purity requires; for a proper subregion, A and its complement $\bar{A}$ share the same surface, so $S_A = S_{\bar{A}}$;



(a) Poincaré slice: γ_A anchored on ∂A, homologous to A via r_A (b) global AdS, pure state: $\gamma_A = \gamma_{\bar{A}}$, so $S_A = S_{\bar{A}}$ (c) black hole, A = whole boundary: γ_A = horizon, $S_A = A_{\mathrm{hor}}/4G$

Figure 21.1: The Ryu–Takayanagi prescription on a bulk time slice. (a) In Poincaré coordinates, A is an interval on the boundary $z = 0$; the surface γ_A (red) is anchored on ∂A and, together with A, bounds the shaded bulk region r_A, which is the homology condition. (b) In global AdS the same surface is anchored on the boundary of the complement $\bar{A}$ and bounds a region with it, so the two entropies agree, as they must for a pure state. (c) With a black hole in the bulk and A the entire boundary, the empty surface is not homologous to A, because the region it would bound contains the horizon; the minimal admissible surface is the horizon itself, and S_A is the Bekenstein–Hawking entropy.

2. if the bulk is a black hole and A is the entire boundary (Figure 21.1(c)), the empty surface is not homologous to A, since the horizon is in the way, and the minimal admissible surface is the horizon: $S = A_{\mathrm{hor}}/4G$, the thermofield-double entropy of Section 18.7. The homology condition is what lets the formula know that a thermal state is mixed;

3. for a small region the surface hugs the boundary and its area diverges as $\mathrm{Area}(\partial A)/\epsilon^{d-2}$, the area law (19.37) of field-theory entanglement, with the bulk cutoff mapped to the CFT cutoff by the UV/IR relation of Section 12.5.

The formula is the leading term in G; quantum corrections and time-dependent states are discussed in Section 21.5.1. These are consistency checks, not tests. The formula earns its keep by reproducing entropies that can be computed independently in the CFT, and we turn to the cleanest case.

21.2 The original setting: an interval in $\mathrm{AdS}_3/\mathrm{CFT}_2$

21.2.1 The geodesic computation

Work in Poincaré AdS_3,

$$\mathrm{d}s^2 \;=\; \frac{L^2}{z^2}\Big(-\,\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}z^2\Big), \tag{21.2}$$

dual to a CFT_2 on the (t, x) plane in its vacuum, and take $A = \{x \in (-\ell/2, \ell/2)\}$ at $t = 0$. A codimension-two surface in three bulk dimensions is a curve, and on the $t = 0$ slice, the hyperbolic plane of radius L, the RT surface is the geodesic joining the endpoints $x = \pm\ell/2$ on the boundary: the semicircle (10.19) of Section 10.5, of regulated length (10.21). Hence

$$S_A \;=\; \frac{\mathrm{Length}}{4G} \;=\; \frac{L}{2G}\,\log\frac{\ell}{\epsilon}. \tag{21.3}$$

The divergence is logarithmic because ∂A is two points (Remark 20.1), and the bulk cutoff ϵ is the CFT cutoff by the UV/IR relation of Section 12.5.

21.2.2 Matching the CFT: Brown–Henneaux and Calabrese–Cardy

The central charge of the CFT dual to AdS_3 gravity follows from the holographic Weyl anomaly of Chapter 14 run in three bulk dimensions, or from the asymptotic Virasoro symmetry of AdS_3, which is how Brown and Henneaux found it before AdS/CFT [69]:

$$c \;=\; \frac{3L}{2G}. \tag{21.4}$$

Substituting into (21.3),

$$S_A \;=\; \frac{c}{3}\,\log\frac{\ell}{\epsilon}, \tag{21.5}$$

the exact vacuum entanglement entropy of an interval in any two-dimensional CFT [104], here obtained as the length of a bulk geodesic.

Since (21.5) is the sharpest test the RT formula faces in this course, we derive the CFT side [104].

1. *Replicas.* By (19.36) we need $\operatorname{Tr}\rho_A^n$, which by Section 19.8 is the Euclidean partition function on the n-sheeted surface $\mathcal{R}_n$, n copies of the plane cut along $A=[0,\ell]$ and glued cyclically across the cut,
$$\operatorname{Tr}\rho_A^n \;=\; \frac{Z[\mathcal{R}_n]}{Z[\mathbb{R}^2]^{\,n}}. \tag{21.6}$$

2. *Twist fields.* The same path integral is n decoupled copies of the CFT on a single plane, with the cyclic gluing implemented by *twist fields* $\sigma_n(0)$ and $\bar\sigma_n(\ell)$ at the branch points, which permute the copies and undo the permutation:
$$\operatorname{Tr}\rho_A^n \;=\; \langle\sigma_n(0)\,\bar\sigma_n(\ell)\rangle_{\mathbb{R}^2}. \tag{21.7}$$
The σ_n are conformal primaries of the n-copy theory, so Chapter 9 fixes this two-point function to be $\propto \ell^{-2\Delta_n}$; it remains to find Δ_n.

3. *The uniformizing map*
$$z \;=\; \Big(\frac{w}{w-\ell}\Big)^{1/n} \tag{21.8}$$
takes $\mathcal{R}_n$ one-to-one onto the plane, since circling $w=0$ n times circles $z=0$ once. On the plane $\langle T(z)\rangle=0$, so on $\mathcal{R}_n$ the one-point function of the stress tensor comes entirely from the Schwarzian term of its transformation law:
$$T(w)=\Big(\frac{\mathrm{d}z}{\mathrm{d}w}\Big)^2 T(z)+\frac{c}{12}\{z;w\} \quad\Longrightarrow\quad \langle T(w)\rangle_{\mathcal{R}_n}=\frac{c}{24}\Big(1-\frac{1}{n^2}\Big)\frac{\ell^2}{w^2(w-\ell)^2}. \tag{21.9}$$

4. *The twist dimension.* In the n-copy picture the total stress tensor is $\langle T_{\rm tot}(w)\rangle = n\,\langle T(w)\rangle_{\mathcal{R}_n}$, while the conformal Ward identity of Chapter 9 gives, for primaries of holomorphic weight h_n at 0 and ℓ, $\langle T_{\rm tot}(w)\rangle = h_n\,\ell^2/w^2(w-\ell)^2$. Comparing,
$$h_n=\bar h_n=\frac{c}{24}\Big(n-\frac{1}{n}\Big), \qquad \Delta_n=h_n+\bar h_n=\frac{c}{12}\Big(n-\frac{1}{n}\Big). \tag{21.10}$$

5. *The entropy.* Therefore $\operatorname{Tr}\rho_A^n = c_n\,(\ell/\epsilon)^{-\frac{c}{6}(n-1/n)}$, with c_n a non-universal constant and $c_1=1$, and the replica formula gives
$$S_A \;=\; -\,\partial_n\Big[-\frac{c}{6}\Big(n-\frac{1}{n}\Big)\log\frac{\ell}{\epsilon}+\log c_n\Big]_{n=1} \;=\; \frac{c}{3}\,\log\frac{\ell}{\epsilon}\;+\;\text{const}, \tag{21.11}$$
in agreement with the geodesic result (21.5).

21.2.3 Finite temperature: BTZ and the thermofield double

At temperature $1/\beta$ the dual geometry is the planar BTZ black hole, the $d = 2$ case of the black brane (18.26),

$$\mathrm{d}s^2 \;=\; \frac{L^2}{z^2}\Big(-f(z)\,\mathrm{d}t^2 + \mathrm{d}x^2 + \frac{\mathrm{d}z^2}{f(z)}\Big), \qquad f(z) \;=\; 1 - \frac{z^2}{z_h^2}, \qquad \beta \;=\; 2\pi z_h. \tag{21.12}$$

Take A an interval of length ℓ at $t = 0$. On the $t = 0$ slice of (21.12) the four steps of Section 10.5 go through with the single replacement $\mathrm{d}z^2 \to \mathrm{d}z^2/f$.

1. The length functional and its first integral are

$$\text{Length} \;=\; L\int \mathrm{d}x\,\frac{\sqrt{1+z'^2/f}}{z}, \qquad z\sqrt{1+z'^2/f} \;=\; z_*, \quad \text{i.e.} \quad z'^2 \;=\; \frac{f\,(z_*^2 - z^2)}{z^2}. \tag{21.13}$$

2. The turning point is fixed by the interval length,

$$\ell \;=\; 2\int_0^{z_*}\frac{z\,\mathrm{d}z}{\sqrt{f\,(z_*^2 - z^2)}} \;=\; z_h\,\operatorname{arccosh}\frac{z_h^2 + z_*^2}{z_h^2 - z_*^2} \qquad\Longrightarrow\qquad z_* \;=\; z_h \tanh\frac{\pi\ell}{\beta}, \tag{21.14}$$

so the geodesic never crosses the horizon and hugs it as $\ell \to \infty$.

3. The regulated length is

$$\text{Length} \;=\; 2L\int_\epsilon^{z_*}\frac{z_*\,\mathrm{d}z}{z\sqrt{f\,(z_*^2 - z^2)}} \;=\; 2L\log\frac{2z_* z_h}{\epsilon\sqrt{z_h^2 - z_*^2}} + O(\epsilon^2) \;=\; 2L\log\Big[\frac{\beta}{\pi\epsilon}\sinh\frac{\pi\ell}{\beta}\Big]. \tag{21.15}$$

Both integrals are elementary in the variable $u = z^2$. With (21.4),

$$S_A \;=\; \frac{c}{3}\,\log\Big[\frac{\beta}{\pi\epsilon}\,\sinh\frac{\pi\ell}{\beta}\Big], \tag{21.16}$$

the **Calabrese–Cardy result for a thermal state**, obtained in the CFT by the conformal map from the plane to the thermal cylinder [104].

- *Why they agree.* BTZ is a quotient of AdS_3, and the coordinate change from (21.12) to (21.2) acts on the boundary as $w = \mathrm{e}^{2\pi(x\pm t)/\beta}$, the exponential map of Chapter 6 that takes the plane to the thermal cylinder.

- *Limits.* For $\ell \ll \beta$ the sinh linearizes and (21.5) returns: short-distance entanglement does not feel the temperature. For $\ell \gg \beta$,

$$S_A \;\simeq\; \frac{\pi c}{3\beta}\,\ell \;+\; \frac{c}{3}\log\frac{\beta}{2\pi\epsilon} \;=\; s_{\text{th}}\,\ell \;+\; \text{const}, \qquad s_{\text{th}} = \frac{\pi c}{3\beta}, \tag{21.17}$$

  extensive, with the thermal entropy density of a 2d CFT: the entanglement entropy of a large region in a thermal state is dominated by thermal mixedness. Geometrically, by (21.14) the geodesic runs along the horizon for a length $\simeq \ell$, so the RT surface converts thermal entropy into horizon area.

- *One whole side.* The eternal BTZ black hole describes two copies of the CFT in the thermofield double state (18.32), joined through the Einstein–Rosen bridge, and is the laboratory for the homology rule. Take A to be the entire constant-t slice of the left boundary. Its entangling surface is empty, so admissible surfaces are closed; the empty surface is excluded, because a bulk region bounded by A alone would have to cross the second asymptotic region; the minimal admissible surface is the bifurcation horizon, and

$$S_{\text{left CFT}} \;=\; \frac{A_{\text{hor}}}{4G} \;=\; s_{\text{th}}\, V, \tag{21.18}$$

with V the regularized boundary volume: the entropy of one side of the thermofield double is the horizon area, as anticipated in Section 18.7. This is the $\ell \to \infty$ limit of (21.17) without its constant term: by (21.14) the geodesic of an interval approaches the horizon, and the constant $\frac{c}{3}\log\frac{\beta}{2\pi\epsilon}$ is the cost of the two endpoints, which the whole boundary does not have.

- *Two intervals, one on each side.* Take $A = A_L \cup A_R$, two intervals of equal length ℓ, one on each boundary, at $t = 0$ (Figure 21.2). Two classes of surfaces compete. *Disconnected*: each interval takes its own single-sided geodesic (21.16), with total $S_{\text{disc}} = \frac{2c}{3}\log[\frac{\beta}{\pi\epsilon}\sinh\frac{\pi\ell}{\beta}]$, growing linearly at large ℓ. *Connected*: two geodesics crossing the bridge, each joining an endpoint of A_L to the matching endpoint of A_R through the bifurcation surface. Each is the curve $x = \text{const}$, a geodesic by the reflection symmetry $x \to 2x_0 - x$ of the slice, of regulated length

$$2L\int_{\epsilon}^{z_h} \frac{\mathrm{d}z}{z\sqrt{f}} \;=\; 2L\,\operatorname{arccosh}\frac{z_h}{\epsilon} \;\simeq\; 2L\log\frac{\beta}{\pi\epsilon}, \tag{21.19}$$

the two halves of the integral being the two exteriors glued at $z = z_h$. The total, $S_{\text{conn}} = \frac{2c}{3}\log(\beta/\pi\epsilon)$, is independent of ℓ.

- *The transition.* For small ℓ, $S_{\text{disc}} < S_{\text{conn}}$, so the minimization in (21.1) selects the disconnected pair: the RT surface of the union is the union of the RT surfaces of A_L and A_R, $S_{A_L\cup A_R} = S_{A_L} + S_{A_R}$, and the mutual information $I(A_L{:}A_R) = S_{A_L} + S_{A_R} - S_{A_L\cup A_R}$ vanishes at order $1/G$. Beyond $\ell_c = \frac{\beta}{\pi}\log(1+\sqrt{2})$, where the two lengths are equal, the connected pair has the smaller length and

$$I(A_L{:}A_R) \;=\; \frac{2c}{3}\log\Big[\sinh\frac{\pi\ell}{\beta}\Big] \;\simeq\; \frac{2\pi c}{3\beta}\,\ell \qquad (\ell \gg \beta): \tag{21.20}$$

the left–right correlations of the thermofield double, detected in Section 18.7 through two-sided correlators, are extensive and are carried by surfaces threading the bridge. The RT surface jumps at ℓ_c while the entropy, the minimum of two smooth functions, is continuous.

21.3 Higher dimensions: strips and spheres

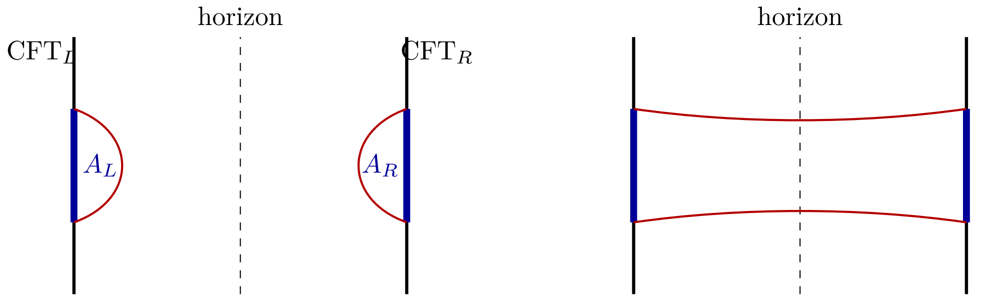
In higher dimensions RT surfaces are minimal-area surfaces. Two regions have surfaces in closed form, the strip by translation symmetry and the ball by conformal symmetry (Remark 21.1), and each exhibits one feature that $d = 2$ cannot show.

1. The strip checks that RT reproduces the area law of Chapter 20, $\text{Area}(\partial A)/\epsilon^{d-2}$ with a non-universal coefficient, and isolates a cutoff-independent finite term. It has no logarithm, because in $d = 4$ the coefficient of the logarithm is an integral of curvature invariants of ∂A [126], which vanish for a flat entangling surface.

2. The ball in $d = 4$ produces a logarithm whose coefficient is fixed by the trace anomaly of the CFT alone, and RT reproduces the holographic anomaly (14.15): this is the $d = 4$ analogue of the Calabrese–Cardy match (21.11).

Work in Poincaré AdS_{d+1} at fixed boundary time,

$$\mathrm{d}s^2 \;=\; \frac{L^2}{z^2}\Big(\mathrm{d}z^2 + \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu\Big). \tag{21.21}$$



(a) disconnected: $S \propto \ell$, $I = 0$ (b) connected: S const, $I \propto \ell$

Figure 21.2: The $t = 0$ spatial slice of the eternal (two-sided) BTZ black hole: two asymptotic boundaries, one for each CFT of the thermofield double, glued across the bifurcation horizon (dashed). For the union of two symmetric intervals $A_L \cup A_R$, the RT prescription compares (a) two single-sided geodesics with (b) two geodesics threading the Einstein–Rosen bridge, and takes the smaller total length. The connected phase dominates for $\ell \gtrsim \beta$, signalling extensive left–right entanglement.

21.3.1 The strip ⋆

Let A be the strip $x^1 \in (-\ell/2, \ell/2)$ times a regularized volume V_{d-2} in the other spatial directions. By translation symmetry the RT surface is a profile $z(x^1)$ times $\mathbb{R}^{d-2}$, with area

$$\text{Area} \;=\; L^{d-1}\, V_{d-2} \int \mathrm{d}x\, \frac{\sqrt{1+z'^2}}{z^{d-1}}, \tag{21.22}$$

the generalization of (10.17). The conserved Hamiltonian gives the first integral

$$\frac{1}{z^{d-1}\sqrt{1+z'^2}} \;=\; \frac{1}{z_*^{d-1}} \qquad \Longrightarrow \qquad z' \;=\; \pm\frac{\sqrt{z_*^{2(d-1)} - z^{2(d-1)}}}{z^{\,d-1}}, \tag{21.23}$$

with the turning point z_* fixed by the width,

$$\frac{\ell}{2} \;=\; \int_0^{z_*} \mathrm{d}z\, \frac{z^{\,d-1}}{\sqrt{z_*^{2(d-1)} - z^{2(d-1)}}} \;=\; z_*\, \frac{\sqrt{\pi}\,\Gamma\big(\frac{d}{2(d-1)}\big)}{\Gamma\big(\frac{1}{2(d-1)}\big)}, \tag{21.24}$$

where $u = (z/z_*)^{2(d-1)}$ reduces the integral to a Beta function. The surface reaches the finite depth $z_* \propto \ell$: wider regions probe deeper, the UV/IR relation of Section 12.5 made quantitative. Converting (21.22) to a z-integral with (21.23) and expanding near $z = 0$,

$$S_{\text{strip}} \;=\; \frac{L^{d-1} V_{d-2}}{4G}\left[\frac{2}{(d-2)\,\epsilon^{\,d-2}} - \frac{\kappa_d}{\ell^{\,d-2}}\right], \qquad \kappa_d \;=\; \frac{(2\sqrt{\pi})^{d-1}}{d-2}\left(\frac{\Gamma\big(\frac{d}{2(d-1)}\big)}{\Gamma\big(\frac{1}{2(d-1)}\big)}\right)^{d-1}, \tag{21.25}$$

again by Beta-function integrals [125]. The first term is the non-universal area-law divergence, proportional to $\text{Area}(\partial A) = 2V_{d-2}$. The second is finite and cutoff independent, since ℓ is the only scale and no local counterterm on the flat ∂A can produce a power of ℓ; its coefficient κ_d is the higher-dimensional analogue of $c/3$ in (21.5), a measure of the number of degrees of freedom, and comparing it with free-field values was the first test of RT in $d > 2$ [125].

21.3.2 The sphere in AdS_5

For a ball of radius R in $d = 4$ the finite term is universal and carries anomaly information. In boundary spherical coordinates,

$$\mathrm{d}s^2\big|_{t=\text{const}} \;=\; \frac{L^2}{z^2}\Big(\mathrm{d}z^2 + \mathrm{d}r^2 + r^2\mathrm{d}\Omega_2^2\Big), \tag{21.26}$$

the RT surface is a profile $z(r)$ with area

$$\text{Area} \;=\; 4\pi L^3 \int \mathrm{d}r\, \frac{r^2\sqrt{1+z'^2}}{z^3}, \tag{21.27}$$

and Euler–Lagrange equation

$$\frac{\mathrm{d}}{\mathrm{d}r}\left[\frac{r^2\, z'}{z^3\sqrt{1+z'^2}}\right] \;=\; -\,\frac{3\,r^2\sqrt{1+z'^2}}{z^4}. \tag{21.28}$$

The hemisphere

$$z(r) \;=\; \sqrt{R^2-r^2} \tag{21.29}$$

is an exact solution: with $z' = -r/z$ one has $\sqrt{1+z'^2} = R/z$, the bracket becomes $-r^3/Rz^3$, and both sides of (21.28) equal $-3r^2R/(R^2-r^2)^{5/2}$. (It is the image of the AdS$_3$ semicircle under the conformal symmetry, see Remark 21.1.) With $r = R\sin u$, $z = R\cos u$, and the cutoff $z=\epsilon$ at $\cos u_{\max} = \epsilon/R$,

$$\text{Area} \;=\; 4\pi L^3 \int_0^{u_{\max}} \mathrm{d}u\,\left(\sec^3 u - \sec u\right) \;=\; 4\pi L^3 \Big[\tfrac{1}{2}\sec u \tan u - \tfrac{1}{2}\log\left(\sec u + \tan u\right)\Big]_0^{u_{\max}}, \tag{21.30}$$

and expanding at $u_{\max}$,

$$\text{Area} \;=\; 2\pi L^3\left[\frac{R^2}{\epsilon^2} - \log\frac{2R}{\epsilon} - \frac{1}{2} + O(\epsilon^2)\right]. \tag{21.31}$$

Dividing by $4G_5$ and using $L^3/G_5 = 2N^2/\pi$, (18.22),

$$S_{\text{ball}} \;=\; \frac{\pi L^3}{2G_5}\left[\frac{R^2}{\epsilon^2} - \log\frac{2R}{\epsilon} - \frac{1}{2}\right] \;=\; N^2\,\frac{R^2}{\epsilon^2} \;-\; N^2\log\frac{R}{\epsilon} \;+\; \cdots. \tag{21.32}$$

The quadratic divergence is the area law with a non-universal coefficient. The logarithm is universal: for a CFT$_4$ the coefficient of $-\log(R/\epsilon)$ for a ball is $4a$, with a the Weyl-anomaly coefficient of (14.2) [127]. Here $4a = N^2$, i.e. $a = N^2/4$, the large-N value of $a = (N^2-1)/4$ for $\mathcal{N}=4$ SYM and the holographic anomaly (14.15) [65].

21.4 Consistency checks: entropy inequalities from surgery

A correct entropy formula must satisfy the theorems of Chapter 19. For RT the hardest of them becomes the easiest.

- *Strong subadditivity.* Theorem 19.6 states $S_{AB} + S_{BC} \geq S_B + S_{ABC}$ for any state. Holographically (Figure 21.3): let A, B, C be adjacent boundary regions and γ_{AB}, γ_{BC} the minimal surfaces for AB and BC. Cutting the two surfaces where they intersect and regluing the pieces gives two new surfaces with the same total area: $\tilde\gamma_B$, anchored on ∂B and homologous to B, and $\tilde\gamma_{ABC}$, anchored on $\partial(ABC)$ and homologous to ABC. These are admissible, so their areas bound the minima defining S_B and S_{ABC} from above (they are never the minima themselves: they are kinked where they were reglued, and rounding off a kink reduces the area, so the inequality is generically strict). Hence

$$S_{AB} + S_{BC} \;=\; \frac{\text{Area}(\gamma_{AB}) + \text{Area}(\gamma_{BC})}{4G} \;=\; \frac{\text{Area}(\tilde\gamma_B) + \text{Area}(\tilde\gamma_{ABC})}{4G} \;\geq\; S_B + S_{ABC}. \tag{21.33}$$

  The alternative regluing, into surfaces homologous to A and to C, proves the second form $S_{AB} + S_{BC} \geq S_A + S_C$. That the Lieb–Ruskai theorem [102] becomes elementary geometry is strong evidence that (21.1) is a consistent entropy formula [127].

(a) minimal surfaces for AB and BC

(b) reglued: candidates for B and ABC

Figure 21.3: Strong subadditivity by surgery. (a) The minimal surfaces for the overlapping regions AB and BC intersect in the bulk. (b) Cutting at the intersection and regluing yields one surface anchored on ∂B and one anchored on $\partial(ABC)$, with the *same total area*; since these are admissible but not minimal for B and ABC, minimality gives $S_{AB} + S_{BC} \geq S_B + S_{ABC}$.

- *Araki–Lieb and the homology constraint.* In a mixed state $S_A \neq S_{\bar{A}}$, and Araki–Lieb (Theorem 19.5) requires $|S_A - S_{\bar{A}}| \leq S_{A\bar{A}}$. Since $A\bar{A}$ is the whole boundary, $S_{A\bar{A}}$ is the entropy of the thermal state itself, $S_{\text{th}} = s_{\text{th}} V$ with V the circumference of the boundary, which RT gives as the horizon area (21.18). The homology rule is what makes RT obey the bound. Figure 21.4 shows the one-sided BTZ slice, an annulus whose inner edge is the horizon,[1] with A an interval of length ℓ. The geodesic γ_A is never admissible for $\bar{A}$, because the region between them has the horizon as a further boundary. The complement has two candidates, γ_A together with the horizon (panel (a)) and the geodesic homotopic to $\bar{A}$ (panel (b)), and RT takes the shorter. Their entropies are $S_A + S_{\text{th}}$ and (21.16) with $\ell \to V - \ell$, so with (21.18), and $\sinh x \simeq \frac{1}{2}\mathrm{e}^x$ for $V \gg \beta$,

$$S_{(\mathrm{b})} - S_{(\mathrm{a})} \;=\; \frac{c}{3}\log\frac{\sinh\frac{\pi(V-\ell)}{\beta}}{\sinh\frac{\pi\ell}{\beta}} \;-\; \frac{\pi c}{3\beta}V \;\simeq\; -\frac{c}{3}\log\left(\mathrm{e}^{2\pi\ell/\beta} - 1\right), \tag{21.34}$$

which is positive for $\ell < \ell_p = \frac{\beta}{2\pi}\log 2$. Hence

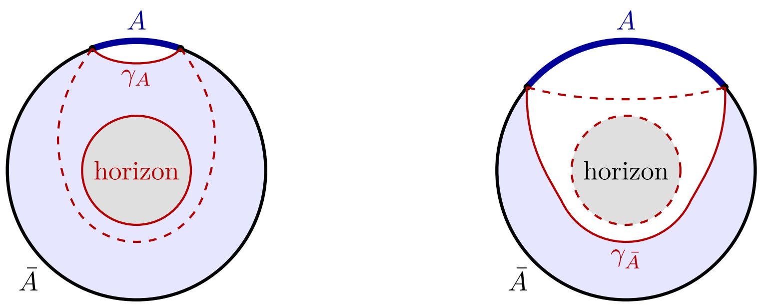
$$S_{\bar{A}} \;=\; S_A + S_{\text{th}} \qquad \text{for} \qquad \ell < \ell_p: \tag{21.35}$$

the complement of a small interval carries the whole thermal mixedness, which RT displays as the horizon inside its homology region, and Araki–Lieb is saturated, the entanglement plateau of [128]. Without the homology constraint RT would give $S_A = S_{\bar{A}}$ always, the signature of a pure state, and would be wrong in mixed states.

21.5 The Lewkowycz–Maldacena derivation

Lewkowycz and Maldacena [45] derived (21.1) from the rules of holography. The idea is to repeat the black-hole computation of Section 20.6.1 for an arbitrary region A. There $\operatorname{Tr}\rho^n$ was the partition function on a thermal circle n times longer, the bulk geometry was the smooth cigar in which that circle shrinks at the horizon, and the entropy came out as the horizon area. Here the circle is the one around ∂A, and the surface on which it shrinks in the bulk will turn out to be γ_A. We present the argument for Einstein gravity and a time-reflection-symmetric state, in four steps (Figure 21.5).

[1]The inner edge is the bifurcation surface of the eternal black hole, where the slice stops because the second exterior has been traced out. For a black hole formed by collapse the slice continues through the horizon to the centre, the empty surface is homologous to the whole boundary, and RT gives $S_{A\bar{A}} = 0$: a pure state, as it must be.



(a) $\ell < \ell_p$: $\gamma_{\bar{A}} = \gamma_A \cup \text{horizon}$ (solid), $S_{\bar{A}} = S_A + S_{\text{th}}$

(b) $\ell > \ell_p$: $\gamma_{\bar{A}}$ homotopic to $\bar{A}$ (solid), $S_{\bar{A}} \neq S_A$

Figure 21.4: The homology rule for the complement $\bar{A}$ of an interval A on a time slice of the one-sided BTZ black hole, with the horizon as inner edge. γ_A alone is never admissible for $\bar{A}$, because the region between them has the horizon as a further boundary. The two admissible surfaces are $\gamma_A \cup$ horizon and the geodesic homotopic to $\bar{A}$; the minimal one is drawn solid, the other dashed, and the shaded region is the homology region of the minimal one. The crossover ℓ_p is given in (21.35).

1. *Boundary replica.* By (19.36) it suffices to know $\operatorname{Tr}\rho_A^n$ for integer n. As in Section 20.3.1, $\operatorname{Tr}\rho_A^n = Z[M_n]/Z[M_1]^n$, where M_1 is the Euclidean boundary geometry and M_n is built from n copies of M_1, each cut open along A and glued cyclically to the next (panel (a)). M_n is smooth except at ∂A, the *branch points*: a small circle around a point of ∂A passes through all n sheets before closing. For A the whole boundary of a thermal state this circle is the thermal circle of length $n\beta$, and we are back in Section 20.6.1. This step is pure CFT.

2. *Bulk saddle.* The dictionary of Chapter 13 gives $Z[M_n] \simeq \mathrm{e}^{-I[B_n]}$, with B_n the least-action bulk solution whose conformal boundary is M_n. In the bulk the circle around ∂A is contractible: it shrinks on a codimension-two surface γ_n anchored on ∂A, as the thermal circle shrinks at the horizon of the cigar. We assume that B_n is smooth at γ_n, the Einstein equations having filled in the branch points as they fill in the tip of the cigar, and that it has the $\mathbb{Z}_n$ symmetry which cyclically permutes the n sheets; γ_n is the fixed locus of this symmetry (panel (b)).

3. *Quotient.* The boundaries M_n of the family B_n change with n, which makes a derivative at $n = 1$ awkward. Pass instead to $\hat{B}_n = B_n/\mathbb{Z}_n$, one sheet with its two edges glued (panel (c)), whose boundary is M_1 for every n. Away from γ_n the quotient is a fundamental domain, one n-th of B_n. At γ_n the angle around the surface, 2π in the smooth B_n, is divided by n, so the quotient has a conical deficit

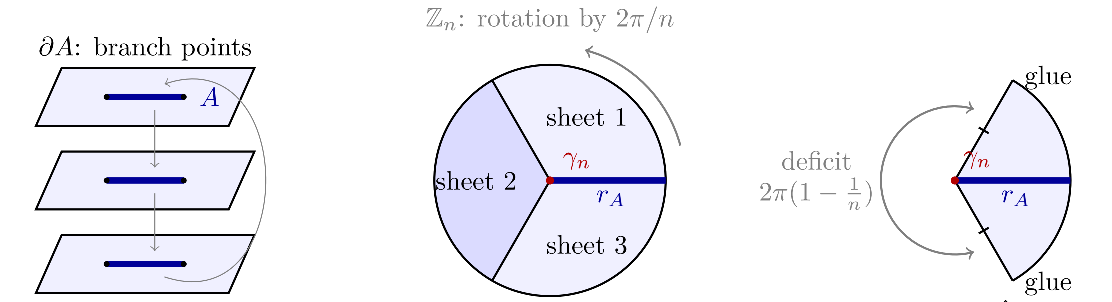
$$\delta_n \;=\; 2\pi\Big(1 - \frac{1}{n}\Big) \tag{21.36}$$

along γ_n. The action is a local integral and γ_n has measure zero in the smooth B_n, so

$$I[B_n] \;=\; n\,\hat{I}_n, \qquad \hat{I}_n \;\equiv\; \text{the action of } \hat{B}_n \text{ with the tip } \gamma_n \text{ left out.} \tag{21.37}$$

Leaving the tip out matters: the cone in $\hat{B}_n$ carries the delta-function curvature (7.30), which is absent in the smooth B_n and must not be counted in $I[B_n]/n$. Since $\hat{I}_1 = I[B_1]$,

$$S \;=\; -\,\partial_n\Big[\log Z[M_n] - n\log Z[M_1]\Big]_{n=1} \;=\; \partial_n\Big[n(\hat{I}_n - \hat{I}_1)\Big]_{n=1} \;=\; \partial_n \hat{I}_n\Big|_{n=1}. \tag{21.38}$$



(a) boundary M_n ($n = 3$): copies of M_1 glued cyclically along A

(b) bulk B_n, plane normal to γ_n: smooth, total angle 2π

(c) quotient $\hat{B}_n = B_n/\mathbb{Z}_n$: one sheet, edges glued, cone along γ_n

Figure 21.5: The gravitational replica trick. (a) The boundary replica manifold M_n, drawn for $n = 3$: n copies of the Euclidean boundary M_1, cut along A and glued cyclically, with branch points at ∂A. (b) The bulk saddle B_n in the plane normal to the fixed surface γ_n: the n sheets meet smoothly, the total angle is 2π, and $\mathbb{Z}_n$ rotates by $2\pi/n$; the sheets are glued along the cut r_A. (c) The quotient $\hat{B}_n = B_n/\mathbb{Z}_n$: one sheet with its edges identified, a cone of deficit (21.36) along γ_n.

The continuation in n is now geometric: $\hat{B}_n$ is the solution of the Einstein equations with boundary M_1 and the deficit (21.36) along a surface anchored on ∂A, which makes sense for real n.

4. *Localization at the cone.* To evaluate $\partial_n \hat{I}_n$ at $n = 1$, put the tip back. By (7.30) the cone contributes
$$I_{\text{tip}} \;=\; -\frac{1}{16\pi G}\cdot 2\,\delta_n\,\text{Area}(\gamma_n) \;=\; -\Big(1-\frac{1}{n}\Big)\frac{\text{Area}(\gamma_n)}{4G}, \tag{21.39}$$
and $I_{\text{full}}[\hat{B}_n] = \hat{I}_n + I_{\text{tip}}$ is the action of $\hat{B}_n$ regarded as a trial configuration for the $n = 1$ problem, with boundary M_1 and the tip included. It differs from the true $n = 1$ saddle B_1 by $O(n-1)$, and B_1 is stationary, so $I_{\text{full}}[\hat{B}_n] = I[B_1] + O((n-1)^2)$.[2] Hence
$$\hat{I}_n \;=\; I_{\text{full}}[\hat{B}_n] - I_{\text{tip}} \;=\; I[B_1] - I_{\text{tip}} + O\big((n-1)^2\big) \;=\; \hat{I}_1 + \Big(1-\frac{1}{n}\Big)\frac{\text{Area}(\gamma_n)}{4G} + O\big((n-1)^2\big). \tag{21.40}$$

Substituting (21.40) into (21.38):

$$\boxed{S \;=\; \partial_n \hat{I}_n\Big|_{n=1} \;=\; \frac{\text{Area}(\gamma)}{4G},} \tag{21.41}$$
with $\gamma = \gamma_{n\to 1}$ the limiting fixed-point surface: the gravitational replica trick derives the Ryu–Takayanagi formula [45].

Three remarks complete the derivation.

1. *Extremality and minimality.* γ_n is not chosen: it is where the circle closes off in the solution B_n. As $n \to 1$ the Einstein equations at the tip reduce to the condition that the $O(n-1)$ term of (21.40) be stationary under deformations of the surface, i.e. that γ be extremal; among extremal surfaces the dominant saddle has least action, hence least area.

[2] Write $\hat{B}_n = B_1 + \delta g$ with $\delta g = O(n-1)$, the deficit (21.36) being $O(n-1)$ and the geometry away from the tip depending smoothly on n. Expanding at fixed boundary M_1, $I_{\text{full}}[B_1 + \delta g] = I[B_1] + \frac{\delta I}{\delta g}\big|_{B_1}\delta g + O(\delta g^2)$, and the linear term vanishes because B_1 solves the equations of motion. The tip must be included for this to apply: the cone is a limit of smooth metrics and, by (7.30), its action with the tip counted is the limit of their actions, so I_{full} is the functional whose first variation vanishes at B_1, whereas $\hat{I}_n$ is not.

2. *Homology.* The n sheets of B_n are glued along a bulk region ending on A and on γ_n, the cut r_A of panels (b) and (c), so γ and A together bound a region: the homology condition of (21.1) is built in.

3. *Assumptions.* The dominant saddle was assumed replica symmetric and the continuation in n to exist. Saddles that break the replica symmetry, the replica wormholes, are the ones invoked by the island proposal of Section 19.10.3.

Remark 21.1 (Spheres without replicas: Casini–Huerta–Myers)**.** For a ball in the vacuum of any CFT there is a derivation that avoids the continuation in n [127]. A conformal transformation maps the causal development of the ball to a static patch of hyperbolic space and the reduced density matrix to a thermal state there, a generalization of the Rindler result (6.9). Holographically the hyperbolic thermal state is dual to a hyperbolic black hole, whose Bekenstein–Hawking entropy, computed as in Chapter 7, maps back to the area of the hemisphere (21.29) over $4G$. For spheres RT is black hole thermodynamics in disguise, which is why (21.28) has an exact solution.

21.5.1 Refinements: HRT, quantum extremal surfaces, islands

Three generalizations connect this lecture to current research.

1. *Covariance.* In time-dependent spacetimes there is no preferred slice, and minimal surfaces are replaced by extremal spacelike surfaces in the Lorentzian geometry, the Hubeny–Rangamani–Takayanagi prescription [129]; the Lewkowycz–Maldacena argument extends to this setting with real-time path integrals.

2. *Quantum corrections.* At one loop the boundary entropy is the area term plus the bulk entanglement entropy of the quantum fields across γ [130], and at the next level of precision one extremizes the generalized entropy of Chapter 20,

$$S_A \;=\; \min_{\gamma}\ \operatorname*{ext}_{\gamma}\left[\frac{\mathrm{Area}(\gamma)}{4G} + S_{\mathrm{bulk,\,out}}(\gamma)\right] \;=\; \min\ \operatorname*{ext}_{\gamma} S_{\mathrm{gen}}(\gamma), \tag{21.42}$$

the quantum extremal surface prescription [131].

3. *Islands.* Applied to an evaporating black hole coupled to a reservoir, (21.42) for the radiation admits, after the Page time, an extremum whose surface lies just inside the horizon, so that the entropy region includes a disconnected island in the interior [113, 114]; the competition between the extrema without and with an island produces the Page curve of Chapter 19, and the island saddles are the replica wormholes mentioned above. The status of this proposal, and its caveats, were stated in Section 19.10.3; the review [115] and the book [14] tell the story in full. What has not been achieved is a bulk identification of the black hole microstates in a general setting: the island formula computes the entropy unitarity requires without exhibiting the states that carry it.

21.6 Where we stand

- *The arc of the course.* It began by dropping a cup of tea into a black hole and asking what became of its entropy. Classical general relativity gave horizons a mechanics shaped like thermodynamics (Chapter 1); taking the analogy seriously forced $S = A/4G$ (Chapter 2), which required a temperature, which Hawking's calculation supplied together with the information problem (Chapters 4–5). The Unruh effect and the Euclidean path integral showed that this thermality is vacuum entanglement seen through a horizon (Chapters 6–7). AdS/CFT (Chapters 8–17) embedded black holes in a unitary quantum theory, counted their states (Chapter 18), and turned the information problem into a sharp question within quantum mechanics (Chapter 19). Finally $A/4G$ and matter entanglement merged into the cutoff-independent S_{gen} (Chapter 20), and the Ryu–Takayanagi formula extended the relation between entropy and area from horizons to arbitrary regions.

- *Established*, in the sense of having been derived several independent ways:

1. black hole thermodynamics: temperature, entropy and the generalized second law follow from quantum field theory in curved spacetime, from Euclidean gravity, and from dual gauge theories, in agreement;

2. holography: with AdS boundary conditions, quantum gravity in a spacetime is an ordinary quantum theory on its boundary, with a dictionary tested from anomalies to Wilson loops to thermodynamics to entanglement;

3. the entanglement–geometry relation: entanglement entropies are computed by extremal surfaces, entropy inequalities become geometric statements, and the generalized entropy behaves as the fine-grained entropy of gravitational subsystems.

- *Open:*

1. the microstates of a Schwarzschild black hole in asymptotically flat space, without supersymmetry or an AdS box, have never been exhibited; we know how many there are, not what they are;

2. de Sitter space has a horizon entropy but no known dual description, and no agreed answer to what its entropy counts;

3. the black hole interior: no proposal, islands included, tells us what an infalling observer meets at the singularity or how that experience is encoded in the boundary theory;

4. the Euclidean gravitational path integral, on which Chapters 7, 18, 20 and this one rest, works far better than we understand.

Every advance recounted in these lectures came from taking gravity's thermodynamic hints at face value, beginning with Bekenstein's in 1972. There is no indication that the hints are exhausted.

Bibliography

[1] E. Witten, *Introduction to Black Hole Thermodynamics*, [arXiv:2412.16795 [hep-th]].

[2] A. Zaffaroni, *Introduction to the AdS-CFT correspondence*, Class. Quant. Grav. **17** (2000) 3571. Extended lecture notes, Lausanne 2008/LACES 2023 version.

[3] R.M. Wald, *General Relativity*, Chicago Univ. Press (1984).

[4] N.D. Birrell and P.C.W. Davies, *Quantum Fields in Curved Space*, Cambridge Univ. Press (1982).

[5] R.M. Wald, *Quantum Field Theory in Curved Space-Time and Black Hole Thermodynamics*, University of Chicago Press (1995).

[6] P. Di Francesco, P. Mathieu and D. Senechal, *Conformal Field Theory*, Springer-Verlag (1997).

[7] S. Rychkov, *EPFL Lectures on Conformal Field Theory in D >= 3 Dimensions*, SpringerBriefs in Physics (2016) [arXiv:1601.05000 [hep-th]].

[8] D. Tong, *String Theory*, [arXiv:0908.0333 [hep-th]].

[9] J. Polchinski, *String theory. Vol. 1: An introduction to the bosonic string*, Cambridge University Press (1998).

[10] J. Polchinski, *String theory. Vol. 2: Superstring theory and beyond*, Cambridge University Press (1998).

[11] O. Aharony, S.S. Gubser, J.M. Maldacena, H. Ooguri and Y. Oz, *Large N field theories, string theory and gravity*, Phys. Rept. **323** (2000) 183 [hep-th/9905111].

[12] K. Skenderis, *Lecture notes on holographic renormalization*, Class. Quant. Grav. **19** (2002) 5849 [hep-th/0209067].

[13] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (2012).

[14] M. Rangamani and T. Takayanagi, *Holographic Entanglement Entropy*, Springer (2017) [arXiv:1609.01287 [hep-th]].

[15] S.W. Hawking, *Gravitational radiation from colliding black holes*, Phys. Rev. Lett. **26** (1971) 1344.

[16] D. Christodoulou, *Reversible and irreversible transformations in black hole physics*, Phys. Rev. Lett. **25** (1970) 1596.

[17] J.M. Bardeen, B. Carter and S.W. Hawking, *The Four laws of black hole mechanics*, Commun. Math. Phys. **31** (1973) 161.

[18] J.D. Bekenstein, *Black holes and entropy*, Phys. Rev. D **7** (1973) 2333.

[19] A. Strominger and C. Vafa, *Microscopic origin of the Bekenstein-Hawking entropy*, Phys. Lett. B **379** (1996) 99 [hep-th/9601029].

[20] T. Regge and J.A. Wheeler, *Stability of a Schwarzschild singularity*, Phys. Rev. **108** (1957) 1063.

[21] D.G. Boulware, *Quantum Field Theory in Schwarzschild and Rindler Spaces*, Phys. Rev. D **11** (1975) 1404.

[22] P. Candelas, *Vacuum Polarization in Schwarzschild Space-Time*, Phys. Rev. D **21** (1980) 2185.

[23] W.G. Unruh, *Notes on black hole evaporation*, Phys. Rev. D **14** (1976) 870.

[24] J.B. Hartle and S.W. Hawking, *Path Integral Derivation of Black Hole Radiance*, Phys. Rev. D **13** (1976) 2188.

[25] S.W. Hawking, *Black hole explosions?*, Nature **248** (1974) 30.

[26] S.W. Hawking, *Particle Creation by Black Holes*, Commun. Math. Phys. **43** (1975) 199.

[27] R. Kubo, *Statistical mechanical theory of irreversible processes. 1. General theory and simple applications in magnetic and conduction problems*, J. Phys. Soc. Jap. **12** (1957) 570.

[28] P.C. Martin and J. Schwinger, *Theory of many particle systems. 1.*, Phys. Rev. **115** (1959) 1342.

[29] D.N. Page, *Particle Emission Rates from a Black Hole: Massless Particles from an Uncharged, Nonrotating Hole*, Phys. Rev. D **13** (1976) 198.

[30] S.W. Hawking, *Breakdown of Predictability in Gravitational Collapse*, Phys. Rev. D **14** (1976) 2460.

[31] G.W. Gibbons and S.W. Hawking, *Cosmological Event Horizons, Thermodynamics, and Particle Creation*, Phys. Rev. D **15** (1977) 2738.

[32] A.A. Starobinsky and S.M. Churilov, *Amplification of electromagnetic and gravitational waves scattered by a rotating black hole*, Sov. Phys. JETP **65** (1974) 1.

[33] J.W. York, *Black hole thermodynamics and the Euclidean Einstein action*, Phys. Rev. D **33** (1986) 2092.

[34] J.J. Bisognano and E.H. Wichmann, *On the Duality Condition for Quantum Fields*, J. Math. Phys. **17** (1976) 303.

[35] R.F. Streater and A.S. Wightman, *PCT, Spin and Statistics, and All That*, Princeton University Press (2000).

[36] R. Haag, *Local Quantum Physics: Fields, Particles, Algebras*, Springer (1996).

[37] G.L. Sewell, *Quantum fields on manifolds: PCT and gravitationally induced thermal states*, Annals Phys. **141** (1982) 201.

[38] B.S. DeWitt, *Quantum gravity: the new synthesis*, in *General Relativity: An Einstein Centenary Survey*, S.W. Hawking and W. Israel eds., Cambridge University Press (1979) 680.

[39] R.C. Tolman, *On the Weight of Heat and Thermal Equilibrium in General Relativity*, Phys. Rev. **35** (1930) 904.

[40] R. Tolman and P. Ehrenfest, *Temperature Equilibrium in a Static Gravitational Field*, Phys. Rev. **36** (1930) 1791.

[41] W. Israel, *Thermo field dynamics of black holes*, Phys. Lett. A **57** (1976) 107.

[42] J.W. York, *Role of conformal three geometry in the dynamics of gravitation*, Phys. Rev. Lett. **28** (1972) 1082.

[43] G.W. Gibbons and S.W. Hawking, *Action Integrals and Partition Functions in Quantum Gravity*, Phys. Rev. D **15** (1977) 2752.

[44] E. Poisson, *A Relativist's Toolkit: The Mathematics of Black-Hole Mechanics*, Cambridge University Press (2009).

[45] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, JHEP **08** (2013) 090 [arXiv:1304.4926 [hep-th]].

[46] M. Banados, C. Teitelboim and J. Zanelli, *Black hole entropy and the dimensional continuation of the Gauss-Bonnet theorem*, Phys. Rev. Lett. **72** (1994) 957 [gr-qc/9309026].

[47] D.V. Fursaev and S.N. Solodukhin, *On the description of the Riemannian geometry in the presence of conical defects*, Phys. Rev. D **52** (1995) 2133 [hep-th/9501127].

[48] G.W. Gibbons, S.W. Hawking and M.J. Perry, *Path Integrals and the Indefiniteness of the Gravitational Action*, Nucl. Phys. B **138** (1978) 141.

[49] J.M. Maldacena, *Eternal black holes in anti-de Sitter*, JHEP **04** (2003) 021 [hep-th/0106112].

[50] K. Osterwalder and R. Schrader, *Axioms for Euclidean Green's Functions*, Commun. Math. Phys. **31** (1973) 83.

[51] L.D. Landau and E.M. Lifshitz, *Mechanics: Course of Theoretical Physics, Vol. 1*, Butterworth-Heinemann (1976).

[52] P. Breitenlohner and D.Z. Freedman, *Stability in Gauged Extended Supergravity*, Annals Phys. **144** (1982) 249.

[53] A.M. Essin and D.J. Griffiths, *Quantum mechanics of the 1/x**2 potential*, Am. J. Phys. **74** (2006) 109.

[54] A. Ishibashi and R.M. Wald, *Dynamics in non-globally-hyperbolic static spacetimes. III. Anti-de Sitter spacetime*, Class. Quant. Grav. **21** (2004) 2981 [hep-th/0402184].

[55] G. 't Hooft, *A Planar Diagram Theory for Strong Interactions*, Nucl. Phys. B **72** (1974) 461.

[56] G. 't Hooft, *Dimensional reduction in quantum gravity*, Conf. Proc. C **930308** (1993) 284 [gr-qc/9310026].

[57] L. Susskind, *The World as a hologram*, J. Math. Phys. **36** (1995) 6377 [hep-th/9409089].

[58] J.M. Maldacena, *The Large N limit of superconformal field theories and supergravity*, Adv. Theor. Math. Phys. **2** (1998) 231 [hep-th/9711200].

[59] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, *Gauge theory correlators from noncritical string theory*, Phys. Lett. B **428** (1998) 105 [hep-th/9802109].

[60] E. Witten, *Anti-de Sitter space and holography*, Adv. Theor. Math. Phys. **2** (1998) 253 [hep-th/9802150].

[61] D.Z. Freedman, S.D. Mathur, A. Matusis and L. Rastelli, *Correlation functions in the CFT(d) / AdS(d+1) correspondence*, Nucl. Phys. B **546** (1999) 96 [hep-th/9804058].

[62] J.L. Cardy, *Is There a c Theorem in Four-Dimensions?*, Phys. Lett. B **215** (1988) 749.

[63] Z. Komargodski and A. Schwimmer, *On Renormalization Group Flows in Four Dimensions*, JHEP **12** (2011) 099 [arXiv:1107.3987 [hep-th]].

[64] M.J. Duff, *Twenty years of the Weyl anomaly*, Class. Quant. Grav. **11** (1994) 1387 [hep-th/9308075].

[65] M. Henningson and K. Skenderis, *The Holographic Weyl anomaly*, JHEP **07** (1998) 023 [hep-th/9806087].

[66] C. Fefferman and C.R. Graham, *Conformal invariants*, Astérisque, hors série (1985) 95.

[67] P. Mansfield and D. Nolland, *One loop conformal anomalies from AdS/CFT in the Schrodinger representation*, JHEP **07** (1999) 028 [hep-th/9906054].

[68] A. Bilal and C. Chu, *A Note on the chiral anomaly in the AdS/CFT correspondence and 1/N**2 correction*, Nucl. Phys. B **562** (1999) 181 [hep-th/9907106].

[69] J.D. Brown and M. Henneaux, *Central Charges in the Canonical Realization of Asymptotic Symmetries: An Example from Three-Dimensional Gravity*, Commun. Math. Phys. **104** (1986) 207.

[70] C.G. Callan, D. Friedan, E.J. Martinec and M.J. Perry, *Strings in Background Fields*, Nucl. Phys. B **262** (1985) 593.

[71] G.T. Horowitz and A. Strominger, *Black strings and P-branes*, Nucl. Phys. B **360** (1991) 197.

[72] J. Polchinski, *Dirichlet Branes and Ramond-Ramond charges*, Phys. Rev. Lett. **75** (1995) 4724 [hep-th/9510017].

[73] S.S. Gubser, I.R. Klebanov and A.W. Peet, *Entropy and temperature of black 3-branes*, Phys. Rev. D **54** (1996) 3915 [hep-th/9602135].

[74] C. Montonen and D.I. Olive, *Magnetic Monopoles as Gauge Particles?*, Phys. Lett. B **72** (1977) 117.

[75] F.A. Dolan and H. Osborn, *On short and semi-short representations for four-dimensional superconformal symmetry*, Annals Phys. **307** (2003) 41 [hep-th/0209056].

[76] D. Anselmi, M.T. Grisaru and A. Johansen, *Universality of the operator product expansions of SCFT in four-dimensions*, Phys. Lett. B **394** (1997) 329 [hep-th/9608125].

[77] N. Gromov, V. Kazakov and P. Vieira, *Exact Spectrum of Planar* $\mathcal{N} = 4$ *Supersymmetric Yang-Mills Theory: Konishi Dimension at Any Coupling*, Phys. Rev. Lett. **104** (2010) 211601 [arXiv:0906.4240].

[78] H.J. Kim, L.J. Romans and P. van Nieuwenhuizen, *The Mass Spectrum of Chiral N=2 D=10 Supergravity on S**5*, Phys. Rev. D **32** (1985) 389.

[79] M. Gunaydin and N. Marcus, *The Spectrum of the S**5 Compactification of the Chiral N=2, D=10 Supergravity and the Unitary Supermultiplets of U(2,2/4)*, Class. Quant. Grav. **2** (1985) L11.

[80] S. Ferrara and C. Fronsdal, *Conformal Maxwell theory as a singleton field theory on AdS(5), IIB three-branes and duality*, Class. Quant. Grav. **15** (1998) 2153 [hep-th/9712239].

[81] S. Lee, S. Minwalla, M. Rangamani and N. Seiberg, *Three point functions of chiral operators in D = 4, N=4 SYM at large N*, Adv. Theor. Math. Phys. **2** (1998) 697 [hep-th/9806074].

[82] K.A. Intriligator, *Bonus symmetries of N=4 superYang-Mills correlation functions via AdS duality*, Nucl. Phys. B **551** (1999) 575 [hep-th/9811047].

[83] V. Pestun, *Localization of gauge theory on a four-sphere and supersymmetric Wilson loops*, Commun. Math. Phys. **313** (2012) 71 [arXiv:0712.2824 [hep-th]].

[84] J.M. Maldacena, *Wilson loops in large N field theories*, Phys. Rev. Lett. **80** (1998) 4859 [hep-th/9803002].

[85] J.K. Erickson, G.W. Semenoff, R.J. Szabo and K. Zarembo, *Static potential in N=4 supersymmetric Yang-Mills theory*, Phys. Rev. D **61** (2000) 105006 [hep-th/9911088].

[86] S. Rey and J. Yee, *Macroscopic strings as heavy quarks in large N gauge theory and anti-de Sitter supergravity*, Eur. Phys. J. C **22** (2001) 379 [hep-th/9803001].

[87] J.K. Erickson, G.W. Semenoff and K. Zarembo, *Wilson loops in N=4 supersymmetric Yang-Mills theory*, Nucl. Phys. B **582** (2000) 155 [hep-th/0003055].

[88] N. Drukker and D.J. Gross, *An Exact prediction of N=4 SUSYM theory for string theory*, J. Math. Phys. **42** (2001) 2896 [hep-th/0010274].

[89] D. Medina-Rincon, A.A. Tseytlin and K. Zarembo, *Precision matching of circular Wilson loops and strings in* $AdS_5\times S^5$, JHEP **05** (2018) 199 [arXiv:1804.08925 [hep-th]].

[90] J.A. Minahan and K. Zarembo, *The Bethe ansatz for N=4 super Yang-Mills*, JHEP **03** (2003) 013 [hep-th/0212208].

[91] N. Beisert and others, *Review of AdS/CFT Integrability: An Overview*, Lett. Math. Phys. **99** (2012) 3 [arXiv:1012.3982].

[92] M.R. Douglas and G.W. Moore, *D-branes, quivers, and ALE instantons*, [hep-th/9603167].

[93] S. Kachru and E. Silverstein, *4-D conformal theories and strings on orbifolds*, Phys. Rev. Lett. **80** (1998) 4855 [hep-th/9802183].

[94] I.R. Klebanov and E. Witten, *Superconformal field theory on three-branes at a Calabi-Yau singularity*, Nucl. Phys. B **536** (1998) 199 [hep-th/9807080].

[95] I.R. Klebanov and M.J. Strassler, *Supergravity and a confining gauge theory: Duality cascades and chi SB resolution of naked singularities*, JHEP **08** (2000) 052 [hep-th/0007191].

[96] S.W. Hawking and D.N. Page, *Thermodynamics of Black Holes in anti-De Sitter Space*, Commun. Math. Phys. **87** (1983) 577.

[97] E. Witten, *Anti-de Sitter space, thermal phase transition, and confinement in gauge theories*, Adv. Theor. Math. Phys. **2** (1998) 505 [hep-th/9803131].

[98] B. Sundborg, *The Hagedorn transition, deconfinement and N=4 SYM theory*, Nucl. Phys. B **573** (2000) 349 [hep-th/9908001].

[99] O. Aharony, J. Marsano, S. Minwalla, K. Papadodimas and M. Van Raamsdonk, *The Hagedorn - deconfinement phase transition in weakly coupled large N gauge theories*, Adv. Theor. Math. Phys. **8** (2004) 603 [hep-th/0310285].

[100] A. Fotopoulos and T.R. Taylor, *Comment on two loop free energy in N=4 supersymmetric Yang-Mills theory at finite temperature*, Phys. Rev. D **59** (1999) 061701 [hep-th/9811224].

[101] S.S. Gubser, I.R. Klebanov and A.A. Tseytlin, *Coupling constant dependence in the thermodynamics of N=4 supersymmetric Yang-Mills theory*, Nucl. Phys. B **534** (1998) 202 [hep-th/9805156].

[102] E.H. Lieb and M.B. Ruskai, *Proof of the strong subadditivity of quantum-mechanical entropy*, J. Math. Phys. **14** (1973) 1938.

[103] H. Casini, *Relative entropy and the Bekenstein bound*, Class. Quant. Grav. **25** (2008) 205021 [arXiv:0804.2182 [hep-th]].

[104] P. Calabrese and J. Cardy, *Entanglement entropy and quantum field theory*, J. Stat. Mech. **0406** (2004) P06002 [hep-th/0405152].

[105] M. Srednicki, *Entropy and area*, Phys. Rev. Lett. **71** (1993) 666 [hep-th/9303048].

[106] L. Bombelli, R.K. Koul, J. Lee and R.D. Sorkin, *A Quantum Source of Entropy for Black Holes*, Phys. Rev. D **34** (1986) 373.

[107] D.N. Page, *Average entropy of a subsystem*, Phys. Rev. Lett. **71** (1993) 1291 [gr-qc/9305007].

[108] S.K. Foong and S. Kanno, *Proof of Page's conjecture on the average entropy of a subsystem*, Phys. Rev. Lett. **72** (1994) 1148.

[109] S. Sen, *Average entropy of a subsystem*, Phys. Rev. Lett. **77** (1996) 1 [hep-th/9601132].

[110] P. Hayden, D.W. Leung and A. Winter, *Aspects of generic entanglement*, Commun. Math. Phys. **265** (2006) 95 [quant-ph/0407049].

[111] D.N. Page, *Information in black hole radiation*, Phys. Rev. Lett. **71** (1993) 3743 [hep-th/9306083].

[112] D.N. Page, *Time Dependence of Hawking Radiation Entropy*, JCAP **09** (2013) 028 [arXiv:1301.4995 [hep-th]].

[113] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, JHEP **09** (2020) 002 [arXiv:1905.08255 [hep-th]].

[114] A. Almheiri, N. Engelhardt, D. Marolf and H. Maxfield, *The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole*, JHEP **12** (2019) 063 [arXiv:1905.08762 [hep-th]].

[115] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *The entropy of Hawking radiation*, Rev. Mod. Phys. **93** (2021) 035002 [arXiv:2006.06872 [hep-th]].

[116] J.S. Dowker, *Quantum Field Theory on a Cone*, J. Phys. A **10** (1977) 115.

[117] L. Susskind and J. Uglum, *Black hole entropy in canonical quantum gravity and superstring theory*, Phys. Rev. D **50** (1994) 2700 [hep-th/9401070].

[118] C.G. Callan and F. Wilczek, *On geometric entropy*, Phys. Lett. B **333** (1994) 55 [hep-th/9401072].

[119] S.N. Solodukhin, *Entanglement entropy of black holes*, Living Rev. Rel. **14** (2011) 8 [arXiv:1104.3712 [hep-th]].

[120] J.D. Bekenstein, *A Universal Upper Bound on the Entropy to Energy Ratio for Bounded Systems*, Phys. Rev. D **23** (1981) 287.

[121] W.G. Unruh and R.M. Wald, *Acceleration Radiation and Generalized Second Law of Thermodynamics*, Phys. Rev. D **25** (1982) 942.

[122] D.D. Blanco, H. Casini, L. Hung and R.C. Myers, *Relative Entropy and Holography*, JHEP **08** (2013) 060 [arXiv:1305.3182 [hep-th]].

[123] A.C. Wall, *A proof of the generalized second law for rapidly changing fields and arbitrary horizon slices*, Phys. Rev. D **85** (2012) 104049 [arXiv:1105.3445 [gr-qc]]. [Erratum: Phys. Rev. D 87, 069904 (2013)].

[124] X. Dong, *The Gravity Dual of Renyi Entropy*, Nature Commun. **7** (2016) 12472 [arXiv:1601.06788 [hep-th]].

[125] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, Phys. Rev. Lett. **96** (2006) 181602 [hep-th/0603001].

[126] S.N. Solodukhin, *Entanglement entropy, conformal invariance and extrinsic geometry*, Phys. Lett. B **665** (2008) 305 [arXiv:0802.3117 [hep-th]].

[127] H. Casini, M. Huerta and R.C. Myers, *Towards a derivation of holographic entanglement entropy*, JHEP **05** (2011) 036 [arXiv:1102.0440 [hep-th]].

[128] V.E. Hubeny, H. Maxfield, M. Rangamani and E. Tonni, *Holographic entanglement plateaux*, JHEP **08** (2013) 092 [arXiv:1306.4004 [hep-th]].

[129] V.E. Hubeny, M. Rangamani and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, JHEP **07** (2007) 062 [arXiv:0705.0016 [hep-th]].

[130] T. Faulkner, A. Lewkowycz and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, JHEP **11** (2013) 074 [arXiv:1307.2892 [hep-th]].

[131] N. Engelhardt and A.C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, JHEP **01** (2015) 073 [arXiv:1408.3203 [hep-th]].