# The Quantum–Geometric Correspondence

## from the Ground Up

A fully self-contained course in fifteen chapters

assuming only undergraduate quantum mechanics,
special relativity, and the will to ask questions

Marc Sperzel

Foundations Track · Quantum–Geometric Correspondence
First edition draft · July 2026

# Preface: how to read this book

## What this book is

This volume teaches one physical theory — the *Quantum–Geometric Correspondence* (QGC) — from the ground up. The theory's claim, stated once here and unpacked over two hundred and fifty pages: quantum mechanics and general relativity are not rivals awaiting a winner, but two complementary restrictions of a single underlying structure, and the seam where they meet is directly measurable as a universal, gravitationally induced loss of quantum coherence,

$$\tau_{\mathrm{dec}} = \frac{\hbar\, d}{GM^2},$$

about 1.6 nanoseconds for a microgram mass held in a superposition one millimetre across. Every symbol in that formula, every step that leads to it, and every honest doubt about it will be built up from scratch.

There is an existing six-lecture course in this repository (`LECTURES/tex/`, the *review track*): 39 dense pages in the register of a mathematical-physics monograph, for readers who already carry graduate quantum field theory, general relativity, and operator algebras. This volume is the *foundations track*. It assumes none of that. If you know undergraduate quantum mechanics (states, bras and kets, operators, expectation values), special relativity at the level of time dilation and four-vectors, and calculus with linear algebra — you have everything you need. General relativity, quantum field theory, entropy and information, von Neumann algebras, modular theory, constrained quantization: each is constructed here, in full, before it is used.

## The reader who asks questions

This book is written for a specific person: smart, relentlessly curious, and unwilling to accept "it can be shown." That reader interrupts. Their interruptions appear throughout as numbered question boxes:

**Q** ***"What do the orange boxes mean?"***

Exactly this. Whenever a sharp reader would object — "why is that step allowed?", "couldn't this just be ordinary decoherence?", "who says the temperature is $2\pi$?" — the objection is raised in a box like this one and answered immediately, in full. These boxes are *not* optional asides; they are load-bearing. The blue **For the curious** boxes, by contrast, are optional deep dives you may skip on a first pass.

Each chapter also ends with a FAQ collecting the larger objections, and Appendix D assembles every question box in the volume into a master index.

## The honesty system

Not everything in this theory has the same epistemic standing, and the notation refuses to blur that. Every non-trivial claim carries a tag:

| | |
|---|---|
| [PROVEN] | mathematically rigorous, unconditional |
| [DERIVED] | follows from the axioms under stated, controlled approximations |
| [CONJECTURED] | precisely formulated, supported, but not proven |
| [MOTIVATED] | physically plausible, matched to phenomenology |
| [ASSUMED] | taken as input, exactly as standard quantum mechanics does |

Cool blue tags mark established ground; warm tags mark open ground. The central identity of the whole framework — that the modular flow of the observer's algebra *is* physical time, $K = 2\pi H_{\text{phys}}$ — is [CONJECTURED]: proven in expectation values and solvable models, open as a full operator equation. The theory's headline rate inherits that tag. You will meet the distinction early and often; it is the most scientifically load-bearing feature of these notes.

## Numbers you can run

Every number in this volume is recomputed from the three constants $G = 6.67430 \times 10^{-11}\,\text{m}^3\text{kg}^{-1}\text{s}^{-2}$, $\hbar = 1.054571817 \times 10^{-34}\,\text{J s}$, $c = 299792458\,\text{m/s}$, and each worked example ends with a Run it line naming the exact script in this repository that checks the result to machine precision. Appendix C explains how to run them. If a number in these notes ever disagrees with the harness, the harness wins.

## The shape of the journey

**Part I** states the problem: why the two best theories in physics contradict each other on a question as simple as "where does a superposed mass gravitate?" **Part II** (Chapters 2–7) builds the complete toolkit, from density matrices to the modular clock of an accelerated observer. **Part III** (Chapters 8–13) states the correspondence, lays down its three axioms, derives Einstein's equations from them, and then walks — with no skipped steps — to the gravitational decoherence rate and the one open construction (R1) that separates conjecture from theorem. **Part IV** (Chapters 14–15) takes the theory outdoors: the cosmological sector, and the laboratory experiments that will, within a decade, either confirm the rate or kill it.

The register throughout is the one Einstein used in 1905: define the measured thing before writing its symbol, interpret every formula in plain words immediately after deriving it, and say — in the main text, not a footnote — exactly where each result stops being valid.

# Contents

# Part I

# The Problem

# 1 The Problem Nobody Ordered

**Synopsis.** Physics runs on two rulebooks. Quantum mechanics governs matter and passes every test; general relativity governs spacetime and passes every test. Ask them one joint question — *where does a mass that is in two places at once gravitate?* — and they give incompatible answers. We work through both standard escape routes and watch each fail: coupling gravity to the quantum *average* contradicts experiment and consistency; quantizing the gravitational field destroys predictivity. The failure is confined to one narrow, sharply identifiable regime, and this book is about the theory that lives there. We close with its headline number, computed from three constants: a microgram mass split by a millimetre should lose quantum coherence in about 1.6 nanoseconds.

**You will be able to.** state precisely why semiclassical gravity fails on superpositions; reproduce the power-counting argument against quantizing the metric; compute the Planck scales and the decoherence benchmark $\tau_{\text{dec}} = \hbar d/GM^2$ from $G$, $\hbar$, $c$ alone.

**Prerequisites.** Undergraduate quantum mechanics (superposition, expectation values), Newtonian gravity, dimensional analysis. Nothing else.

## 1.1 Two rulebooks

It is well known that physics, as actually practised, rests on two theories that have never been made to share a single set of rules.

The first is *quantum mechanics*. Operationally it says: the complete description of a physical system is a vector $|\psi\rangle$ in a complex vector space; if $|\psi_1\rangle$ and $|\psi_2\rangle$ are possible states, so is any superposition $c_1 |\psi_1\rangle + c_2 |\psi_2\rangle$; and measurable quantities are represented by operators whose statistics the state fixes. The superposition rule is not a curiosity of small things — it is the load-bearing wall. Every semiconductor, every chemical bond, every interference fringe is the superposition principle at work, and no experiment has ever caught it failing. Molecules of more than 2,000 atoms (mass around $4 \times 10^{-23}$ kg) have been placed in superpositions of position and interfered, and the fringes appeared exactly where quantum mechanics said they would.

The second is *general relativity*. Operationally it says: spacetime is not a fixed stage but a dynamical geometry $g_{\mu\nu}$; matter tells geometry how to curve, through Einstein's field equations

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{1.1}$$

and geometry tells matter how to move. Here $T_{\mu\nu}$ is the stress–energy tensor — the density and flux of energy and momentum — and $G_{\mu\nu}$ is a package of second derivatives of the geometry that we will construct honestly in Chapter 3. For now only one feature matters: *the right-hand side is a definite, classical function of position and time.* One configuration of matter in, one geometry out. General relativity, too, passes every test: planetary orbits, the bending of starlight, GPS clock corrections, binary-pulsar energy loss, gravitational waves, black-hole images.

So we have two rulebooks, each flawless in its own courtroom. The trouble begins the moment a single witness must appear in both.

## 1.2 The question a superposition forces

Let us take a small but macroscopic body — a grain of mass $M$ — and prepare it in a superposition of two locations a distance $d$ apart:

$$|\psi\rangle \;=\; \frac{1}{\sqrt{2}}\Big(|L\rangle + |R\rangle\Big), \tag{1.2}$$

where $|L\rangle$ means "the grain is at the left site" and $|R\rangle$ "at the right site." Quantum mechanics has no objection; states like (1.2) are prepared daily with atoms and molecules, and nothing in the theory puts a mass limit on the procedure.

Now we ask the joint question: *what is the gravitational field of this grain?*

The question is fair — the grain has mass, and everything with mass gravitates — and it is operational: place a small test mass nearby, wait, and watch which way it drifts. Yet neither rulebook can process it alone. Quantum mechanics describes the grain but says nothing about geometry; general relativity describes geometry but demands a *definite* matter configuration on the right-hand side of (1.1), and a superposition is precisely a refusal to be definite.

There are two textbook escape routes. We now take each seriously enough to watch it fail. The failures are instructive: between them they will mark out, with surprising precision, the territory this book explores.

## 1.3 First exit: gravity couples to the average

### 1.3.1 The semiclassical equation

The oldest proposal (Møller and Rosenfeld, around 1960) keeps geometry classical and feeds it the quantum *expectation value* of the matter distribution:

$$G_{\mu\nu}[g] \;=\; \frac{8\pi G}{c^4}\,\langle\psi|\,\hat{T}_{\mu\nu}\,|\psi\rangle\,. \tag{1.3}$$

In words: the source of gravity is not the matter itself but its quantum average. This is called *semiclassical gravity*, and it is less naive than it may look — when matter is in a state with small fluctuations (a planet, a brick), the average is essentially the actual configuration, and (1.3) reduces to ordinary general relativity. All the successes listed above survive.

For our superposed grain, though, (1.3) makes a sharp and strange prediction. Let us work it out in the Newtonian regime, where gravity is weak and slow and the full equation (1.3) collapses to something we can solve in three lines: the Poisson equation $\nabla^2\Phi = 4\pi G\,\langle\hat{\rho}\rangle$ for the gravitational potential $\Phi$, sourced by the expectation value of the mass density.

**Worked Example 1.1 — The gravitational field of a superposed mass**

Put the left site at $\vec{x}_L = (-d/2, 0, 0)$ and the right site at $\vec{x}_R = (+d/2, 0, 0)$. In the branch $|L\rangle$ the mass density is that of a point mass at $\vec{x}_L$; in $|R\rangle$, at $\vec{x}_R$. For the equal superposition

(1.2) the expectation value is

$$\langle\hat\rho(\vec x)\rangle \;=\; \tfrac{1}{2}\,M\,\delta^3(\vec x-\vec x_L)+\tfrac{1}{2}\,M\,\delta^3(\vec x-\vec x_R). \tag{1.4}$$

(The cross terms $\langle L|\,\hat\rho\,|R\rangle$ vanish because the two locations are macroscopically distinct: the wavefunctions do not overlap.) The Poisson equation is linear, so the potential is the sum of two half-strength point potentials:

$$\Phi(\vec x) \;=\; -\frac{GM/2}{|\vec x-\vec x_L|} \;-\; \frac{GM/2}{|\vec x-\vec x_R|}. \tag{1.5}$$

Semiclassical gravity says: the grain in two places gravitates like *half a grain in each place.* Now put a test mass at the midpoint, $\vec x = 0$. By symmetry the two pulls in (1.5) cancel exactly: the test mass feels *zero force* and stays put. Contrast the "either–or" expectation: if the grain is really at $L$ (or really at $R$) the test mass accelerates toward the occupied site with magnitude $a = GM/(d/2)^2 = 4GM/d^2$, picking left or right with probability 1/2 each. The two predictions differ not in some fine decimal but categorically: *no motion* versus *motion, direction random.*

So the semiclassical equation is not vague about superpositions — it makes a definite prediction, the *mean-field* prediction. The test mass falls toward where the grain is on average, which is a place the grain never is. The prediction is wrong. We know it is wrong on two independent grounds, one experimental and one structural, and it is worth being precise about both, because the way it fails tells us what any repair must accomplish.

### 1.3.2 The experimental ground: Page and Geilker

In 1981 Page and Geilker performed the experiment — crude, but decisive for the question asked. A Cavendish torsion balance measured the gravitational pull of two heavy masses whose placement (one arrangement or a very different one) was decided by a quantum random process. If gravity were sourced by the expectation value of the full quantum state — which, absent any collapse of the wavefunction, retains *both* placements with equal weight — the torsion balance should have responded to the *average* of the two arrangements. It did not. It responded, every time, to the arrangement actually realised in the laboratory.

**Q 1.1** ***"Doesn't that just show the wavefunction collapsed when the random choice was made? Then the expectation value is the collapsed one, and the semiclassical equation survives."***

Correct — and that is precisely the fork in the road. Page and Geilker rule out the combination *semiclassical sourcing + no collapse*: if the state never collapses, its expectation value keeps both branches and gravity should have followed the mean, which it measurably did not. To save (1.3) you must invoke a genuine, physical collapse somewhere between the quantum die-roll and the torsion balance. But then the right-hand side of (1.3) *jumps* discontinuously at the moment of collapse, and we are about to see (next subsection) that the left-hand side is mathematically forbidden from jumping. Either way the semiclassical equation, taken as fundamental, is dead: without collapse it contradicts experiment; with collapse it contradicts itself. The equation survives only in the humbler role it plays to this day — an excellent *approximation* whenever matter fluctuations are negligible.

### 1.3.3 The structural ground: the left side cannot jump

The deeper failure is internal, and it will echo through this whole book, so let us state it carefully even though its full machinery arrives only in Chapter 3.

The geometric object $G_{\mu\nu}$ on the left of (1.3) is not an arbitrary package of derivatives. It obeys, as a mathematical *identity* — true for any geometry whatsoever, the way $\nabla \cdot (\nabla \times \vec{A}) = 0$ is true for any vector field — a conservation law called the contracted Bianchi identity:

$$\nabla^{\mu} G_{\mu\nu} \;=\; 0. \tag{1.6}$$

Consequently anything set equal to $G_{\mu\nu}$ must be conserved too: $\nabla^{\mu}\langle \hat{T}_{\mu\nu}\rangle = 0$, always, exactly. A conserved quantity cannot jump. But a wavefunction collapse is a jump: before the measurement the average mass distribution is (1.4), half left and half right; after it, all the mass sits on one side. Energy has teleported across the distance $d$ discontinuously, and (1.6) forbids the geometry from keeping up. The semiclassical coupling is therefore not merely inaccurate for superpositions — combined with collapse it is *inconsistent*, a theory whose two sides make contradictory demands.

[PROVEN] — this is not a subtle or contested point; it is nineteenth-century-style mathematics applied to twentieth-century equations, and we will derive (1.6) in full in Chapter 3.

**Q 1.2** ***"Maybe gravity is sourced branch by branch: in the $L$-branch the field points to the left mass, in the $R$-branch to the right. No mean field, no jump. Problem solved?"***

Notice what you have just done. "The gravitational field in the $L$-branch" and "the gravitational field in the $R$-branch" are two *different* geometries, and the world, on this proposal, is in a superposition of them. You have put spacetime itself into a quantum state — which is to concede the entire point at issue. The single classical metric of general relativity is gone, and the honest follow-up questions come flooding in: what is the state space of geometries? what are the observables? do superposed geometries interfere, and if so, what suppresses the interference for the Earth? This branch-wise picture is in fact where the theory of this book *starts* (Chapter 8 builds exactly the map $\sum_n c_n |\psi_n\rangle \mapsto \sum_n c_n |\psi_n\rangle \otimes |g_n\rangle$), and its central quantitative claim is the answer to the last question: the interference between macroscopically distinct geometries dies at a calculable rate, and the rate is the 1.6 nanoseconds on the cover.

## 1.4 Second exit: quantize the metric

If feeding gravity a classical average fails, the obvious alternative is to stop treating geometry classically at all: promote the metric to a quantum field, exactly as was done — with historic success — for electromagnetism. Write the geometry as flat spacetime plus a fluctuation,

$$g_{\mu\nu} \;=\; \eta_{\mu\nu} + \sqrt{32\pi G}\, h_{\mu\nu}, \tag{1.7}$$

and quantize $h_{\mu\nu}$: its quanta are *gravitons*, massless spin-2 cousins of the photon. At low orders this works and reproduces everything it should. The obstruction appears when one asks for corrections, and it can be exhibited with nothing more than dimensional analysis.

### 1.4.1 Power counting from scratch

Let us adopt, for this argument only, the particle physicist's units $\hbar = c = 1$, in which every quantity carries a dimension of mass to some power: energies are masses, lengths are inverse masses. (We restore $\hbar$ and $c$ the moment the argument ends; Worked Example 1.2 does it explicitly.) In these units Newton's constant has dimension

$$[\,G\,] \;=\; \text{mass}^{-2}, \qquad \text{write} \qquad G \;=\; \frac{1}{m_{\mathrm{P}}^2}, \tag{1.8}$$

which *defines* a mass scale $m_{\mathrm{P}}$, the Planck mass. You can check the dimension directly from Newton's law $F = GM_1M_2/r^2$: force is energy/length $=$ mass$^2$ in these units, the two masses contribute mass$^2$, $r^{-2}$ contributes mass$^2$, so $G$ must carry mass$^{-2}$.

Now the key question of quantum field theory: when we compute a process as a perturbation series (a sum over Feynman diagrams, organised by the number of closed loops), how do successive terms behave? Each extra order in perturbation theory brings one more factor of the coupling $G$, and each closed loop brings an integral over the unrestricted momentum $k$ running around it. Since $G$ has dimension mass$^{-2}$, and every term in the series for a given process must have the same total dimension, the extra $G$ must be compensated by *two extra powers of momentum* in the loop integral. Cutting the integrals off at a large momentum $\Lambda$ (a placeholder for "the highest energy where we still trust the theory"), the $L$-loop contribution scales as

$$(\text{amplitude at } L \text{ loops}) \;\sim\; (G\,\Lambda^2)^L \;=\; \left(\frac{\Lambda}{m_{\mathrm{P}}}\right)^{2L}. \tag{1.9}$$

Read (1.9) twice; it is the whole story. If $\Lambda \ll m_{\mathrm{P}}$, higher loops are ever *smaller* and the theory is predictive — gravity as an effective quantum field theory works fine at low energy. But as $\Lambda \to m_{\mathrm{P}}$ every order contributes equally, and beyond it the series explodes, each order worse than the last. Divergences of ever higher type appear at every order, and each new type of divergence must be cancelled by a new counterterm — a new free parameter that experiment alone can fix. Infinitely many orders, infinitely many parameters: a theory that can fit anything and predict nothing. This is what *non-renormalizable* means, and it is a statement about (1.8) — about the dimension of $G$ — not about anyone's lack of ingenuity.

**For the curious — Divergences, counterterms, and what "non-renormalizable" does and does not mean**

A loop integral like $\int^{\Lambda} d^4k\, k^{-2}$ grows as $\Lambda^2$: it *diverges* as the cutoff is removed. That by itself is not fatal — quantum electrodynamics diverges too. The cure, *renormalization*, absorbs each divergence into a redefinition of a parameter already present in the theory (a mass, a charge). QED needs only a finite list of such redefinitions, ever, at every order: it is renormalizable, because its coupling $e$ is dimensionless and the analogue of (1.9) does not grow with $L$. Gravity's coupling has negative mass dimension, so each order generates divergences of a genuinely *new* shape, needing counterterms that were not in Einstein's action — terms with more and more derivatives of the metric. The one-loop check was done by 't Hooft and Veltman (1974): for pure gravity the dangerous terms happen to vanish on shell. The two-loop computation (Goroff and Sagnotti, 1986) closed the loophole: a genuinely new divergence appears, with coefficient $\frac{209}{2880}(16\pi^2)^{-2}$ in front of a cube of

curvature tensors, and nothing in Einstein's theory can absorb it. [PROVEN]
None of this says quantum gravity is impossible. It says the *naive quantization of the metric* fails above $m_{\mathrm{P}}$, and whatever happens there — string theory, loop quantization, something nobody has thought of — involves physics beyond Einstein's action. The theory in this book takes no position on that ultraviolet question; its wager is that the *measurable* quantum-gravitational physics lies elsewhere entirely.

### 1.4.2 How big is the wall?

**Worked Example 1.2 — The Planck scales from $G$, $\hbar$, $c$**

Let us restore units and compute where the wall (1.9) stands. Out of the three constants

$$G = 6.67430\times10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}, \qquad \hbar = 1.054571817\times10^{-34}\,\mathrm{J\,s}, \qquad c = 299792458\,\mathrm{m/s},$$

exactly one combination has the dimension of each mechanical quantity. Demanding $[\,m_{\mathrm{P}}\,] = \mathrm{kg}$ with $m_{\mathrm{P}} = G^a\hbar^b c^d$ and matching powers of kg, m, s gives $a = -\frac{1}{2}$, $b = \frac{1}{2}$, $d = \frac{1}{2}$:

$$m_{\mathrm{P}} = \sqrt{\frac{\hbar c}{G}} = \sqrt{\frac{(1.054571817\times10^{-34})(2.99792458\times10^{8})}{6.67430\times10^{-11}}} = 2.176\times10^{-8}\ \mathrm{kg}, \tag{1.10}$$

$$\ell_{\mathrm{P}} = \sqrt{\frac{\hbar G}{c^3}} = 1.616\times10^{-35}\ \mathrm{m}, \qquad t_{\mathrm{P}} = \frac{\ell_{\mathrm{P}}}{c} = 5.391\times10^{-44}\ \mathrm{s}. \tag{1.11}$$

The Planck energy is $E_{\mathrm{P}} = m_{\mathrm{P}}c^2 = 1.956\times10^{9}$ J $\approx 1.22\times10^{19}$ GeV. For orientation: the Large Hadron Collider reaches about $10^4$ GeV. The wall stands fifteen orders of magnitude above the most powerful machine ever built; a collider probing $\ell_{\mathrm{P}}$ with present technology would have to be roughly the size of a galaxy.

▶ RUN IT: `TOOLS/verify-paper-numerics.py` (constants recomputed from $G, \hbar, c$; never hardcoded)

Note the strange position of the Planck *mass* in (1.10): unlike $\ell_{\mathrm{P}}$ and $t_{\mathrm{P}}$, which are absurdly small, $m_{\mathrm{P}} \approx 22$ micrograms is the mass of a small grain of sand — a thoroughly human-scale quantity, about $1.3\times10^{19}$ proton masses. Planck-scale physics is hopelessly remote in *energy per particle*, yet oddly nearby in *mass*. Hold that thought; it is not a coincidence, and the theory ahead lives on it.

## 1.5 The regime both theories forfeit

Let us take stock, because the two failures fit together with unusual precision.

> *Semiclassical gravity* fails only when matter fluctuations are large — a *heavy* mass in a *genuine superposition* of distinguishable configurations. For definite or nearly definite sources it is excellent.
> *Quantized gravity* (as an effective field theory) fails only at energies near $m_{\mathrm{P}}c^2$ per quantum. For any laboratory process it is predictive — but for a slow, heavy grain in superposition

> it predicts graviton-exchange effects so minute (we will compute them in Chapter 12: suppressed relative to the effect in this book by roughly $10^{34}$) that it effectively predicts *nothing observable*.

The territory that neither theory covers is therefore not some exotic Planckian fireball. It is a bench-top situation: *a mass large enough to gravitate measurably, superposed across a distance large enough to matter, at ordinary energies*. No curvature singularities, no colliders the size of galaxies — a grain, a beam splitter, and patience. That is the problem nobody ordered: the twentieth century's two triumphant rulebooks, each valid to fantastic precision on its own turf, jointly fall silent on a question you can ask about laboratory apparatus.

**Q 1.3** ***"Why did nobody just do the experiment, then?"***

Because "a mass large enough and a superposition wide enough" is brutally hard to arrange. Superpositions of heavy objects are destroyed by *ordinary* decoherence — stray gas molecules, blackbody photons, vibrations — at rates that grow ferociously with size. The state of the art (as of the mid-2020s) is interference of $\sim 4 \times 10^{-23}$ kg molecules and ground-state control of $\sim 10^{-17}$ kg mechanical oscillators, while the benchmark grain of this book weighs $10^{-9}$ kg. The gap is closing on a logarithmic march — levitated nanospheres, matter-wave interferometers of increasing grasp — and Chapter 15 is devoted to the crossing point: the mass–separation–time window where the theory ahead predicts an effect that *cannot be refrigerated away*, with a five-sigma protocol an experimentalist could run. The honest timeline is years, not decades: roughly 2028–2035.

## 1.6 The wager of this book

Every response to the impasse chooses something to give up. String theory and loop quantum gravity give up the simplicity of the field content or the smoothness of spacetime, and buy consistency at the Planck scale. Collapse models (we will meet them properly in Chapters 12 and 15) give up the exactness of quantum mechanics itself, adding a stochastic term that kills superpositions by hand. The theory of this book — the *Quantum–Geometric Correspondence* (QGC) — gives up something different and, at first hearing, more modest: the assumption that quantum mechanics and general relativity are two theories in need of merging.

Its stance: there is one underlying structure — one algebra of observables carrying one state, objects we will build with complete care in Chapters 6–7 — and "quantum mechanics" and "general relativity" are what that single structure looks like under two *complementary restrictions*, the way a cylinder casts a circular shadow on one wall and a rectangular one on the other. Neither shadow is wrong; neither is the object. On this reading the impasse of §1.3–1.4 is a category error — like asking which shadow is the real cylinder — and the productive question becomes: *what happens at the seam*, where a system is heavy enough for its geometric shadow to matter and quantum enough for its matter shadow to superpose? (The cylinder is an analogy and will be retired on contact with mathematics: Chapter 8 states the same sentence as a formal dictionary between restrictions of one algebra.)

The theory's answer at the seam is quantitative. When a mass $M$ sits in a superposition of two locations separated by $d$, the two branches carry measurably different geometries, and the correspondence forces the quantum coherence between the branches to decay — not because of gas molecules or photons, but as a matter of principle — at the timescale

$$\tau_{\text{dec}} = \frac{\hbar\, d}{G M^2}, \tag{1.12}$$

set by the same three constants as Worked Example 1.2. Notice its shape before trusting it: coherence dies *faster* for heavier masses (as $M^{-2}$) and *slower* for wider splits (as $d$ — wider superpositions live longer, which will surprise you until Chapter 11 explains it). Both features are fingerprints, distinguishable in the laboratory from every mundane decoherence source, all of which grow with $d$.

**Worked Example 1.3 — The benchmark: one microgram, one millimetre**

Insert $M = 1\,\mu\text{g} = 10^{-9}$ kg and $d = 1$ mm $= 10^{-3}$ m into (1.12), using the constants of Worked Example 1.2:

$$\tau_{\text{dec}} = \frac{(1.054571817 \times 10^{-34})\,(10^{-3})}{(6.67430 \times 10^{-11})\,(10^{-9})^2} = \frac{1.054571817 \times 10^{-37}}{6.67430 \times 10^{-29}} = 1.58 \times 10^{-9}\text{ s} \;\approx\; 1.6\text{ ns}. \tag{1.13}$$

The energy scale behind it — the gravitational self-energy that distinguishes the two branches, a quantity we will derive rather than posit in Chapter 11 — is

$$E_{\text{G}} = \frac{GM^2}{d} = \frac{(6.67430 \times 10^{-11})(10^{-18})}{10^{-3}} = 6.7 \times 10^{-26}\text{ J}, \qquad \tau_{\text{dec}} = \frac{\hbar}{E_{\text{G}}}. \tag{1.14}$$

A microgram grain, coherently split by a millimetre, keeps its quantum character for about a nanosecond and a half — no matter how perfect the vacuum, no matter how cold the apparatus.

▶ Run it: `TOOLS/verify-paper-numerics.py` (`cross-paper benchmark X:` $\tau_{\text{dec}}$, $E_{\text{G}}$)

**Q 1.4 *"You just wrote down the headline formula with no derivation. Why should I believe a single symbol of it?"***

You should not — yet. Equation (1.12) is, at this point in the book, a *promissory note*, and the honest status of its ingredients differs and must be kept straight. That the energy scale $E_{\text{G}} = GM^2/d$ of (1.14) governs the seam is [PROVEN] — it follows from classical gravity by a computation you will verify yourself. That the decoherence *rate* equals $C \cdot E_{\text{G}}/\hbar$ with $C$ of order one rests on an identity between two kinds of time — modular time and physical time, $K = 2\pi H_{\text{phys}}$, the very heart of the theory — which is [CONJECTURED]: proven in expectation values and in every solvable model, open as a full operator equation, with the residue isolated in one sharply-posed construction (Chapter 13). The coefficient $C$ itself is pinned to the window $[2/\pi, 1]$. Chapters 11–13 exist to turn the note into whatever it can honestly be turned into, in front of you, with no skipped steps — and Chapter 15 describes the experiments that will settle what mathematics so far has not. If a levitated-nanosphere interferometer holds coherence for a hundred times $\tau_{\text{dec}}$, the theory in this book is dead. That is a feature. Few approaches to quantum gravity offer their own executioner a date and an address.

## 1.7 The road ahead

The strategy of the volume follows from the diagnosis. If QM and GR are two restrictions of one structure, we must first own both restrictions — not at the level of slogans but at the level where calculations can be checked.

**Part II** builds the toolkit from the floor up. Chapter 2 recasts quantum mechanics in the language the seam demands — density matrices, entanglement, entropy — where *decoherence* gets its precise meaning as the decay of interference terms. Chapter 3 builds general relativity operationally, from the equivalence principle to Einstein's equations, with special care for *horizons* — Rindler, Schwarzschild, de Sitter — and for the Raychaudhuri equation, the workhorse of Part III. Chapter 4 puts quantum fields on spacetime and computes the Unruh effect in full: an accelerated observer finds the vacuum at temperature $T = \hbar a/2\pi c k_\mathrm{B}$, the first exact hint that geometry and quantum statistics are entangled at the root. Chapter 5 collides the toolkits: black-hole entropy, the Bekenstein bound, and the generalized entropy $S_\mathrm{gen} = A/4\ell_\mathrm{P}^2 + S_\mathrm{ext}$ that becomes Axiom I. Chapters 6 and 7 then build the mathematics the correspondence is actually written in — algebras of observables and Tomita–Takesaki modular theory — entirely from finite-dimensional examples, up to the crossed product that gives the observer a clock.

**Part III** is the theory itself: the correspondence stated exactly (Chapter 8), the three axioms with their independence countermodels (Chapter 9), Einstein's equations *derived* from entanglement equilibrium (Chapter 10), and then the long walk to the rate: the Wheeler–DeWitt constraint and the influence functional (Chapter 11), the rate itself with its honest $G^1$-versus-$G^2$ ledger (Chapter 12), and the one open construction R1 (Chapter 13).

**Part IV** leaves the bench. Chapter 14 applies the same machinery to the cosmological horizon and finds, with no parameters to turn, the acceleration scale $a_0 = cH_0/2\pi$ that galactic rotation curves have been whispering for forty years. Chapter 15 ends where physics must: at the experiments, with the five-sigma protocol and the list of results that would kill the theory outright.

## Chapter FAQ

**Q: Is the conflict of §1.2 not just the measurement problem in disguise?** They are adjacent but distinct. The measurement problem asks why experiments have single outcomes at all; it exists with or without gravity. The conflict here is sharper and more technical: two specific, spectacularly confirmed dynamical equations make contradictory demands on one observable (the gravitational field of a superposition). You could solve the measurement problem tomorrow and this conflict would remain; conversely, the theory in this book takes no position on interpretation — the Born rule enters as [ASSUMED], exactly as in ordinary quantum mechanics (Chapter 2 makes this precise).

**Q: Gravity between laboratory masses is fantastically weak. How can a gravitational effect on coherence be as fast as nanoseconds?** Because (1.12) does not compete with gravitational *forces*; it is a dephasing clock. The relevant quantity is the energy $E_\mathrm{G}$ by which the two branches' gravitational configurations differ, and a quantum phase winds at the rate $E_\mathrm{G}/\hbar$ — and $\hbar$ is small, so even $10^{-26}$ J winds phases in nanoseconds. Weak energy, fast clock. The full mechanism — including why this is a genuine loss of coherence and not a removable overall phase — is the business of Chapter 11.

**Q: Doesn't "heavier ⇒ faster decoherence" just say big things are classical because they are big — common sense, not new physics?** The qualitative direction is common sense; the *law* is not. Ordinary environmental decoherence also kills big superpositions, but with rates that depend on temperature, pressure, and geometry, and that can in principle be engineered away — better vacuum, colder walls. Equation (1.12) is an *uncoolable floor*: it depends only on $M$, $d$, and constants of nature. The experimental programme of Chapter 15 is precisely the hunt for that floor beneath the removable noise.

**Q: Which existing approach is this closest to — and is the formula (1.12) even new?** The timescale $\hbar d/GM^2$ itself has a distinguished history: it is the Diósi–Penrose scale, proposed on heuristic grounds in the 1980s–90s as the lifetime of superposed geometries. What the present framework claims as its own is not the number but the *derivation and its consequences*: the scale emerges from an axiomatized correspondence rather than being postulated, quantum mechanics stays exactly linear (no stochastic collapse terms added by hand), the coefficient is pinned to a window with sharp experimental discriminators against collapse models, and the same three axioms yield, with no further input, the cosmological results of Chapter 14. Where the predictions of QGC and collapse models diverge — and they measurably do — Chapter 15 takes sides.

## Exercises

**Exercise 1.1.** An electron ($M = 9.109 \times 10^{-31}$ kg) is placed in a superposition of two paths separated by $d = 1\,\mu$m — routine in electron interferometry. Compute $\tau_{\text{dec}}$ from (1.12). Compare with the age of the universe ($4.35 \times 10^{17}$ s). What does the answer say about why quantum mechanics was discovered on light particles?

**Exercise 1.2.** (a) A future levitated-nanosphere experiment can hold a coherent superposition with $d = 1\,\mu$m for about one second before ordinary decoherence intervenes. What mass $M$ puts the gravitational timescale (1.12) exactly at $\tau_{\text{dec}} = 1$ s? (b) Evaluate $\tau_{\text{dec}}$ for a grain of $M = 10^{-15}$ kg at the same separation. (c) Express the answer to (a) in units of the Planck mass (1.10) and comment on the "oddly nearby" remark of §1.4.1.

## Solutions

*Solution (Exercise 1.1).* With $M^2 = 8.298 \times 10^{-61}\,\text{kg}^2$,

$$\tau_{\text{dec}} = \frac{\hbar d}{GM^2} = \frac{(1.054571817 \times 10^{-34})(10^{-6})}{(6.67430 \times 10^{-11})(8.298 \times 10^{-61})} = \frac{1.0546 \times 10^{-40}}{5.539 \times 10^{-71}} = 1.90 \times 10^{30}\ \text{s},$$

about $6 \times 10^{22}$ years — some $4 \times 10^{12}$ times the age of the universe. Gravitational decoherence is utterly irrelevant for electrons: their interference patterns are safe forever, which is precisely why quantum mechanics could be discovered, and confirmed to exquisite precision, entirely on light particles. The $M^{-2}$ in (1.12) hides the seam from every experiment ever done — and uncovers it abruptly as $M$ climbs toward the nanogram range.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: lecture-notes checks)`

*Solution (Exercise 1.2).* (a) Set $\tau_{\text{dec}} = \hbar d/GM^2 = 1$ s and solve for $M$:

$$M = \sqrt{\frac{\hbar d}{G\,\tau_{\text{dec}}}} = \sqrt{\frac{(1.054571817 \times 10^{-34})(10^{-6})}{(6.67430 \times 10^{-11})(1)}} = \sqrt{1.580 \times 10^{-30}} = 1.26 \times 10^{-15}\ \text{kg}.$$

About a femtogram — a nanoparticle of roughly 50 nm radius at ordinary solid densities: exactly the class of object levitated interferometry is learning to control, which is why Chapter 15's five-sigma protocol lives at this mass. (b) For $M = 10^{-15}$ kg, $GM^2 = 6.674 \times 10^{-41}$ J m and $\tau_{\text{dec}} = 1.0546 \times 10^{-40}/6.674 \times 10^{-41} = 1.58$ s: the gravitational clock and the practical coherence time collide — the experiment becomes decisive rather than hopeless. (c) $M/m_{\text{P}} = 1.26 \times 10^{-15}/2.176 \times 10^{-8} \approx 5.8 \times 10^{-8}$. The decisive experiment sits eight orders of magnitude *below* the Planck mass — and fifteen orders above the electron. The seam is not at the Planck scale; it is in the middle of the mass ladder, within engineering reach.

▶ RUN IT: `TOOLS/verify-paper-numerics.py (section N: lecture-notes checks)`

**What to carry forward**

1. Semiclassical gravity ($G_{\mu\nu} = 8\pi G\langle\hat{T}_{\mu\nu}\rangle/c^4$) is internally inconsistent on superpositions: without collapse it contradicts Page–Geilker; with collapse it contradicts the Bianchi identity $\nabla^\mu G_{\mu\nu} = 0$. It survives only as a mean-field approximation.

2. Quantizing the metric fails by power counting: $[G] = \text{mass}^{-2}$ forces $L$-loop amplitudes to grow as $(\Lambda/m_{\text{P}})^{2L}$ — predictivity dies at the Planck scale (Goroff–Sagnotti). Below it, graviton effects on a superposed grain are unobservably small.

3. The no-man's-land is bench-top, not Planckian: a heavy mass in a wide superposition. There QGC issues its promissory note $\tau_{\text{dec}} = \hbar d/GM^2 = \hbar/E_{\text{G}}$ — 1.6 ns at one microgram and one millimetre — with the energy scale [PROVEN] and the rate [CONJECTURED], to be earned step by step in Part III and executed or exonerated in the laboratory.

# Part II

# The Toolkit

# 2 Quantum Mechanics as Information

**Synopsis.** Undergraduate quantum mechanics runs on state vectors $|\psi\rangle$. The physics of this book runs on a more honest object, the density matrix, which describes what an observer with access to *part* of the world can actually know. We construct it operationally, prove the two tools that make it tractable (partial trace, Schmidt decomposition), and then watch quantum coherence die in a fully solvable model: *decoherence* is the decay of off-diagonal matrix elements, and its magnitude is an overlap of environment states — a formula we will meet again, verbatim, when the environment is spacetime itself. Entropy then quantifies what was lost, and one inequality (Klein's) is proved in full because half of Part III leans on it.

**You will be able to.** compute reduced density matrices and Schmidt decompositions by hand; derive the visibility law $\mathcal{V}(t) = |\langle E_L(t)|E_R(t)\rangle|$ and the exponential decay it produces for many-particle environments; compute von Neumann and relative entropies and prove $S(\rho\|\sigma) \geq 0$.

**Prerequisites.** Chapter 1; bra-ket algebra, matrix diagonalization, Taylor expansion. Everything else is built here.

## 2.1 Why state vectors are not enough

Quantum mechanics, as first taught, assigns to every system a state vector $|\psi\rangle$ and to every observable a Hermitian operator $A$, with statistics $\langle A\rangle = \langle\psi| A |\psi\rangle$. For the problems of this book that description fails — not because it is wrong, but because it answers a question we are rarely entitled to ask. It presumes the describer has the *whole* system in view. The observer at the seam of quantum mechanics and gravity never does: she sees the grain but not every graviton, the laboratory but not the cosmological horizon. We need the mechanics of *partial* knowledge, and we should build it operationally: what is the most general assignment of statistics to measurements that quantum mechanics permits?

Two distinct situations force the same generalisation.

*Situation one: classical ignorance.* A machine emits a qubit in state $|\psi_1\rangle$ with probability $p_1$, or $|\psi_2\rangle$ with probability $p_2$. Anyone measuring an observable $A$ on the output sees the mixture of statistics

$$\langle A\rangle \;=\; p_1 \langle\psi_1| A |\psi_1\rangle + p_2 \langle\psi_2| A |\psi_2\rangle\,. \tag{2.1}$$

No single vector $|\psi\rangle$ reproduces (2.1) for all $A$: a superposition $c_1 |\psi_1\rangle + c_2 |\psi_2\rangle$ produces *interference* cross terms $c_1^* c_2 \langle\psi_1| A |\psi_2\rangle + \ldots$ that the ensemble does not have.

*Situation two: entanglement.* The system is one half of a pair prepared in a joint state; the observer measures only her half. We will compute in §2.2 what she sees, and it will again fail to be $\langle\psi| A |\psi\rangle$ for any vector $|\psi\rangle$ of her half alone.

Both situations are covered by one object. Rewrite (2.1) using the trace — recall $\mathrm{Tr}(X) = \sum_n \langle n| X |n\rangle$ over any orthonormal basis, and its cyclic property $\mathrm{Tr}(XY) = \mathrm{Tr}(YX)$:

$$\langle\psi_i| A |\psi_i\rangle = \mathrm{Tr}(|\psi_i\rangle\langle\psi_i|\, A) \qquad \Longrightarrow \qquad \langle A\rangle = \mathrm{Tr}(\rho A), \quad \rho \,\equiv\, \sum_i p_i \,|\psi_i\rangle\langle\psi_i|\,. \tag{2.2}$$

The operator $\rho$ is the *density matrix* (or density operator, or simply the *state*), and (2.2) says: everything any measurement can reveal about the system is stored in $\rho$. Its defining properties follow directly from the construction:

**Definition 2.1** (Density matrix)**.** A density matrix on a Hilbert space $\mathcal{H}$ is an operator $\rho$ that is (i) Hermitian, $\rho^\dagger = \rho$; (ii) positive, $\langle\phi|\,\rho\,|\phi\rangle \geq 0$ for all $|\phi\rangle$; (iii) normalised, $\mathrm{Tr}\,\rho = 1$. A state is *pure* if $\rho = |\psi\rangle\langle\psi|$ for some vector (equivalently $\rho^2 = \rho$), and *mixed* otherwise. The *purity* $\mathrm{Tr}\,\rho^2 \leq 1$ measures which, with equality exactly for pure states.

Each property is inherited from (2.2): hermiticity and positivity because each $|\psi_i\rangle\langle\psi_i|$ is a projector and the $p_i$ are non-negative; normalisation because the $p_i$ sum to one. The physical reading of positivity is worth one line: it says every probability the state assigns — $\mathrm{Tr}(\rho P)$ for a projector $P$ — comes out non-negative, which is the least a probability can do.

### Worked Example 2.1 — Every qubit state is a point in a ball

For a single qubit, the Hermitian $2 \times 2$ matrices are spanned by the identity and the three Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

each traceless with $\sigma_j^2 = \mathbb{1}$. Any density matrix can therefore be written with a real vector $\vec{r}$ as

$$\rho \;=\; \tfrac{1}{2}(\mathbb{1} + \vec{r}\cdot\vec{\sigma}), \qquad r_j = \mathrm{Tr}(\rho\,\sigma_j) = \langle\sigma_j\rangle. \tag{2.3}$$

The eigenvalues of $\vec{r}\cdot\vec{\sigma}$ are $\pm|\vec{r}|$ (square it: $(\vec{r}\cdot\vec{\sigma})^2 = |\vec{r}|^2\mathbb{1}$), so the eigenvalues of $\rho$ are $(1 \pm |\vec{r}|)/2$. Positivity demands $|\vec{r}| \leq 1$: the state space is a solid ball, the *Bloch ball.* Purity is $\mathrm{Tr}\rho^2 = (1 + |\vec{r}|^2)/2$: pure states live on the surface $|\vec{r}| = 1$, the maximally mixed state $\rho = \mathbb{1}/2$ ($\vec{r} = 0$, eigenvalues $\frac{1}{2}, \frac{1}{2}$) at the centre. A state with, say, $|\vec{r}| = 0.8$ has eigenvalues 0.9 and 0.1 — we will compute its entropy in §2.5.

**Q 2.1** ***"Is a mixed state just a fancy bookkeeping for ignorance? If I knew which $|\psi_i\rangle$ the machine emitted, there would be no $\rho$."***

For situation one, yes — and then the decomposition $\{p_i, |\psi_i\rangle\}$ reflects your ignorance and is not unique (the maximally mixed qubit is an equal mixture of $|0\rangle, |1\rangle$, but equally of $|+\rangle, |-\rangle$; same $\rho$, same physics). But situation two is the profound one: when your qubit is half of an entangled pair, its reduced state is mixed *and there is no fact of the matter you are ignorant of* — the pair as a whole is in a single, definite pure state. No one, anywhere, holds hidden information whose disclosure would purify your qubit. Mixedness from entanglement is objective. And here is the consequence to hold onto for later: since the ensemble reading and the entanglement reading produce the *same* $\rho$, no local measurement can distinguish them. That indistinguishability is what protects relativity from entanglement — if your distant partner's choice of measurement changed your local $\rho$, you could receive her message instantaneously. It also foreshadows a theme of Chapter 7: what an observer can know is determined by the *algebra of observables she can reach*, not by the global state.

## 2.2 Composite systems and the partial trace

### 2.2.1 Tensor products in five minutes

Two systems, $A$ with basis $\{|i\rangle_A\}$ $(i = 1, \ldots, d_A)$ and $B$ with basis $\{|a\rangle_B\}$, are described jointly on the *tensor product* space $\mathcal{H}_A \otimes \mathcal{H}_B$: the $d_A d_B$-dimensional space spanned by the pairs $|i\rangle_A \otimes |a\rangle_B$ (written $|i, a\rangle$ or $|i\rangle\,|a\rangle$), with inner product $\langle i, a|j, b\rangle = \delta_{ij}\delta_{ab}$ and operators acting factor-wise, $(X \otimes Y)\,|i, a\rangle = (X\,|i\rangle) \otimes (Y\,|a\rangle)$. The general joint state is

$$|\Psi\rangle = \sum_{i,a} c_{ia}\,|i\rangle_A\,|a\rangle_B\,, \tag{2.4}$$

and here quantum mechanics springs its central trap: generically (2.4) is *not* a product $|\alpha\rangle_A\,|\beta\rangle_B$ — the coefficient matrix $c_{ia}$ generically has rank $> 1$. Such states are called *entangled*, and the whole of this book happens inside that word.

### 2.2.2 What the local observer sees

Suppose the pair is in the state (2.4) but our observer can measure only system $A$: her observables are $A \otimes \mathbb{1}_B$. Her statistics are $\langle A\rangle = \langle\Psi|\,A \otimes \mathbb{1}\,|\Psi\rangle$, and we now ask: is there a density matrix $\rho_A$ *of system* $A$ *alone* such that $\langle A\rangle = \mathrm{Tr}_A(\rho_A A)$ for every $A$? Insert (2.4) and compute directly:

$$\langle\Psi|\,A \otimes \mathbb{1}\,|\Psi\rangle = \sum_{i,a}\sum_{j,b} c^*_{ia}c_{jb}\,\langle i|\,A\,|j\rangle\,\underbrace{\langle a|b\rangle}_{\delta_{ab}} = \sum_{i,j}\Big(\sum_a c^*_{ia}c_{ja}\Big)\,\langle i|\,A\,|j\rangle\,. \tag{2.5}$$

The bracketed matrix is the answer: defining $(\rho_A)_{ji} \equiv \sum_a c_{ja}c^*_{ia}$, equation (2.5) reads $\langle A\rangle = \mathrm{Tr}(\rho_A A)$. This $\rho_A$ is called the *reduced density matrix*, the operation that produces it the *partial trace*:

$$\rho_A \;=\; \mathrm{Tr}_B\,|\Psi\rangle\langle\Psi| \;\equiv\; \sum_a {}_B\langle a|\;\Psi\rangle\langle\Psi\;|a\rangle_B\,. \tag{2.6}$$

Two remarks, both load-bearing. First, $\rho_A$ is not an approximation or a convenience: it is the *unique* operator reproducing every local measurement, so it is the complete physical truth available at $A$. Second, the derivation used nothing about $B$ except an orthonormal basis — $B$ may be a single spin or the rest of the universe.

**Worked Example 2.2 — Tracing out half of a Bell pair**

Take two qubits in the Bell state

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A\,|0\rangle_B + |1\rangle_A\,|1\rangle_B),$$

i.e. $c_{00} = c_{11} = 1/\sqrt{2}$, $c_{01} = c_{10} = 0$. Apply (2.6) term by term. Writing $|\Phi^+\rangle\langle\Phi^+|$ out gives four terms; sandwiching with ${}_B\langle 0|\cdots|0\rangle_B$ keeps only those with $|0\rangle_B$ on both sides, and similarly for $|1\rangle_B$:

$$ {}_B\langle 0|\,\Phi^+\rangle\langle\Phi^+\,|0\rangle_B = \tfrac{1}{2}\,|0\rangle_A\langle 0|_A\,, \qquad {}_B\langle 1|\,\Phi^+\rangle\langle\Phi^+\,|1\rangle_B = \tfrac{1}{2}\,|1\rangle_A\langle 1|_A\,, $$
$$ \rho_A = \tfrac{1}{2}\,|0\rangle\langle 0| + \tfrac{1}{2}\,|1\rangle\langle 1| = \tfrac{1}{2}\,\mathbb{1}. \tag{2.7} $$

The cross terms $|0\rangle_A\,\langle 1|_A$ carried $\langle 1|0\rangle_B = 0$ and died. Physically: the half of a maximally entangled pair is, on its own, the *most random state a qubit can be in* — the centre of the Bloch ball of Worked Example 2.1 — even though the pair jointly is in a single pure state known exactly. Perfect global knowledge, zero local knowledge: that inversion of classical intuition is entanglement's signature, and §2.5 will give it a number ($\ln 2$).

**Q 2.2** ***"Where did the information go? The pair is pure — someone must still "have" the missing half of the story."***

Nobody has it locally. The missing information sits in the *correlations*: joint measurements (compare notes on both qubits) find perfect agreement in the $z$-basis *and* in the $x$-basis, and no local record could produce both. You can check the second claim in one line: rewriting $|\Phi^+\rangle$ in the $\pm$ basis ($|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$) gives $|\Phi^+\rangle = (|+\rangle\,|+\rangle + |-\rangle\,|-\rangle)/\sqrt{2}$ — the same correlated form. Information stored in no place but in the relations between places: this is not a metaphor but bookkeeping, and the bookkeeping device — how much of a pure state's definiteness is locally invisible — is the entanglement entropy of §2.5. When, in Chapter 5, an *area* shows up as the answer to "how much can be hidden behind a horizon," it is exactly this ledger being balanced.

## 2.3 Schmidt decomposition and purification

The coefficient matrix $c_{ia}$ of (2.4) admits a normal form that makes every entanglement question diagonal. The tool is the singular value decomposition (SVD) of linear algebra: any complex matrix $c$ can be written $c = U\Sigma V^\dagger$ with $U, V$ unitary and $\Sigma$ diagonal, real, non-negative.

**Theorem 2.2** (Schmidt decomposition). *Every pure state of a bipartite system can be written*

$$ |\Psi\rangle \;=\; \sum_{k=1}^{r} \sqrt{\lambda_k}\;|u_k\rangle_A\,|v_k\rangle_B\,, \qquad \lambda_k > 0, \quad \sum_k \lambda_k = 1, \tag{2.8} $$

*with* $\{|u_k\rangle\}$ *orthonormal in* $\mathcal{H}_A$ *and* $\{|v_k\rangle\}$ *orthonormal in* $\mathcal{H}_B$*. The number* $r$ *(the Schmidt rank) and the coefficients* $\lambda_k$ *are unique.*

*Derivation.* Insert the SVD $c_{ia} = \sum_k U_{ik}\,\sigma_k\,V^*_{ak}$ into (2.4) and regroup:

$$ |\Psi\rangle = \sum_k \sigma_k \Big(\sum_i U_{ik}\,|i\rangle_A\Big)\Big(\sum_a V^*_{ak}\,|a\rangle_B\Big) \equiv \sum_k \sigma_k\,|u_k\rangle_A\,|v_k\rangle_B\,. $$

The vectors $|u_k\rangle$ are orthonormal because $U$ is unitary ($\langle u_k|u_l\rangle = \sum_i U^*_{ik}U_{il} = \delta_{kl}$), likewise $|v_k\rangle$; normalisation of $|\Psi\rangle$ forces $\sum_k \sigma_k^2 = 1$, so set $\lambda_k = \sigma_k^2$. Uniqueness follows from the next display: the $\lambda_k$ are the eigenvalues of a reduced density matrix, which is basis-independent.

Tracing (2.8) with (2.6) — the cross terms die on $\langle v_k|v_l\rangle = \delta_{kl}$ exactly as in Worked

Example 2.2 — gives the payoff:

$$\rho_A = \sum_k \lambda_k \left|u_k\right\rangle\!\left\langle u_k\right| , \qquad \rho_B = \sum_k \lambda_k \left|v_k\right\rangle\!\left\langle v_k\right| . \tag{2.9}$$

Both reduced states have the *same* spectrum $\{\lambda_k\}$ — remarkable, since $A$ and $B$ may be nothing alike (one atom versus a cloud of $10^{23}$). Everything either party can locally know is fixed by one probability distribution, and the state is entangled precisely when $r > 1$.

Read (2.9) backwards and you get *purification*: given any mixed state $\rho_A = \sum_k \lambda_k \left|u_k\right\rangle\!\left\langle u_k\right|$ of $A$, pick any auxiliary system $B$ with dimension $\geq r$ and define $|\Psi\rangle$ by (2.8). Then $\rho_A$ is the reduced state of the *pure* $|\Psi\rangle$. Every mixed state, without exception, can be viewed as half of a pure entangled state. Mixedness is never fundamental; it is always, possibly fictitiously, traced-out entanglement. This innocent-looking remark returns in Chapter 6 wearing formal dress (the GNS construction) and in Chapter 7 it becomes the engine of the whole theory: the purifying partner of a thermal state is what modular theory acts on.

## 2.4 Decoherence: how interference dies

We now have the tools to watch the central phenomenon of this book happen in slow motion, in a model simple enough to solve exactly and rich enough to contain the full mechanism.

### 2.4.1 Which-path information kills fringes

Return to the superposed grain of Chapter 1, $|\psi\rangle = (|L\rangle + |R\rangle)/\sqrt{2}$, and let it interact with *one* environment particle — a gas molecule, a photon — that scatters differently off the two branches. Before the interaction the joint state is a product; the interaction correlates them:

$$\frac{|L\rangle + |R\rangle}{\sqrt{2}} \otimes |E_0\rangle \;\longrightarrow\; \frac{1}{\sqrt{2}}\Big(|L\rangle \otimes |E_L\rangle + |R\rangle \otimes |E_R\rangle\Big), \tag{2.10}$$

where $|E_L\rangle$, $|E_R\rangle$ are the (normalised) states the environment ends in, given the branch. Nothing non-unitary happened — (2.10) is ordinary Schrödinger evolution — and yet the grain's local physics has changed. Compute its reduced density matrix by the partial trace (2.6). In the $\{|L\rangle, |R\rangle\}$ basis:

$$\rho_{\text{grain}} = \frac{1}{2}\begin{pmatrix} 1 & \langle E_R|E_L\rangle \\ \langle E_L|E_R\rangle & 1 \end{pmatrix}. \tag{2.11}$$

(The diagonal entries carry $\langle E_L|E_L\rangle = \langle E_R|E_R\rangle = 1$; the off-diagonal ones inherit the environment overlap, exactly as the cross terms in Worked Example 2.2 inherited $\langle 1|0\rangle = 0$.)

Every interference experiment on the grain measures, in the end, some observable with matrix elements between $|L\rangle$ and $|R\rangle$, and (2.11) says all such matrix elements are multiplied by one complex number. Its modulus is the *visibility* — the contrast of the fringes the grain can still produce:

$$\mathcal{V} \;=\; |\langle E_L|E_R\rangle|. \tag{2.12}$$

In words: *coherence survives exactly to the extent that the environment cannot tell the branches apart.* If the two scattered states are identical ($\langle E_L|E_R\rangle = 1$), full fringes; if they are orthogonal

($\langle E_L|E_R\rangle = 0$) — if the environment has recorded complete which-path information — the fringes are gone and (2.11) is an even classical mixture. This is *decoherence*: not a new dynamical law, not a collapse, but the arithmetic of entanglement viewed from one side.

Equation (2.12) is, at bottom, the only formula this book has. In Chapter 11 the "environment" will be the gravitational field the grain itself sources, $|E_{L,R}\rangle$ will be two coherent states of that field differing by where the mass sits, and the program of Part III is nothing but the honest computation of their overlap $\langle E_L|E_R\rangle$ — with the twist that for gravity, unlike for gas molecules, a constraint *forces* the record to exist.

### 2.4.2 Many environment particles: the exponential law

Real environments are not one molecule. Let $N$ of them scatter independently, each acquiring its own small piece of which-path information:

$$|E_L\rangle = \bigotimes_{k=1}^{N} |e_L^{(k)}\rangle\,, \qquad |E_R\rangle = \bigotimes_{k=1}^{N} |e_R^{(k)}\rangle \qquad \Longrightarrow \qquad \mathcal{V} = \prod_{k=1}^{N} |\langle e_L^{(k)}|e_R^{(k)}\rangle|. \tag{2.13}$$

The overlap of product states factorises — check it on two factors: $(\langle e_L^{(1)}|\otimes\langle e_L^{(2)}|)(|e_R^{(1)}\rangle\otimes|e_R^{(2)}\rangle) = \langle e_L^{(1)}|e_R^{(1)}\rangle\,\langle e_L^{(2)}|e_R^{(2)}\rangle$. Now suppose each collision is almost uninformative, $|\,\langle e_L^{(k)}|e_R^{(k)}\rangle\,| = 1-\epsilon$ with $\epsilon \ll 1$. Then

$$\mathcal{V} = (1-\epsilon)^N = e^{N\ln(1-\epsilon)} \;\approx\; e^{-N\epsilon}, \tag{2.14}$$

using $\ln(1-\epsilon) \approx -\epsilon$. Collisions arrive at some rate $\gamma$, so $N = \gamma t$ and $\mathcal{V}(t) = e^{-t/\tau_{\text{env}}}$ with $\tau_{\text{env}} = 1/(\gamma\epsilon)$: visibility decays *exponentially*, with a timescale set by how often the environment looks ($\gamma$) times how much it learns per look ($\epsilon$). No single collision does anything dramatic — each overlap is 0.999 — but which-path information is cumulative and unforgiving: at $\epsilon = 10^{-3}$, a thousand collisions reduce the fringes to $1/e$, and a microgram grain in room air suffers that many collisions in about $10^{-17}$ seconds. This is why the quantum-to-classical transition *looks* like a law of nature: for macroscopic bodies in ordinary environments it is faster than anything else in physics.

**Q 2.3 *"Nothing was destroyed — the evolution (2.10) is unitary. Couldn't I undo it and get my fringes back?"***

In principle, yes, and the principle matters. The global state remains pure; the coherence has not vanished but *relocated* into grain–environment correlations, and a sufficiently godlike experimenter who reflected every scattered molecule back to undo (2.10) would revive the fringes (for one or two photons, such "quantum eraser" experiments are real). In practice the records fly away at the speed of sound or light into $10^{23}$ degrees of freedom, and no laboratory operation confined to the grain can ever recontract them: for all local purposes the coherence is gone, which is what (2.11) states with no philosophy attached. Now mark the question this raises for Part III, because it is sharper than it looks: ordinary decoherence is *contingent* — better vacuum, fewer records. Is gravitational which-path information contingent too? The grain's mass sources a field that permeates all space; can the record it leaves be screened, undone, refrigerated away? The theory of this book answers no — a constraint of the theory (Chapter 11) ties the field to the mass configuration with no gauge freedom left to hide it — and that "no" is precisely what makes the floor of Chapter 15 uncoolable, and the theory falsifiable.

### 2.4.3 Pointer states, and a first decoherence functional

One more lesson from the solvable model, because Part III inherits it whole. Why did the environment record *position* — why $|L\rangle$ versus $|R\rangle$ and not, say, $(|L\rangle \pm |R\rangle)/\sqrt{2}$? Because the interaction between grain and environment couples to position: a molecule bounces off the grain *where the grain is*. States that the interaction Hamiltonian leaves unentangled — its (approximate) eigenbasis — are the ones that survive observation; superpositions of them are the ones that decay. Zurek named them *pointer states*, and the selection of a preferred basis by the form of the coupling is *einselection*. Gravity, as we will see, couples to energy–momentum — which is why gravitational decoherence selects superpositions of *mass configurations* as its victims, and why the grain of Chapter 1 was the right thing to worry about.

**Worked Example 2.3 — A fully solvable dephasing model**

Let a central qubit couple to $N$ environment spins through

$$H = \frac{\sigma_z}{2} \otimes \sum_{k=1}^{N} g_k \, \sigma_z^{(k)},$$

the simplest Hamiltonian that measures $\sigma_z$ of the centre. Start in $|+\rangle \otimes \bigotimes_k |+\rangle_k$ with $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. Because $H$ is diagonal in the central $z$-basis, the branches $|0\rangle, |1\rangle$ evolve the environment oppositely: spin $k$ precesses by the angle $\mp g_k t$ depending on the branch. Each factor overlap is

$$\langle e_0^{(k)}(t) | e_1^{(k)}(t) \rangle = \langle + | \, e^{+ i g_k t \, \sigma_z^{(k)}} \, | + \rangle = \cos(g_k t)$$

(expand $|+\rangle$ in the $z$-basis: the two phases $e^{\pm i g_k t}$ average to the cosine). By (2.13) the visibility is

$$\mathcal{V}(t) = \prod_{k=1}^{N} |\cos(g_k t)| = \exp\Big[\textstyle\sum_k \ln|\cos g_k t|\Big] \approx \exp\Big[-\frac{t^2}{2} \textstyle\sum_k g_k^2\Big], \tag{2.15}$$

the last step for early times, $\ln\cos x \approx -x^2/2$. Two features deserve flags. *First*: the decay begins *quadratically* in $t$, not linearly — a universal short-time behaviour called the Gaussian onset (nothing can decay linearly at $t = 0$ if the global evolution is unitary; we prove this in general in Chapter 12, where the doubled-time ratio of the decay exponent, $\Gamma(2t)/\Gamma(t)$, built from it becomes an experimental discriminator between QGC and collapse models). *Second*: the exponent is a sum over environment modes of $(\text{coupling})^2 \times (\text{a function of } t)$ — exactly the shape, mode sum and all, that the gravitational decoherence functional of Chapter 11 will have, with $g_k$ replaced by the gravitational coupling of mode $k$ to the mass difference. This worked example is the whole book in a toy.

**What decoherence does and does not explain.** Decoherence explains why no experiment confined to the grain can exhibit interference between macroscopically distinct branches, and it selects which basis survives. It does *not* explain why a single outcome

> occurs, nor whence the probabilities: after (2.11) with $\langle E_L|E_R\rangle = 0$, the state is a mixture whose weights we *interpret* as probabilities via the Born rule $p_n = |c_n|^2$ — an input, not an output. The status ledger of this book therefore carries, permanently: Born rule [ASSUMED] — exactly as every laboratory application of quantum mechanics assumes it. QGC is a theory of dynamics at the quantum–gravity seam, not a solution of the measurement problem, and it never pretends otherwise.

## 2.5 Entropy: the ledger of missing information

Decoherence moves information; entropy counts it. We build the counter from its classical root, because the intuition transfers intact.

### 2.5.1 From surprise to Shannon

How much information does a probability distribution $\{p_i\}$ withhold? Shannon's answer: quantify the *surprise* of outcome $i$ as $-\ln p_i$ (rare outcomes surprise more; certain ones, $p = 1$, not at all; independent surprises add because probabilities multiply), and define the entropy as the expected surprise,

$$S_{\rm Sh}(\{p_i\}) \;=\; -\sum_i p_i \ln p_i \;\geq 0. \tag{2.16}$$

It vanishes only when one $p_i = 1$ (no missing information) and is maximal, $\ln d$, for the uniform distribution over $d$ outcomes (everything missing). We use natural logarithms throughout: entropy in *nats*, with $\ln 2 = 0.693$ nats to the bit; thermodynamics will later multiply by $k_{\rm B}$.

### 2.5.2 Von Neumann entropy

A quantum state is a density matrix, and a density matrix is — by the spectral theorem — a probability distribution $\{\lambda_k\}$ over its orthonormal eigenvectors. Von Neumann's entropy is Shannon's, applied to that canonical distribution:

$$S(\rho) \;=\; -\mathrm{Tr}\,(\rho \ln \rho) \;=\; -\sum_k \lambda_k \ln \lambda_k. \tag{2.17}$$

Its immediate properties, each one line from (2.17): $S = 0$ exactly for pure states (spectrum $\{1, 0, \dots\}$); $S \leq \ln d$ with equality for $\rho = \mathbb{1}/d$; and $S(U\rho U^\dagger) = S(\rho)$ for unitary $U$ (the spectrum does not move), so closed-system evolution creates no entropy — whatever entropy appears locally is imported through entanglement.

For the half of the Bell pair, (2.7) has spectrum $\{\frac{1}{2}, \frac{1}{2}\}$ and $S = \ln 2$: one full bit locally missing, the promised number for "perfect global knowledge, zero local knowledge." For the Bloch vector of length 0.8 (Worked Example 2.1), spectrum $\{0.9, 0.1\}$: $S = -0.9 \ln 0.9 - 0.1 \ln 0.1 = 0.325$ nats — partially known, partially opaque. When $\rho_A$ is the reduced state of a pure bipartite state, $S(\rho_A)$ is called the *entanglement entropy*, and by the equal spectra of (2.9), $S(\rho_A) = S(\rho_B)$ always — the ledger balances no matter how mismatched the parties.

▶ RUN IT: `TOOLS/verify-paper-numerics.py (section N: ch2 entropy values)`

### 2.5.3 Relative entropy and Klein's inequality

The most consequential member of the family measures not missing information but *distinguishability*: how surely can the state $\rho$ be told apart from a reference state $\sigma$?

$$S(\rho\|\sigma) \;=\; \operatorname{Tr}\rho\ln\rho - \operatorname{Tr}\rho\ln\sigma. \tag{2.18}$$

Its operational meaning (which we quote, not prove): after $n$ copies, the probability of mistaking $\rho$ for $\sigma$ decays as $e^{-nS(\rho\|\sigma)}$. Zero means indistinguishable; infinite means one measurement settles it. We flag now why this quantity, of all of them, carries Part III: when Chapter 6 forces us into systems with infinitely many degrees of freedom, $S(\rho)$ itself diverges and dies, but $S(\rho\|\sigma)$ — a *comparison* — survives with full rigor, and Axiom II of Chapter 9 is a statement about it. Everything therefore rests on one fact:

**Theorem 2.3** (Klein's inequality)**.** *$S(\rho\|\sigma) \geq 0$, with equality if and only if $\rho = \sigma$.*

*Derivation.* Diagonalise both states: $\rho = \sum_i p_i \,|i\rangle\langle i|$ and $\sigma = \sum_a q_a \,|a\rangle\langle a|$, with $\{|i\rangle\}$, $\{|a\rangle\}$ two orthonormal bases. Insert into (2.18), evaluating the trace in the $\{|i\rangle\}$ basis. The first term is $\sum_i p_i \ln p_i$. For the second, $\langle i|\ln\sigma\,|i\rangle = \sum_a P_{ia}\ln q_a$ with $P_{ia} \equiv |\,\langle i|a\rangle\,|^2$; hence

$$S(\rho\|\sigma) = \sum_i p_i\Big(\ln p_i - \sum_a P_{ia}\ln q_a\Big).$$

The matrix $P_{ia}$ is *doubly stochastic*: $\sum_a P_{ia} = 1$ (row: completeness of $\{|a\rangle\}$) and $\sum_i P_{ia} = 1$ (column: completeness of $\{|i\rangle\}$). Now use concavity of the logarithm (Jensen's inequality) on each row: $\sum_a P_{ia}\ln q_a \leq \ln r_i$ where $r_i \equiv \sum_a P_{ia}q_a$ is a probability distribution ($r_i \geq 0$; $\sum_i r_i = \sum_a q_a = 1$ by column stochasticity). Therefore

$$S(\rho\|\sigma) \;\geq\; \sum_i p_i \ln\frac{p_i}{r_i}.$$

The right side is a *classical* relative entropy, and it is non-negative by the elementary bound $\ln x \leq x - 1$ (with equality only at $x = 1$): $\sum_i p_i \ln(r_i/p_i) \leq \sum_i p_i(r_i/p_i - 1) = \sum_i r_i - \sum_i p_i = 0$. Equality throughout requires $r_i = p_i$ and no Jensen gap, which forces $P$ to be a permutation on the support — i.e. common eigenbasis and equal spectra — i.e. $\rho = \sigma$.

One corollary shows the leverage. Take $\sigma = \mathbb{1}/d$: then $S(\rho\|\mathbb{1}/d) = \ln d - S(\rho) \geq 0$ proves the maximum-entropy bound quoted below (2.17) — and this pattern, extremal statements about entropy recast as positivity of a relative entropy, is exactly how Einstein's equations will be derived in Chapter 10.

Finally, for two parties the natural correlation measure is the *mutual information*

$$I(A\!:\!B) \;=\; S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \;=\; S(\rho_{AB}\,\|\,\rho_A\otimes\rho_B) \;\geq\; 0, \tag{2.19}$$

where the second equality is a two-line expansion of (2.18) (do it: $\ln(\rho_A\otimes\rho_B) = \ln\rho_A\otimes\mathbb{1} + \mathbb{1}\otimes\ln\rho_B$, and the partial traces marginalise) and the inequality is Theorem 2.3. For the Bell pair: $I = \ln 2 + \ln 2 - 0 = 2\ln 2$ — correlations worth two bits, the maximum two qubits can share, twice what any classical pair of coins can manage ($\leq \ln 2$). That factor of two is entanglement, counted.

**Q 2.4 *"You keep saying entropy will "diverge and die" in field theory. Why should I believe the fix is relative entropy rather than some better regularisation of $S$ itself?"***

Because the divergence is physical, not an artifact. In a quantum field theory every region of space is entangled with its complement at every length scale, down to arbitrarily short distances — we will see this explicitly in the Rindler calculation of Chapter 4 — so the entanglement entropy of a region is genuinely infinite, and any cutoff that renders it finite imports arbitrariness (the notorious scheme-dependence of the "area term" in Chapter 5). Relative entropy sidesteps the infinity honestly: the divergent parts of $\mathrm{Tr}\rho \ln \rho$ and $\mathrm{Tr}\rho \ln \sigma$ are the same short-distance entanglement and cancel in the difference (2.18), leaving a finite, cutoff-independent, mathematically rigorous quantity (Araki's construction, Chapter 7, defines it with no density matrices at all). The pattern — *only differences and comparisons of entropy are physical* — is Axiom I's fine print and one of the honest lessons of the whole subject.

## Chapter FAQ

**Q: Does decoherence pick out a unique pointer basis, or just approximately?** Approximately, with the approximation superbly good for macroscopic couplings. The criterion (the "predictability sieve"): pointer states are those that, under the actual interaction, entangle *slowest* with the environment. For position-coupled interactions they are well-localised wavepackets, not exact position eigenstates (which would have infinite momentum spread and decohere in other ways). Nothing in Part III hangs on exactness; the gravitational coupling selects mass-configuration branches with the same robust approximateness.

**Q: Whose entropy is it? If $S(\rho_A)$ depends on the cut between $A$ and $B$, is entropy subjective?** Entropy is *relational*: it is a property of a state *and* a choice of accessible observables (here, the split $A|B$). Once the split is fixed — by the physics, e.g. by what an observer bounded by a horizon can reach — the entropy is as objective as any expectation value. Making "the observables an observer can reach" into the primary mathematical object is precisely the move of Chapter 6, and the theory of this book is built on the version of quantum mechanics in which the algebra, not the global Hilbert space, comes first.

**Q: Why natural logs and nats instead of bits, and where is $k_\mathrm{B}$?** Conventions only. Information theory prefers $\log_2$ (bits); thermodynamics uses $k_\mathrm{B} \ln$ (joules per kelvin); mathematics uses ln (nats). They differ by constant factors: 1 bit $= \ln 2 = 0.693$ nats; $S_\mathrm{thermo} = k_\mathrm{B} S_\mathrm{nats}$. We keep nats and restore $k_\mathrm{B}$ only when meeting laboratory temperatures, first in the Unruh formula of Chapter 4.

**Q: All your measurements are projective. Don't real detectors need something more general?** Yes — the general framework (POVMs) matters for quantum information practice, but every POVM is a projective measurement on a larger system (one more use of purification), and nothing in this book needs the generality: the seam physics of Part III is about states and dynamics, with measurements entering only as idealised readouts. We stay projective and lose nothing.

## Exercises

**Exercise 2.1.** (a) For $|\Psi(\theta)\rangle = \cos\theta\,|00\rangle + \sin\theta\,|11\rangle$, find $\rho_A$ and the entanglement entropy $S(\theta)$; evaluate at $\theta = \pi/6$ and locate the maximum. (b) Now let the environment keep only a *partial* record: $|\Psi\rangle = (|0\rangle|e_0\rangle + |1\rangle|e_1\rangle)/\sqrt{2}$ with real overlap $\langle e_0|e_1\rangle = \cos\varphi$. Compute $\rho_A$, the fringe visibility $\mathcal{V}$ available to an interferometer acting on $A$ alone, the eigenvalues of $\rho_A$, and $S(\rho_A)$; evaluate at $\varphi = \pi/3$. Check the two limits $\varphi = 0$ and $\varphi = \pi/2$ and state the complementarity between $S$ and $\mathcal{V}$ that your formulas exhibit.

**Exercise 2.2.** A grain's superposition suffers independent environment collisions, each leaving which-path overlap $1-\epsilon$ with $\epsilon = 10^{-3}$. (a) How many collisions reduce the visibility to $1/e$? (Give both the short answer from (2.14) and the exact one from $(1-\epsilon)^N = e^{-1}$.) (b) If collisions arrive at $\gamma = 10^{20}\,\mathrm{s}^{-1}$ (microgram grain in room air), what is $\tau_{\mathrm{env}}$, and what does comparing it with the $\tau_{\mathrm{dec}} = 1.6$ ns of Chapter 1 say about the experimental strategy of Chapter 15?

## Solutions

*Solution (Exercise 2.1).* (a) The state is already in Schmidt form (2.8) with $\lambda_1 = \cos^2\theta$, $\lambda_2 = \sin^2\theta$, so $\rho_A = \mathrm{diag}(\cos^2\theta, \sin^2\theta)$ and $S(\theta) = -\cos^2\theta\ln\cos^2\theta - \sin^2\theta\ln\sin^2\theta$. At $\theta = \pi/6$: $\lambda = (0.75, 0.25)$, so $S = -0.75\ln 0.75 - 0.25\ln 0.25 = 0.562$ nats. The maximum is at $\theta = \pi/4$ (equal $\lambda$'s), $S = \ln 2 = 0.693$ nats: the Bell state. (Note the visibility law in passing: here the environment records are the orthogonal $|0\rangle_B, |1\rangle_B$, so $\mathcal{V} = |\langle 0_B|1_B\rangle| = 0$ — perfect which-path recording, no local fringes at any $\theta$.) (b) By the computation that gave (2.11),

$$\rho_A = \frac{1}{2}\begin{pmatrix} 1 & \cos\varphi \\ \cos\varphi & 1 \end{pmatrix}, \qquad \mathcal{V} = |\langle e_0|e_1\rangle| = \cos\varphi.$$

Its eigenvalues are $(1\pm\cos\varphi)/2$ (Bloch vector of length $\cos\varphi$ along $x$; Worked Example 2.1), so

$$S(\rho_A) = -\frac{1+\mathcal{V}}{2}\ln\frac{1+\mathcal{V}}{2} - \frac{1-\mathcal{V}}{2}\ln\frac{1-\mathcal{V}}{2}.$$

At $\varphi = \pi/3$: $\mathcal{V} = \frac{1}{2}$, eigenvalues $(0.75, 0.25)$, $S = 0.562$ nats — the same number as (a), reached by a different route, because both describe one bit recorded to the same partial degree. Limits: $\varphi = 0$ gives $\mathcal{V} = 1$, $S = 0$ (no record, full fringes); $\varphi = \pi/2$ gives $\mathcal{V} = 0$, $S = \ln 2$ (complete record, no fringes). The complementarity is monotone and exact: $S$ is a strictly decreasing function of $\mathcal{V}$ — every nat the environment learns is paid for in fringe contrast, with the exchange rate set by the formulas above. Gravitational decoherence (Chapter 11) is this exercise with $|e_{0,1}\rangle$ two coherent states of the gravitational field.

▶ RUN IT: `TOOLS/verify-paper-numerics.py (section N: ch2 entropy values)`

*Solution (Exercise 2.2).* (a) From $e^{-N\epsilon} = e^{-1}$: $N = 1/\epsilon = 1000$ collisions. Exactly: $N = 1/[-\ln(1-10^{-3})] = 999.5$ — the approximation is good to $5\times10^{-4}$, as expected since the error term is $\epsilon/2$. (b) $\tau_{\mathrm{env}} = 1/(\gamma\epsilon) = 1/(10^{20}\times10^{-3}) = 10^{-17}$ s. Air decoheres the grain $10^8$ times faster than gravity would ($\tau_{\mathrm{dec}} = 1.6\times10^{-9}$ s): in any ordinary environment the gravitational effect is invisible, buried under removable noise. The strategy of Chapter 15 follows: push $\gamma\epsilon$ down (ultra-high vacuum, cryogenics, small polarisability) until $\tau_{\mathrm{env}} \gg \tau_{\mathrm{dec}}$, so that the *uncoolable* gravitational floor is the first thing the fringes feel. The experiment is a race between engineering ($\tau_{\text{env}}$, improvable without limit in principle) and a constant of nature ($\tau_{\text{dec}}$, not improvable at fixed $M, d$) — which is what makes it decisive.

▶ RUN IT: `TOOLS/verify-paper-numerics.py (section N: ch2 collision counts)`

**What to carry forward**

1. The observer's true state is the density matrix $\rho$ ($\langle A \rangle = \mathrm{Tr}\rho A$); the reduced state (2.6) is the unique local truth, and mixedness from entanglement is objective — no hidden information anywhere would purify it.
2. Decoherence is entanglement seen from one side: interference terms are multiplied by the environment overlap, $\mathcal{V} = |\langle E_L | E_R \rangle|$, which factorises over independent records and decays exponentially — with a universal Gaussian onset (2.15) whose mode-sum form is exactly what gravitational decoherence will inherit. The Born rule stays [ASSUMED].
3. Entropy counts what entanglement hides: $S(\rho) = -\mathrm{Tr}\rho \ln \rho$, equal for the two halves of any pure state; relative entropy $S(\rho \| \sigma) \geq 0$ (Klein, proved) measures distinguishability and is the only member of the family that will survive, rigorously, in quantum field theory — which is why the axioms of Part III are written in it.

# 3 General Relativity, Operationally

**Synopsis.** Gravity is not a force but the geometry of spacetime, and this chapter builds that claim from one operational fact: inside a freely falling box, gravity disappears. We ride that fact from the equivalence principle to gravitational redshift (gravity is the geometry of *time*), to the metric and geodesics, to curvature and Einstein's equations — and we discharge the promissory note of Chapter 1 by proving the contracted Bianchi identity $\nabla^\mu G_{\mu\nu} = 0$ that killed semiclassical gravity. The second half assembles the geometric toolkit the rest of the book runs on: the *Rindler horizon* an accelerated observer cannot see behind (load-bearing for the Unruh temperature of Chapter 4), the Schwarzschild horizon of a black hole, the expanding FRW universe with its de Sitter horizon at distance $c/H$ (everything Chapter 14 needs), and the Raychaudhuri equation for a bundle of light rays — the single equation that will turn entanglement into Einstein's equations in Chapter 10.

**You will be able to.** derive gravitational redshift from the equivalence principle; write the geodesic equation and read off the Christoffel symbols; construct the Riemann, Ricci, and Einstein tensors and prove $\nabla^\mu G_{\mu\nu} = 0$; compute the Schwarzschild radius of any mass; set up Rindler and FRW coordinates and locate their horizons; and derive the null Raychaudhuri equation in full.

**Prerequisites.** Chapters 1–2; special relativity at the undergraduate level (four-vectors, the Minkowski metric $\eta_{\mu\nu} = \mathrm{diag}(-1, +1, +1, +1)$, time dilation, proper time); vector calculus and matrix algebra. *General* relativity is built here from scratch, assuming none of it.

## 3.1 The one fact gravity is made of

Newton wrote two laws in which the same symbol $m$ appears for what look like two unrelated things. In $F = ma$ the mass is *inertia* — the resistance of a body to being accelerated by any force whatever. In $F = GMm/r^2$ the mass is *gravitational charge* — the strength with which a body couples to gravity, the analogue of electric charge in Coulomb's law. There is no reason in Newton's framework why these two numbers should be equal. A body with twice the gravitational charge but the same inertia would fall twice as fast; the theory permits it. Yet every experiment ever done finds the two masses equal to spectacular precision — Eötvös, then the lunar laser-ranging tests, now to parts in $10^{15}$. Every body, whatever its composition, falls with the same acceleration in a given gravitational field.

Einstein took this coincidence not as an accident to be tolerated but as the deepest clue in physics, and raised it — in the manner of a postulate — to a principle. Let us state it the way he did, operationally, in terms of what an observer sealed in a box can measure.

**Principle 3.1** (Equivalence principle)**.** *No experiment performed inside a small enough, sealed laboratory can distinguish* *(a) the laboratory at rest in a uniform gravitational field of strength $g$, from*

*(b) the laboratory accelerating with acceleration $a = g$ through empty space, far from all masses; and equivalently, no experiment can distinguish*

*(c) a laboratory in free fall in a gravitational field, from*

*(d) a laboratory floating in empty space with no field at all.*

The two halves say the same thing twice. In (c)–(d), free fall *cancels* gravity: the astronaut and her pen fall together, so relative to her the pen floats, exactly as it would in deep space. Because inertial and gravitational mass are equal, this cancellation is universal — it works for every object at once, which no ordinary force can manage (an electric field lifts charges but not neutral atoms). Gravity is the one "force" you can transform away everywhere at once by choosing to fall. That is the operational content of Principle 3.1, and everything in this chapter is its consequence.

The word "small" in the statement is not a hedge; it is the physics. Over a large box, free fall does *not* perfectly cancel gravity: two pens dropped on opposite sides of a falling elevator over the Earth drift slightly *toward* each other, because they fall along radii that converge at the Earth's centre. This residual, uncancellable relative acceleration is the *tidal* field, and — we will see in §3.4 — it is precisely what no change of coordinates can remove. Uniform gravity is fake, a choice of accelerated frame; tidal gravity is real, and its name in the finished theory is *curvature*.

**Q 3.1** ***"If gravity is just an accelerated frame, why can't I transform away the Earth's gravity everywhere at once and be done with it?"***

Because the Earth's field is not uniform — it points toward the centre, so it aims in different directions at different places. You can cancel it at one point (jump, and locally gravity vanishes for you), and to first order in a small neighbourhood of that point. But there is no single accelerated frame that cancels a field pointing north here and south on the far side of the planet simultaneously. The part of gravity that *is* a choice of frame — the uniform part, at a point — is what the equivalence principle lets you delete. The part that survives in every frame — the convergence of the field lines, the tide — is genuine, coordinate-independent geometry. This split, "locally removable versus globally real," is the mathematical spine of the whole subject: it is the difference between the Christoffel symbols of §3.3 (removable) and the Riemann tensor of §3.4 (not).

## 3.2 Gravity is the geometry of time

Before any metric, any tensor, we can already extract from Principle 3.1 a prediction that overturns the Newtonian picture: *clocks run at different rates at different gravitational potentials.* We derive it in full, using nothing but the equivalence principle and the ordinary Doppler effect of special relativity.

### 3.2.1 Redshift from the accelerating elevator

Consider case (b) of the principle: a rocket (an elevator, a sealed box) of height $h$ accelerating with constant acceleration $g$ through empty space, in the $+z$ direction. At the floor is an emitter;

at the ceiling, a receiver. The floor emits a light wave of frequency $\nu_{\text{em}}$ straight up. We ask what frequency the ceiling receives, and then invoke the principle to reinterpret the answer as a statement about gravity.

**Worked Example 3.1 — Gravitational redshift from acceleration**

Work in the momentarily inertial frame in which the rocket is at rest at the instant of emission ($t = 0$). The light travels the height $h$ at speed $c$, so it reaches the ceiling after a light-travel time

$$\Delta t = \frac{h}{c}. \tag{3.1}$$

This is the time for the signal to cross the box, computed in the frame where the box was instantaneously still.

During that interval the rocket has been accelerating, so by the time the light arrives the ceiling is no longer at rest in our inertial frame: it has picked up a velocity

$$\Delta v = g\,\Delta t = \frac{gh}{c} \tag{3.2}$$

in the $+z$ direction — *away* from where the light came from. The receiver is therefore receding from the (now past) emission event, and it sees the light Doppler-*shifted to lower frequency*. For $\Delta v \ll c$ the first-order Doppler formula gives the fractional shift

$$\frac{\nu_{\text{rec}} - \nu_{\text{em}}}{\nu_{\text{em}}} = -\frac{\Delta v}{c} = -\frac{gh}{c^2}. \tag{3.3}$$

The received frequency is lower: light climbing the accelerating box is redshifted.

Now apply Principle 3.1: the accelerating rocket in empty space is indistinguishable from a box at rest in a uniform gravitational field $g$, with the floor at gravitational potential $\Phi_{\text{floor}}$ and the ceiling higher up at $\Phi_{\text{ceil}} = \Phi_{\text{floor}} + gh$. Writing $\Delta\Phi = gh$ for the potential difference climbed, the shift (3.3) becomes

$$\boxed{\frac{\Delta\nu}{\nu} = -\frac{\Delta\Phi}{c^2}} \tag{3.4}$$

Light climbing *out* of a gravitational potential well loses frequency in proportion to the potential climbed; light falling in gains it. The apparatus never knew whether it was accelerating or sitting in a field — the principle guarantees the same answer — so this is a statement about gravity.

Equation (3.4) was confirmed in the laboratory by Pound and Rebka in 1960, sending gamma rays up a 22.5 m tower at Harvard and measuring a fractional shift of $2.5 \times 10^{-15}$, and it is the effect GPS satellites must correct for every day. But its deep meaning is not the number; it is what frequency *is*. Frequency is the ticking rate of a clock: a wave of frequency $\nu$ delivers $\nu$ crests per second, and each crest is one tick. If the ceiling receives fewer crests per second than the floor emitted, the only consistent reading is that *the floor clock is running slow compared to the ceiling clock*. Deeper in the potential well, time itself runs slower.

$$\frac{d\tau_{\text{floor}}}{d\tau_{\text{ceil}}} = 1 - \frac{\Delta\Phi}{c^2} < 1 \qquad \text{(clocks deeper in the well tick slower).} \tag{3.5}$$

This is the sentence that reorganises everything: *gravity is the geometry of time.* A gravitational field is a place where the rate of time varies from point to point, and — as we are about to see — a falling body is simply following the route that ages it the most. Nothing here required the full machinery of general relativity; the equivalence principle plus special-relativistic Doppler was enough. The machinery exists to make this quantitative and to add the geometry of *space* to the geometry of time. [PROVEN]

**Q 3.2** ***"You used the flat-space Doppler formula inside a gravitational field. Isn't that circular — assuming special relativity to derive a correction to it?"***

It is not circular, because the derivation lives entirely in case (b): a rocket accelerating through *genuinely empty, flat* spacetime, where special relativity holds exactly and the Doppler formula is unimpeachable. The only physical input beyond special relativity is Principle 3.1, which lets us *reinterpret* the flat-space answer as a gravitational one. We never assumed a gravitational Doppler formula; we derived it. The subtlety you are sensing is real but lies elsewhere: the argument is exact only for a *uniform* field over a *small* height, precisely the regime where gravity is locally indistinguishable from acceleration. For strong fields (§3.7) the exact result replaces $\Delta\Phi/c^2$ by a metric factor, but the linearised law (3.4) is its weak-field limit, and (3.4) is what the Pound–Rebka experiment measured.

## 3.3 Proper time, the metric, and free fall

To make "the geometry of time" quantitative we need the object that assigns a duration to every worldline and a length to every path. That object is the metric.

### 3.3.1 Proper time and the metric

In special relativity the proper time $\tau$ along a worldline — the time a clock carried along it actually reads — is built from the Minkowski line element. For two neighbouring events separated by $(dt, dx, dy, dz)$,

$$-c^2 d\tau^2 = ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 = \eta_{\mu\nu}\, dx^\mu dx^\nu, \tag{3.6}$$

with $x^\mu = (ct, x, y, z)$, $\eta_{\mu\nu} = \mathrm{diag}(-1, +1, +1, +1)$, and the Einstein summation convention (repeated upper–lower indices are summed 0 to 3) in force from here on. The interval $ds^2$ is what all inertial observers agree on; a moving clock reads less time ((3.6) with $dx \neq 0$ gives smaller $|d\tau|$), which is ordinary time dilation.

General relativity makes exactly one change to (3.6), but it is the change that contains all of gravity: the constant matrix $\eta_{\mu\nu}$ is replaced by a *field* $g_{\mu\nu}(x)$ — a symmetric $4 \times 4$ matrix that varies from point to point:

$$ds^2 = g_{\mu\nu}(x)\, dx^\mu dx^\nu, \qquad g_{\mu\nu} = g_{\nu\mu}. \tag{3.7}$$

The metric $g_{\mu\nu}(x)$ *is* the gravitational field. Operationally it is defined by what it does: it is the machine that, fed a small displacement $dx^\mu$, returns the proper time or proper distance between the two events, so it is measured by carrying clocks and rulers around and recording their readings. Where $g_{00}$ differs from $-1$, clocks tick at an altered rate — that is the redshift of §3.2, now encoded in a single component. The equivalence principle reappears as a theorem about (3.7): at any single point $p$ one can always choose coordinates (a freely falling frame) in which $g_{\mu\nu}(p) = \eta_{\mu\nu}$ and its first derivatives vanish, so that gravity disappears locally. What cannot be removed is the *second* derivatives — the tides.

**Worked Example 3.2 — The metric of Newtonian gravity**

Let us find $g_{00}$ in the weak, slow, static regime and check it against §3.2. Consider a clock at rest ($dx = dy = dz = 0$) at a point where the metric is $g_{00}(x)$. By (3.7) the proper time it records over a coordinate-time interval $dt$ is

$$d\tau \;=\; \sqrt{-g_{00}/c^2}\, c\, dt \;=\; \sqrt{-g_{00}}\, dt.$$

Comparing two static clocks at potentials $\Phi_1$ and $\Phi_2$, the ratio of their rates is $\sqrt{g_{00}(1)/g_{00}(2)}$. But we already know this ratio from (3.5): to first order it is $1 + (\Phi_1 - \Phi_2)/c^2$. Matching the two, write $g_{00} = -(1 + 2\Phi/c^2)$ and expand $\sqrt{-g_{00}} = \sqrt{1 + 2\Phi/c^2} \approx 1 + \Phi/c^2$; the ratio of two such factors is indeed $1 + (\Phi_1 - \Phi_2)/c^2$ to first order. Hence

$$g_{00} \;=\; -\Big(1 + \frac{2\Phi}{c^2}\Big), \tag{3.8}$$

with $\Phi$ the Newtonian gravitational potential ($\Phi = -GM/r$ outside a mass $M$). One number encodes the entire Newtonian field: the 00 component of the metric *is* the gravitational potential, rescaled. We will need this identification twice — to reduce Einstein's equations to Newton's in §3.5, and as the anchor of the whole weak-field program.

### 3.3.2 Free fall is the straightest possible path

We now answer: how does a body move in the geometry (3.7)? The equivalence principle already told us — a freely falling body feels no force; it is the unaccelerated, "straight" worldline of the geometry. In a curved geometry the straightest available path is called a *geodesic*, and there is a clean variational way to characterise it, which for timelike worldlines has a startling physical reading.

**Principle 3.2** (Geodesic principle / extremal ageing)**.** *A freely falling body travels the worldline that* extremises the proper time *between two events:*

$$\tau[\gamma] \;=\; \int_\gamma d\tau \;=\; \frac{1}{c}\int_\gamma \sqrt{-g_{\mu\nu}\, dx^\mu dx^\nu} \;=\; \textit{extremum}. \tag{3.9}$$

The physical content is worth stating before the algebra: *of all the ways to get from event $A$ to event $B$, a freely falling clock takes the one that makes it age the most.* A thrown ball rises and falls not to minimise some abstract action but because that arc is the path of maximal ageing through the warped time of §3.2 — climbing higher buys the clock faster ticking (less time dilation from the potential), at the cost of the speed needed to get there and back (more time dilation from motion), and the actual trajectory is the optimal trade. This is the "geometry of time" made into a law of motion.

To turn (3.9) into an equation of motion, parametrise the worldline by proper time $\tau$ and apply the calculus of variations. The Euler–Lagrange equations for the Lagrangian $L = \frac{1}{2}g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu$ (with $\dot{x}^\mu = dx^\mu/d\tau$; extremising $L$ is equivalent to extremising the proper time and is smoother to differentiate) read $\frac{d}{d\tau}\frac{\partial L}{\partial\dot{x}^\alpha} - \frac{\partial L}{\partial x^\alpha} = 0$. Compute the two pieces: $\partial L/\partial\dot{x}^\alpha = g_{\alpha\nu}\dot{x}^\nu$ and $\partial L/\partial x^\alpha = \frac{1}{2}(\partial_\alpha g_{\mu\nu})\dot{x}^\mu\dot{x}^\nu$, so

$$g_{\alpha\nu}\ddot{x}^\nu + (\partial_\mu g_{\alpha\nu})\dot{x}^\mu\dot{x}^\nu - \tfrac{1}{2}(\partial_\alpha g_{\mu\nu})\dot{x}^\mu\dot{x}^\nu = 0.$$

Symmetrise the middle term in $\mu \leftrightarrow \nu$ (legal because $\dot{x}^\mu\dot{x}^\nu$ is symmetric), multiply through by the inverse metric $g^{\alpha\beta}$ (defined by $g^{\alpha\beta}g_{\beta\gamma} = \delta^\alpha_\gamma$), and the result is the *geodesic equation*:

$$\frac{d^2x^\rho}{d\tau^2} + \Gamma^\rho{}_{\mu\nu}\,\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 0, \qquad \Gamma^\rho{}_{\mu\nu} = \tfrac{1}{2}\,g^{\rho\sigma}\left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right). \tag{3.10}$$

The quantities $\Gamma^\rho{}_{\mu\nu}$ are the *Christoffel symbols*. Read the equation as Newton's second law for gravity: the "acceleration" $d^2x^\rho/d\tau^2$ is not zero (as it would be for a free particle in flat space, a straight line) but is fixed by the $\Gamma$'s, which are built entirely from first derivatives of the metric. The Christoffel symbols *are* the gravitational "force" per unit mass — and, crucially, they are the removable part: at any point one can choose the freely falling frame in which all $\Gamma^\rho{}_{\mu\nu}$ vanish, which is the equivalence principle again. A quantity that can be set to zero by a change of coordinates cannot be the true, invariant field. The invariant field is one derivative up, in §3.4.

### Worked Example 3.3 — Recovering Newton's $\ddot{\vec{x}} = -\nabla\Phi$

Take a slow body ($|\dot{\vec{x}}| \ll c$, so $dx^i/d\tau \approx 0$ and $dx^0/d\tau \approx c$) in the weak static field (3.8). In the geodesic equation (3.10) the dominant term in the sum is $\mu = \nu = 0$:

$$\frac{d^2x^i}{d\tau^2} \;\approx\; -\Gamma^i{}_{00}\,(dx^0/d\tau)^2 \;=\; -\Gamma^i{}_{00}\,c^2.$$

Now evaluate $\Gamma^i{}_{00}$ from (3.10). For a static field $\partial_0 g_{\mu\nu} = 0$, so $\Gamma^i{}_{00} = \frac{1}{2}g^{ij}(2\partial_0 g_{j0} - \partial_j g_{00}) = -\frac{1}{2}\partial_i g_{00}$ (using $g^{ij} \approx \delta^{ij}$ to leading order). With $g_{00} = -(1 + 2\Phi/c^2)$, $\partial_i g_{00} = -\frac{2}{c^2}\partial_i\Phi$, so $\Gamma^i{}_{00} = \frac{1}{c^2}\partial_i\Phi$. Substituting,

$$\frac{d^2x^i}{dt^2} \;=\; -\partial_i\Phi \qquad\Longleftrightarrow\qquad \ddot{\vec{x}} = -\nabla\Phi. \tag{3.11}$$

Newton's law of gravitation drops out of the geometry, with the identification (3.8) doing all the work: gravitational "force" is the gradient of the time-time component of the metric. General relativity does not contradict Newton; it explains him, and tells us $\Phi$ was the metric all along.

**Q 3.3** ***"So the Christoffel symbols are the gravitational force. But you just said they can be made zero by choosing coordinates. How can the force be real if I can transform it away?"***

You have put your finger on the conceptual heart of the theory. In Newtonian physics the gravitational force at a point is an absolute fact. In general relativity it is not: what you call "the gravitational force" depends on your state of motion, and a freely falling observer measures it to be zero (she floats). This is not a paradox — it is the equivalence principle stated in the language of $\Gamma$. The genuinely coordinate-*independent* facts of gravity are the ones no observer can transform away: the tidal effects, the relative acceleration of nearby freely falling bodies. Those are governed by derivatives of the $\Gamma$'s, and they assemble into the Riemann curvature tensor of the next section. A useful slogan: $\Gamma$ *is how gravity looks;* $R$ *is what gravity is.*

## 3.4 Curvature: the part of gravity you cannot escape

We have twice promised that the invariant field lives one derivative above the Christoffel symbols. We now construct it operationally, from the one gravitational effect no freely falling observer can make vanish: the failure of parallel transport to commute.

### 3.4.1 Parallel transport and the loop test

In flat space, to compare a vector at point $A$ with a vector at a distant point $B$ you slide it over, keeping it parallel to itself — "parallel transport." On a curved surface this operation depends on the path. Carry an arrow that points south, starting at the equator: walk it north to the pole along one meridian, then it points along that meridian; instead walk it a quarter of the way around the equator first and *then* north to the pole, and it arrives rotated by 90°. Same start, same end, different answer — the discrepancy is the operational signature of curvature, and it is detectable entirely from within the surface, with no need to see it embedded in a higher space. This is why a two-dimensional creature confined to the sphere could discover its curvature by transporting arrows around loops.

The covariant version of "slide it keeping it parallel" is the *covariant derivative* $\nabla_\mu$, which corrects the ordinary derivative for the twisting of the coordinate basis encoded in $\Gamma$:

$$\nabla_\mu V^\rho \;=\; \partial_\mu V^\rho + \Gamma^\rho{}_{\mu\sigma} V^\sigma, \qquad \nabla_\mu V_\rho \;=\; \partial_\mu V_\rho - \Gamma^\sigma{}_{\mu\rho} V_\sigma. \tag{3.12}$$

The $\Gamma$ terms are exactly what make $\nabla_\mu V^\rho$ transform as a proper tensor even though $\partial_\mu V^\rho$ does not; a vector is parallel-transported along a curve when its covariant derivative along the curve vanishes. Note in passing that the geodesic equation (3.10) is just $u^\mu \nabla_\mu u^\rho = 0$ — a geodesic parallel-transports its own velocity, i.e. it never turns, the precise sense in which it is "straight."

### 3.4.2 The Riemann tensor

The loop test made quantitative: transport a vector around an infinitesimal closed loop spanned by the two directions $\mu$ and $\nu$, and the vector comes back rotated by an amount linear in the loop's area. The rotation per unit area is the *Riemann curvature tensor*, and the cleanest definition is that it measures the failure of two covariant derivatives to commute:

$$[\nabla_\mu, \nabla_\nu]\, V^\rho \;=\; R^\rho{}_{\sigma\mu\nu}\, V^\sigma, \qquad R^\rho{}_{\sigma\mu\nu} = \partial_\mu \Gamma^\rho{}_{\nu\sigma} - \partial_\nu \Gamma^\rho{}_{\mu\sigma} + \Gamma^\rho{}_{\mu\lambda}\Gamma^\lambda{}_{\nu\sigma} - \Gamma^\rho{}_{\nu\lambda}\Gamma^\lambda{}_{\mu\sigma}. \tag{3.13}$$

The explicit form is obtained by writing out the commutator on the left with (3.12) and collecting terms; the ordinary derivatives cancel and the $\Gamma$'s survive. Read the structure: $R$ is *first derivatives of* $\Gamma$ *plus* $\Gamma$*-squared* — that is, second derivatives of the metric. This is the level at which gravity becomes invariant. If spacetime is flat, one can find coordinates with $g_{\mu\nu} = \eta_{\mu\nu}$ everywhere, all $\Gamma = 0$, hence $R = 0$; conversely $R^\rho{}_{\sigma\mu\nu} = 0$ everywhere is exactly the condition that coordinates flattening the metric exist. *Curvature is nonzero if and only if gravity is real* — not a frame artifact. The tidal field of §3.1 is $R$: two nearby geodesics separated by $\xi^\mu$ accelerate apart according to the equation of geodesic deviation $\frac{D^2\xi^\rho}{d\tau^2} = -R^\rho{}_{\mu\sigma\nu}u^\mu \xi^\sigma u^\nu$, whose right side is nonzero precisely when $R$ is.

### 3.4.3 Ricci, scalar, and Einstein tensors

The Riemann tensor has 20 independent components in four dimensions — more than we need to build a field equation. We distil it by *contraction*, summing a pair of indices. The trace on the first and third indices gives the *Ricci tensor*; a further trace gives the *Ricci scalar*:

$$R_{\mu\nu} \;\equiv\; R^\rho{}_{\mu\rho\nu}, \qquad R \;\equiv\; g^{\mu\nu} R_{\mu\nu}. \tag{3.14}$$

The Ricci tensor is symmetric, $R_{\mu\nu} = R_{\nu\mu}$, and it captures the part of curvature that a ball of freely falling test particles feels as a change of *volume*: $R_{\mu\nu}u^\mu u^\nu$ controls how fast a small comoving cloud of dust contracts, a fact we will use verbatim in the Raychaudhuri equation of §3.9. The scalar $R$ is a single number at each point, the most compressed measure of curvature. From these two we assemble the object Einstein's equations need, the *Einstein tensor*:

$$G_{\mu\nu} \;\equiv\; R_{\mu\nu} - \tfrac{1}{2}\, R\, g_{\mu\nu}. \tag{3.15}$$

This is the "package of second derivatives of the geometry" that Chapter 1 promised to construct honestly; it is symmetric, and it is built to satisfy a conservation identity we now prove.

### 3.4.4 The contracted Bianchi identity — discharging the note

In Chapter 1 the downfall of semiclassical gravity was the identity $\nabla^\mu G_{\mu\nu} = 0$ (equation (1.6) there), which we asserted holds for *any* geometry, like $\nabla \cdot (\nabla \times \vec{A}) = 0$. We now redeem that promissory note.

**Theorem 3.3** (Contracted Bianchi identity)**.** *For any metric whatsoever, the Einstein tensor is covariantly divergence-free:*

$$\nabla^\mu G_{\mu\nu} \;=\; 0. \tag{3.16}$$

*Derivation.* The starting point is a symmetry of the Riemann tensor that follows directly from its definition (3.13) by antisymmetrising the covariant derivative over three indices. In a locally inertial frame at a point (where all $\Gamma = 0$, though their derivatives need not vanish) the second-derivative structure of (3.13) gives, upon cyclic antisymmetrisation of the last three indices,

$$\nabla_\lambda R_{\rho\sigma\mu\nu} + \nabla_\mu R_{\rho\sigma\nu\lambda} + \nabla_\nu R_{\rho\sigma\lambda\mu} \;=\; 0. \tag{3.17}$$

This is the (full, second) *Bianchi identity*; each term is a covariant derivative, so an equation true in the inertial frame at a point is true as a tensor equation everywhere. Now contract it twice. First contract $\rho$ with $\mu$ using the inverse metric (i.e. raise $\rho$ and set it equal to $\mu$), and use $g^{\rho\mu}\nabla_\lambda R_{\rho\sigma\mu\nu} = \nabla_\lambda R_{\sigma\nu}$ together with the antisymmetry $R_{\rho\sigma\mu\nu} = -R_{\sigma\rho\mu\nu}$ (which turns the middle term's contraction into $-\nabla_\mu R^\mu{}_{\sigma\nu\lambda}$ reindexed):

$$\nabla_\lambda R_{\sigma\nu} - \nabla_\mu R^\mu{}_{\sigma\nu\lambda} - \nabla_\nu R_{\sigma\lambda} \;=\; 0. \tag{3.18}$$

(The signs come from the index antisymmetries of $R$; the middle term is the contraction of $\nabla_\mu$ on the raised first index.) Now contract a second time, $\sigma$ with $\nu$, using $g^{\sigma\nu}$ and $g^{\sigma\nu} R_{\sigma\nu} = R$:

$$\nabla_\lambda R - \nabla_\mu R^\mu{}_\lambda - \nabla_\nu R^\nu{}_\lambda \;=\; \nabla_\lambda R - 2\nabla_\mu R^\mu{}_\lambda \;=\; 0. \tag{3.19}$$

Divide by 2 and rearrange: $\nabla_\mu R^\mu{}_\lambda = \frac{1}{2}\nabla_\lambda R$. Finally take the divergence of the Einstein tensor (3.15), raising an index and using $\nabla_\mu g^\mu{}_\lambda = 0$ (the metric is covariantly constant, which is what (3.12) was built to ensure):

$$\nabla_\mu G^\mu{}_\lambda = \nabla_\mu R^\mu{}_\lambda - \tfrac{1}{2}\,\nabla_\lambda R = \tfrac{1}{2}\nabla_\lambda R - \tfrac{1}{2}\nabla_\lambda R = 0.$$

The two halves cancel exactly. That exact cancellation — not an approximation, not a physical law, but an algebraic identity forced by the symmetries of $R$ — is (3.16).

The factor of $\frac{1}{2}$ in the definition (3.15) of $G_{\mu\nu}$ is now revealed as no accident: it is the unique coefficient that makes the two terms in the last line cancel. The Einstein tensor is, so to speak, *engineered* to be divergence-free. [PROVEN] — and with it the structural argument of Chapter 1 is complete: anything set equal to $G_{\mu\nu}$ inherits $\nabla^\mu(\cdot) = 0$ and therefore cannot jump, which is exactly the noose that caught semiclassical gravity's collapsing source.

**Q 3.4 *"You proved $\nabla^\mu G_{\mu\nu} = 0$ is a pure identity. Then why is energy conservation ($\nabla^\mu T_{\mu\nu} = 0$) said to be a physical law? Aren't they the same equation once $G = 8\pi G\, T/c^4$?"***

Exactly the same equation — and that coincidence is one of the most beautiful facts in physics. On the geometry side, $\nabla^\mu G_{\mu\nu} = 0$ is an identity, true for every metric, proved above with no physics input. On the matter side, $\nabla^\mu T_{\mu\nu} = 0$ is local energy–momentum conservation, a physical law. Einstein's equation $G_{\mu\nu} = (8\pi G/c^4) T_{\mu\nu}$ *welds* the two: the geometric identity *forces* the matter to conserve energy–momentum. You cannot write down a source that violates conservation and still solve the field equations, because the left side is mathematically incapable of matching it. Gravity does not merely respond to energy; its very structure enforces that energy is conserved. This is also why, in Chapter 1, a collapsing wavefunction (which teleports energy) is incompatible with a classical geometry: the geometry's identity refuses to carry a non-conserved source.

## 3.5 Einstein's equations

We have the left-hand side. The right-hand side is the stress–energy tensor $T_{\mu\nu}$ — the density and flux of energy and momentum, with $T_{00}$ the energy density, $T_{0i}$ the momentum density, $T_{ij}$ the stresses — and it obeys $\nabla^\mu T_{\mu\nu} = 0$ (conservation) as a physical law. Einstein's field equations equate the two:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} \;=\; \frac{8\pi G}{c^4}\, T_{\mu\nu}. \tag{3.20}$$

Matter tells spacetime how to curve (right to left); the curvature tells matter how to move, through the geodesic equation (3.10). The term $\Lambda g_{\mu\nu}$ is the *cosmological constant*, the unique other divergence-free tensor available (since $\nabla_\mu g^\mu{}_\nu = 0$); it is a constant energy density of the vacuum itself, negligible in the laboratory and dominant in cosmology (Chapter 14). We set it aside until then.

### 3.5.1 Why this equation, and no other

Einstein's equations are often *quoted*; let us instead see why the form (3.20) is essentially forced. We want an equation "curvature = matter." Demand of the left-hand side four things, each physically motivated:

- it is a symmetric two-index tensor (to match $T_{\mu\nu}$);
- it is built from the metric and its first two derivatives (so that, like every other field equation in physics, it is second order — fixing the metric requires initial data on $g$ and $\dot{g}$, no more);
- it is *divergence-free*, $\nabla^\mu(\cdot) = 0$ (so that it can equal the conserved $T_{\mu\nu}$, per the reader question above);
- it is linear in the second derivatives (locality and the superposition of weak fields).

A theorem of Lovelock (1971) says that *in four spacetime dimensions the only tensor meeting all four demands* is

$$a\, G_{\mu\nu} + b\, g_{\mu\nu}, \tag{3.21}$$

for constants $a, b$. [PROVEN] There is no freedom left: the left side *must* be a multiple of the Einstein tensor plus a multiple of the metric. The constant $a$ is fixed to $8\pi G/c^4$ by demanding the Newtonian limit below; $b = \Lambda$ is the cosmological constant. Einstein's equations are not one choice among many — they are the only equation of their kind that four in-principle requirements permit.

### 3.5.2 Reduction to Newton

**Worked Example 3.4 — Einstein → Poisson in the weak, slow limit**

Take the weak-field static metric with $g_{00} = -(1 + 2\Phi/c^2)$ (equation (3.8)) and non-relativistic matter, for which the energy density dominates: $T_{00} \approx \rho c^2$ (rest energy), all other components negligible. It is cleanest to use the trace-reversed form of (3.20) (with $\Lambda = 0$), obtained by taking its trace, $R = -(8\pi G/c^4)T$ with $T = g^{\mu\nu}T_{\mu\nu}$, and substituting back:

$$R_{\mu\nu} = \frac{8\pi G}{c^4}\Big(T_{\mu\nu} - \tfrac{1}{2} T g_{\mu\nu}\Big). \tag{3.22}$$

For dust, $T = g^{00}T_{00} \approx -\rho c^2$, so the 00 component of the right side is $(8\pi G/c^4)(\rho c^2 - \frac{1}{2}(-\rho c^2)(-1)) = (8\pi G/c^4)\cdot \frac{1}{2}\rho c^2 = 4\pi G\rho/c^2$. For the left side we need $R_{00}$ in the weak static field. From the definition (3.14)–(3.13), keeping only terms linear in $\Phi$ and dropping $\Gamma^2$ and time derivatives, $R_{00} \approx \partial_i \Gamma^i{}_{00} = \partial_i(\frac{1}{c^2}\partial_i\Phi) = \frac{1}{c^2}\nabla^2\Phi$ (using $\Gamma^i{}_{00} = \partial_i\Phi/c^2$ from Worked Example 3.3). Equating the two:

$$\frac{1}{c^2}\nabla^2\Phi = \frac{4\pi G\rho}{c^2} \qquad \Longrightarrow \qquad \nabla^2\Phi = 4\pi G\rho.$$

This is exactly Newton's field equation (the Poisson equation) that Chapter 1 used to compute the field of a superposed mass. The constant $8\pi G/c^4$ in (3.20) was chosen precisely so that this reduction lands on $4\pi G\rho$ with no stray factors. Einstein contains Newton exactly, in the weak-slow corner.

With the toolkit assembled — metric, geodesic, curvature, field equation — we now build the four specific geometries the rest of the book requires. Each carries a *horizon*, and horizons are where quantum information and geometry will meet.

## 3.6 Rindler space: the horizon of an accelerated observer

The simplest horizon in physics belongs not to a black hole but to a uniformly accelerated observer in ordinary flat spacetime. This construction is load-bearing: it is the arena of the Unruh effect (Chapter 4) and of the Bisognano–Wichmann theorem (Chapter 7) that underlies the whole correspondence.

### 3.6.1 The uniformly accelerated worldline

Let an observer move along the $x$-axis of Minkowski space with constant *proper* acceleration $a$ — meaning the acceleration she feels, the reading of an accelerometer she carries, is the constant $a$ (a rocket burning at a fixed thrust). This is *not* constant coordinate acceleration $d^2x/dt^2$, which would drive her past $c$; relativistically the worldline is a hyperbola. Solving $du^\mu/d\tau$ constant in magnitude $a$ with $u^\mu u_\mu = -c^2$ gives

$$ct(\tau) = \frac{c^2}{a}\sinh\frac{a\tau}{c}, \qquad x(\tau) = \frac{c^2}{a}\cosh\frac{a\tau}{c}, \tag{3.23}$$

with $\tau$ the observer's proper time. Squaring and subtracting, $x^2 - c^2t^2 = (c^2/a)^2$: a hyperbola asymptotic to the light-cone lines $x = \pm ct$. The physical reading: no matter how long she accelerates, her worldline hugs but never crosses the line $x = ct$.

### 3.6.2 Rindler coordinates and the horizon

Adapt coordinates to a whole family of such observers, one for each acceleration. Label each hyperbola by a coordinate $\rho$ and use the proper time along it (rescaled) as a time coordinate $\eta$:

$$ct = \rho\sinh\frac{a\eta}{c}, \qquad x = \rho\cosh\frac{a\eta}{c}, \tag{3.24}$$

so the observer of (3.23) sits at fixed $\rho = c^2/a$. Compute the flat metric $ds^2 = -c^2dt^2 + dx^2$ in these coordinates by differentiating (3.24) ($c\,dt = \sinh(a\eta/c)\,d\rho + (a\rho/c)\cosh(a\eta/c)\,d\eta$, similarly for $dx$, and combine):

$$ds^2 \;=\; -\left(\frac{a\rho}{c^2}\right)^2 c^2\,d\eta^2 + d\rho^2 \;+\; dy^2 + dz^2. \tag{3.25}$$

This is the *Rindler metric*, and it describes exactly the same flat spacetime — just as seen by accelerated observers. It is static (nothing depends on the time $\eta$), and its clocks run at position-dependent rates: the factor $g_{\eta\eta} = -(a\rho/c)^2$ shows that clocks at larger $\rho$ ("higher up" in the direction of acceleration) tick faster, which is the gravitational redshift of §3.2 reappearing, since by the equivalence principle acceleration *is* a gravitational field.

The decisive feature is at $\rho \to 0$. There $g_{\eta\eta} \to 0$: a clock sitting at $\rho = 0$ does not tick at all, and the surface $\rho = 0$ (which in Minkowski coordinates (3.24) is $x = \pm ct$, the light-cone through the origin) is a *horizon*. Signals emitted from beyond it ($x < ct$) never reach the accelerated observer: to catch her they would have to outrun the light ray $x = ct$ that she forever chases but never overtakes.

> **The Rindler horizon.** An eternally accelerated observer can receive signals only from the wedge $x > |ct|$ (the "right Rindler wedge"). The null surface $x = ct$ is her *horizon*: everything behind it is causally hidden from her, permanently, even though the spacetime is perfectly flat and empty. Acceleration alone manufactures a hidden region.

### 3.6.3 Surface gravity and a temperature preview

The rate at which the redshift factor $g_{\eta\eta}$ falls to zero at the horizon defines the *surface gravity* $\kappa$. For the Rindler metric (3.25), writing the norm of the timelike Killing vector $\partial_\eta$ as $V = \sqrt{-g_{\eta\eta}} = a\rho/c$ and computing $\kappa =$ (the acceleration of the redshift factor as measured at the horizon), one finds

$$\kappa \;=\; a \qquad \text{(for the reference observer at } \rho = c^2/a\text{).} \tag{3.26}$$

The surface gravity of the Rindler horizon is just the observer's proper acceleration. The reason this is load-bearing rather than a curiosity: in Chapter 4 we will find that a detector following this worldline registers a thermal bath at temperature

$$T \;=\; \frac{\hbar\kappa}{2\pi c\, k_{\mathrm{B}}} \;=\; \frac{\hbar a}{2\pi c\, k_{\mathrm{B}}} \qquad \text{(Unruh; derived in Ch. 4).} \tag{3.27}$$

The horizon of §3.6 is not merely a causal boundary — it radiates, and the appearance of $\kappa$ in a temperature is the first thread of the entire quantum–geometric correspondence.

**Q 3.5** ***"The spacetime is empty and flat. How can there be a horizon and a temperature? Surely flat space has no gravity and no thermodynamics."***

Both facts are real and both are about the *observer*, not the spacetime. The horizon is relative: to an inertial observer (constant velocity) there is no horizon at all — she eventually receives every signal. Only the eternally accelerating observer, forever chasing a light ray, walls off a region she can never hear from. Flat spacetime has no *curvature* ($R = 0$ everywhere), but it can still have observer-dependent *horizons*, because a horizon is a causal notion (what can reach whom) not a curvature notion. The temperature is likewise observer-dependent, and its origin is that the accelerated observer's notion of "vacuum" differs from the inertial one — the flat vacuum, restricted to her wedge, is a thermal (mixed) state. That the state looks thermal is the content of the Bisognano–Wichmann theorem (Chapter 7), and it is the technical heart of why an area, a horizon, and an entropy always travel together in this book.

## 3.7 The Schwarzschild horizon: gravity that traps light

We now leave flat space for the field of a single spherical mass $M$. Solving the vacuum Einstein equations $G_{\mu\nu} = 0$ (empty space outside the mass) with spherical symmetry gives a unique answer, found by Schwarzschild in 1916 within months of Einstein's field equations. We quote the solution and then read its physics; deriving it from (3.20) is a standard but lengthy exercise we do not need in full.

$$ds^2 = -\Big(1 - \frac{r_s}{r}\Big)c^2dt^2 + \Big(1 - \frac{r_s}{r}\Big)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \qquad r_s \equiv \frac{2GM}{c^2}. \tag{3.28}$$

The single length $r_s$, the *Schwarzschild radius*, controls everything. Far away ($r \gg r_s$) the factor $1 - r_s/r \to 1$ and the metric becomes flat Minkowski in spherical coordinates: a distant observer sees no gravity, as she should. Closer in, $g_{00} = -(1 - r_s/r)$ differs from $-1$, and comparing with the Newtonian $g_{00} = -(1 + 2\Phi/c^2)$ of (3.8) identifies $\Phi = -GM/r$, the correct Newtonian potential — the strong-field metric reduces to Newton in its outer reaches, exactly as required.

The drama is at $r = r_s$. There $g_{00} = -(1 - r_s/r) \to 0$: by the redshift law (3.5) a clock lowered to $r_s$ ticks infinitely slowly as seen from far away, and light climbing out of $r_s$ is redshifted to zero frequency — infinite redshift. The surface $r = r_s$ is the *event horizon*: light emitted just inside it cannot escape to infinity, because to do so it would have to climb out of a well so deep that its frequency, and with it its energy, would be redshifted away entirely. Anything crossing inward can never return; it is a one-way membrane. (The apparent blow-up of $g_{rr} = (1 - r_s/r)^{-1}$ at $r_s$ is a coordinate artifact, removable by better coordinates; the genuine, coordinate-invariant singularity where $R^\rho{}_{\sigma\mu\nu}R_\rho{}^{\sigma\mu\nu}$ diverges sits at $r = 0$, hidden behind the horizon.)

**Worked Example 3.5 — The Schwarzschild radius of the Sun and of one kilogram**

The Schwarzschild radius is $r_s = 2GM/c^2$. For the Sun, $M_\odot = 1.989 \times 10^{30}$ kg:

$$r_s(M_\odot) = \frac{2\,(6.67430 \times 10^{-11})(1.989 \times 10^{30})}{(2.99792458 \times 10^8)^2} = \frac{2.655 \times 10^{20}}{8.988 \times 10^{16}} = 2.95 \times 10^3\ \text{m} \approx 2.95\ \text{km}. \tag{3.29}$$

If the entire Sun were compressed inside a sphere of radius 3 km it would become a black hole; as it is, $r_s$ sits far inside the Sun's actual 700,000 km radius, so no horizon forms and the external field is the ordinary weak-field one. For a 1 kg mass,

$$r_s(1\,\text{kg}) = \frac{2\,(6.67430 \times 10^{-11})(1)}{8.988 \times 10^{16}} = 1.48 \times 10^{-27}\ \text{m}. \tag{3.30}$$

A kilogram's Schwarzschild radius is twelve orders of magnitude smaller than a proton: everyday masses are nowhere near their horizons, which is why gravity is feeble and geometry near-flat in the laboratory. The horizon is a genuine feature of the equations, but forming one requires compressing mass to absurd density — another reminder that the quantum–gravity seam of this book lives far from any black hole.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch3 Schwarzschild radii)`

The Schwarzschild horizon shares with the Rindler horizon the two features that matter for Part III: it is a null surface hiding a region, and it carries a surface gravity $\kappa = c^4/(4GM)$

(computed in Chapter 5) whose associated Hawking temperature $T = \hbar\kappa/2\pi c k_{\mathrm{B}}$ is the exact analogue of the Unruh temperature (3.27). Every horizon in this book — Rindler, Schwarzschild, and the cosmological one below — tells the same story: a causal boundary, a surface gravity, a temperature, and (we will see in Chapter 5) an entropy proportional to its area.

## 3.8 FRW cosmology: the expanding universe

The last geometry is the universe as a whole. We build it thoroughly, because Chapter 14 — where the same three axioms yield the MOND acceleration scale $a_0 = cH_0/2\pi$ — introduces *no* new geometry beyond what we assemble here.

### 3.8.1 The homogeneous, isotropic metric

On the largest scales the universe looks the same everywhere (homogeneous) and the same in every direction (isotropic) — the *cosmological principle*, an observational input confirmed by galaxy surveys and the microwave background. The most general metric consistent with these two symmetries is the Friedmann–Robertson–Walker (FRW) metric. Restricting to the spatially flat case (which observation favours and which we use throughout):

$$ds^2 \;=\; -c^2 dt^2 + a(t)^2 (dx^2 + dy^2 + dz^2). \tag{3.31}$$

Here $t$ is cosmic time (read by clocks comoving with the matter), the spatial coordinates $(x, y, z)$ are *comoving* labels that ride along with the expansion, and the single function $a(t)$ — the *scale factor* — stretches all physical distances uniformly. Two galaxies at fixed comoving separation $\Delta x$ are at physical distance $a(t)\Delta x$, which grows as $a$ grows: space itself expands, and the galaxies recede without moving through it.

The expansion rate is measured by the *Hubble rate*

$$H(t) \;\equiv\; \frac{\dot a}{a}, \tag{3.32}$$

the fractional growth of distances per unit time. Its physical meaning is Hubble's law: a galaxy at physical distance $D = a\,\Delta x$ recedes at velocity $\dot D = \dot a\,\Delta x = HD$ — recession speed proportional to distance, the observation that founded modern cosmology. Today $H_0 \approx 2.2 \times 10^{-18}\,\mathrm{s}^{-1}$ (about $70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$).

### 3.8.2 The Friedmann equation

Feed the FRW metric (3.31) into Einstein's equations (3.20) with a perfect-fluid source (energy density $\rho c^2$, isotropic pressure $p$), and the 00 component collapses to a single ordinary differential equation for $a(t)$ — the *Friedmann equation*:

$$H^2 \;=\; \left(\frac{\dot a}{a}\right)^2 \;=\; \frac{8\pi G}{3}\rho \;+\; \frac{\Lambda c^2}{3}. \tag{3.33}$$

The expansion rate squared is set by the energy density of matter plus the cosmological constant. Its physical reading is a cosmic energy balance: the "kinetic" expansion $H^2$ is driven by the "potential" $\rho$ and $\Lambda$. Different contents give different histories: matter ($\rho \propto a^{-3}$, diluting as volume grows) yields decelerating expansion; the constant $\Lambda$ does not dilute and comes to dominate, which is the fate of our universe and the subject of the next subsection. [PROVEN] (standard consequence of (3.20)).

### 3.8.3 de Sitter space and the cosmological horizon

Suppose the energy density is dominated by the cosmological constant, so the right side of (3.33) is a constant: $H^2 = \Lambda c^2/3 \equiv \text{const}$. Then $\dot{a}/a = H$ is constant and integrates to

$$a(t) \;=\; a_0\, e^{Ht}, \tag{3.34}$$

*exponential* expansion. This is *de Sitter space*, the maximally symmetric solution with constant positive curvature, and it is the geometry of both the inflationary early universe and the accelerating late universe we inhabit. Physically: distances double every $\ln 2/H$ of time, forever.

De Sitter space has a horizon, and it is the cosmological cousin of Rindler's. Consider light emitted now from a distant galaxy toward us. Because space between us and the galaxy is stretching exponentially, sufficiently distant light can never arrive: the intervening space grows faster than the light can cross it. The boundary sits at the physical distance where recession speed $HD$ equals $c$:

$$D_H \;=\; \frac{c}{H} \qquad \text{(the de Sitter / Hubble horizon).} \tag{3.35}$$

Galaxies beyond $D_H = c/H$ recede faster than light (permissibly — it is space expanding, not motion through space) and are causally cut off from us. Like the Rindler and Schwarzschild horizons, the Hubble horizon is a null surface hiding a region, it carries a surface gravity, and — the punchline of Chapter 14 — an entropy and a temperature $T = \hbar H/2\pi k_B$. The single length $c/H$ is the cosmological analogue of $r_s$, and the acceleration scale it defines, $a_0 \sim cH$, is where the MOND phenomenology of Part IV comes from.

### 3.8.4 Conformal time

One coordinate change makes the causal structure — who can signal whom — transparent, and Chapter 14 uses it. Define *conformal time* $\eta$ by absorbing the scale factor into the time coordinate:

$$d\eta \;=\; \frac{dt}{a(t)} \qquad \Longrightarrow \qquad ds^2 \;=\; a(\eta)^2\big(-c^2 d\eta^2 + dx^2 + dy^2 + dz^2\big). \tag{3.36}$$

In these coordinates the metric is flat Minkowski times an overall factor $a(\eta)^2$ (a *conformal* factor). Because light rays satisfy $ds^2 = 0$ and the overall factor cannot change where $ds^2$ vanishes, null rays travel on 45° lines $\Delta x = \pm c\,\Delta\eta$ exactly as in flat space: conformal time trivialises the propagation of light through an expanding universe. For de Sitter, integrating (3.36) with (3.34) gives $\eta = -e^{-Ht}/(a_0 H)$, running over $(-\infty, 0)$: the horizon at $\eta \to 0^-$ is the exponential expansion seen as a finite conformal boundary. This is the entire geometric apparatus Chapter 14 needs; we introduce no more.

**Q 3.6 "Galaxies beyond $c/H$ recede faster than light. Doesn't that violate the rule that nothing outruns light?"**

No, and the distinction is exactly the one the equivalence principle taught us to make. Special relativity's speed limit is *local*: nothing can move through space, past you, faster than $c$ — and no galaxy does; each one sits essentially still relative to its own local space (its comoving coordinates are fixed). What grows is the *space between* galaxies, and there is no law limiting how fast space itself may stretch. The recession "velocity" $HD$ is not a velocity through spacetime; it is the rate of accumulation of new space, and it can exceed $c$ without any signal ever outrunning a light ray locally. This is why the Hubble horizon is a genuine causal boundary (distant light truly cannot reach us) without contradicting relativity — precisely as the Rindler horizon walled off a region in flat space without anything travelling faster than light. Horizons are about the global integration of light paths, not about local speeds.

## 3.9 The Raychaudhuri equation: gravity focuses light

We close with the single equation that Chapter 10 will use to turn entanglement into Einstein's equations. It governs how a bundle of neighbouring light rays — a *null congruence* — expands or contracts as it propagates, and it is the sharpest quantitative form of the statement "gravity attracts." We derive it in full.

### 3.9.1 Setup: the expansion of a ray bundle

Consider a family of light rays (null geodesics) filling a small region, each labelled by an affine parameter $\lambda$ along it, with tangent vector $k^\mu = dx^\mu/d\lambda$ satisfying $k^\mu k_\mu = 0$ (null) and $k^\nu \nabla_\nu k^\mu = 0$ (geodesic). Picture a small circular cross-section of the bundle, a bundle of parallel rays like a flashlight beam. As the rays propagate, this cross-section can change in three independent ways, encoded in the gradient $\nabla_\nu k_\mu$ of the ray field:

- its *area* can grow or shrink — the *expansion* $\theta$, the fractional rate of change of cross-sectional area, $\theta = \nabla_\mu k^\mu$;
- its *shape* can distort at fixed area (circle to ellipse) — the *shear* $\sigma_{\mu\nu}$, the symmetric trace-free part;
- it can *rotate* — the *twist* $\omega_{\mu\nu}$, the antisymmetric part.

The expansion $\theta$ is what we track: $\theta > 0$ means the beam is spreading (area growing), $\theta < 0$ means it is converging (rays focusing toward a point).

### 3.9.2 Derivation of the evolution equation

We compute how $\theta$ changes along the rays, $d\theta/d\lambda = k^\nu \nabla_\nu \theta$. Start from $\theta = \nabla_\mu k^\mu$ and differentiate along the ray:

$$\frac{d\theta}{d\lambda} = k^\nu \nabla_\nu \nabla_\mu k^\mu. \tag{3.37}$$

Swap the order of the two covariant derivatives using the definition of the Riemann tensor (3.13), which is precisely the machine that measures the failure of covariant derivatives to commute:

$\nabla_\nu \nabla_\mu k^\mu = \nabla_\mu \nabla_\nu k^\mu + R_\nu{}^\mu{}_{\dots}$; carrying the index bookkeeping,

$$k^\nu \nabla_\nu \nabla_\mu k^\mu = k^\nu \nabla_\mu \nabla_\nu k^\mu - R_{\mu\nu} k^\mu k^\nu. \tag{3.38}$$

(The Riemann tensor contracted on two indices became the Ricci tensor $R_{\mu\nu}$; the sign is fixed by the definition (3.14).) In the first term on the right, move $k^\nu$ inside the derivative and use the geodesic equation $k^\nu \nabla_\nu k^\mu = 0$ to drop one piece: $k^\nu \nabla_\mu \nabla_\nu k^\mu = \nabla_\mu (k^\nu \nabla_\nu k^\mu) - (\nabla_\mu k^\nu)(\nabla_\nu k^\mu) = -(\nabla_\mu k^\nu)(\nabla_\nu k^\mu)$. The surviving quadratic term is the "velocity gradient squared," and it decomposes into the three kinematic pieces named above. Writing the traceless-symmetric part as the shear scalar $\sigma^2 = \sigma_{\mu\nu}\sigma^{\mu\nu} \geq 0$ and the antisymmetric part as the twist $\omega^2 = \omega_{\mu\nu}\omega^{\mu\nu} \geq 0$, and separating the trace (which gives the $\theta^2$ term with a dimension-dependent coefficient — for a null congruence the transverse space is $d-2$ dimensional), we arrive at:

$$\frac{d\theta}{d\lambda} = -\frac{\theta^2}{d-2} - \sigma_{\mu\nu}\sigma^{\mu\nu} + \omega_{\mu\nu}\omega^{\mu\nu} - R_{\mu\nu}\, k^\mu k^\nu. \tag{3.39}$$

This is the *Raychaudhuri equation* for a null congruence (in $d$ spacetime dimensions; $d = 4$ gives the familiar $-\theta^2/2$). Every sign on the right is doing physical work.

### 3.9.3 What it says: gravity attracts, precisely

Read (3.39) term by term. The $-\theta^2/(d-2)$ term is self-focusing: once a beam is converging ($\theta < 0$), this term makes $d\theta/d\lambda$ more negative still — convergence accelerates, independent of any matter. The shear term $-\sigma^2 \leq 0$ also drives convergence: distorting a beam focuses it. The twist term $+\omega^2 \geq 0$ opposes focusing — rotation resists collapse, the centrifugal effect — but for a bundle emanating from a point (or any hypersurface-orthogonal congruence) the twist vanishes, $\omega_{\mu\nu} = 0$, by a theorem (Frobenius).

That leaves the last term, the only one carrying the geometry's response to matter: $-R_{\mu\nu}k^\mu k^\nu$. Through Einstein's equations (3.20), $R_{\mu\nu}k^\mu k^\nu$ is controlled by $T_{\mu\nu}k^\mu k^\nu$ — the energy density that matter presents to a light ray. For all ordinary matter this is non-negative (the *null energy condition*, $T_{\mu\nu}k^\mu k^\nu \geq 0$, satisfied by every classical field we know), which makes $-R_{\mu\nu}k^\mu k^\nu \leq 0$: matter always *focuses* light, never defocuses it.

> **Gravity focuses light.** With vanishing twist and ordinary matter, *every* term on the right of (3.39) is $\leq 0$: $d\theta/d\lambda \leq 0$. A bundle of light rays, once it starts converging, is driven inexorably to a focus (a caustic, $\theta \to -\infty$) in finite affine parameter. This is the mathematical engine behind gravitational lensing, the Penrose singularity theorems, and — in Chapter 10 — the derivation of Einstein's equations themselves from the requirement that a horizon's area (i.e. its entanglement entropy) obey the second law.

The reason this specific equation earns a section of its own: in Chapter 10, the term $R_{\mu\nu}k^\mu k^\nu$ will be equated, through the first law of entanglement thermodynamics, to a variation of entropy across a local Rindler horizon, and demanding that the equality hold for *every* such horizon (every observer, every acceleration) will force $R_{\mu\nu}$ to take the exact form of Einstein's equations. The Raychaudhuri equation is the bridge from "how light bends" to "why spacetime obeys (3.20)." It is, quietly, the most important equation in the book after the correspondence itself.

**Q 3.7** ***"You said twist $\omega^2$ appears with a plus sign and opposes focusing — but then dropped it. When am I allowed to drop it, and does the theory ever need the case where I can't?"***

You may drop the twist exactly when the congruence is *hypersurface-orthogonal* — when the rays are everywhere perpendicular to a family of surfaces, like the wavefronts of a beam emitted from a two-surface. Frobenius's theorem guarantees $\omega_{\mu\nu} = 0$ in that case, and it is precisely the case that arises at horizons: the null generators of a Rindler, Schwarzschild, or cosmological horizon are hypersurface-orthogonal by construction, so the twist genuinely vanishes there and the focusing theorem applies cleanly. This is why Chapter 10 can use the twist-free Raychaudhuri equation without apology — it works on exactly the local horizons the derivation of Einstein's equations is built from. Rotating congruences (nonzero $\omega$) matter for spinning black holes and fluid vorticity, but not for the entanglement-equilibrium argument, which lives on instantaneously non-rotating local horizons.

## Chapter FAQ

**Q: Is the equivalence principle exactly true, or only approximately?** The version we used — gravity is locally indistinguishable from acceleration — is exact *in the limit of a small enough laboratory*, and this is not a defect but the definition of "local." Over any finite region, tidal effects (curvature, $R^\rho{}_{\sigma\mu\nu} \neq 0$) betray real gravity and break the equivalence with a uniformly accelerated frame. So the principle is exactly a statement that gravity's *first-derivative* part (the $\Gamma$'s) is removable and its *second-derivative* part (the $R$'s) is not. Everything in the chapter is that one sentence unfolded.

**Q: Why the signature $(-,+,+,+)$ and not $(+,-,-,-)$?** Pure convention; nothing physical depends on it. With $(-,+,+,+)$ a timelike interval has $ds^2 < 0$ and proper time is $d\tau = \sqrt{-ds^2}/c$; spatial distances are positive, which is convenient for the geometry-first viewpoint of this book. The opposite convention flips signs on $\eta_{\mu\nu}$, $g_{\mu\nu}$, and many intermediate quantities but leaves every measurable — proper times, redshifts, curvatures — unchanged. We fix $(-,+,+,+)$ once and never revisit it.

**Q: The Schwarzschild metric blows up at $r = r_s$. Isn't the horizon a singularity where physics breaks down?** No — that blow-up is a fault of the coordinates, not the spacetime. The curvature invariant $R^\rho{}_{\sigma\mu\nu}R_\rho{}^{\sigma\mu\nu}$ is perfectly finite at $r = r_s$ (it goes as $r_s^2/r^6$, smooth there) and diverges only at $r = 0$. An observer falling through the horizon notices nothing locally special — no infinite tide, no wall — which is exactly what the equivalence principle predicts, since the horizon is a null surface she crosses freely. Better coordinates (Eddington–Finkelstein, Kruskal) make the metric manifestly smooth at $r_s$. The horizon is a global causal boundary, not a local catastrophe.

**Q: If the universe is expanding, is the Earth (or an atom, or a ruler) expanding too?** No. The expansion (3.31) applies to the average cosmic fluid on the largest scales, where matter is smoothly spread. Bound systems — atoms held by electromagnetism, the solar system held by gravity, galaxies held by their own gravity — decoupled from the Hubble flow long ago and do not expand: their internal forces vastly overwhelm the cosmological "stretching," which is not even a force so much as the absence of one. Only the comoving distances between unbound, freely falling galaxies grow with $a(t)$. Your ruler is safe; the cosmological principle is a statement about the coarse-grained universe, not about anything you can hold.

## Exercises

**Exercise 3.1.** A GPS satellite orbits at altitude $h = 2.0 \times 10^7$ m above the Earth's surface (Earth mass $M_\oplus = 5.97 \times 10^{24}$ kg, radius $R_\oplus = 6.37 \times 10^6$ m). (a) Using the gravitational redshift law (3.4) with the Newtonian potential $\Phi = -GM_\oplus/r$, compute the fractional rate difference $\Delta\nu/\nu$ between a clock on the satellite and a clock on the ground due to gravity alone (ignore the satellite's orbital motion). (b) Over one day, how many nanoseconds does the satellite clock gain on the ground clock? (c) Multiply by $c$: how many metres of positioning error per day would result if this effect were not corrected? Comment on why general relativity is a practical engineering necessity.

**Exercise 3.2.** An observer accelerates with proper acceleration $a = 9.81\,\mathrm{m/s^2}$ (one Earth gravity). (a) Using the Unruh preview (3.27), $T = \hbar a/2\pi c k_\mathrm{B}$ with $k_\mathrm{B} = 1.381 \times 10^{-23}\,\mathrm{J/K}$, compute the temperature of the Rindler bath she experiences. (b) What acceleration would be needed to reach $T = 1$ K? Express it in units of $g = 9.81\,\mathrm{m/s^2}$. (c) Comment on why the Unruh effect, though a firm prediction, is so hard to observe directly.

## Solutions

*Solution (Exercise 3.1).* (a) The potential at radius $r$ is $\Phi(r) = -GM_\oplus/r$. Ground clock at $r_1 = R_\oplus = 6.37 \times 10^6$ m, satellite at $r_2 = R_\oplus + h = 2.637 \times 10^7$ m. The satellite is *higher* in the well (less negative $\Phi$), so its clock runs *faster*; the fractional rate difference is

$$\frac{\Delta\nu}{\nu} = \frac{\Phi(r_2) - \Phi(r_1)}{c^2} = \frac{GM_\oplus}{c^2}\Big(\frac{1}{r_1} - \frac{1}{r_2}\Big).$$

Compute $GM_\oplus/c^2 = (6.6743 \times 10^{-11})(5.97 \times 10^{24})/(8.988 \times 10^{16}) = 4.43 \times 10^{-3}$ m. Then $1/r_1 - 1/r_2 = 1.570 \times 10^{-7} - 3.792 \times 10^{-8} = 1.191 \times 10^{-7}\,\mathrm{m^{-1}}$, giving $\Delta\nu/\nu = (4.43 \times 10^{-3})(1.191 \times 10^{-7}) = 5.28 \times 10^{-10}$. (b) Over one day ($8.64 \times 10^4$ s) the satellite clock gains $5.28 \times 10^{-10} \times 8.64 \times 10^4 = 4.56 \times 10^{-5}$ s $= 45.6\,\mu$s $= 4.56 \times 10^4$ ns. (c) Multiplying by $c$: $4.56 \times 10^{-5}\,\mathrm{s} \times 3 \times 10^8\,\mathrm{m/s} \approx 1.4 \times 10^4$ m — about 14 km of positioning error *per day*. Uncorrected, GPS would be useless within minutes. General relativity is not an esoteric refinement here; it is load-bearing infrastructure, the geometry of time (3.5) measured to the nanosecond by every phone.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch3 GPS redshift)`

*Solution (Exercise 3.2).* (a) With $a = 9.81\,\mathrm{m/s^2}$:

$$T = \frac{\hbar a}{2\pi c\, k_\mathrm{B}} = \frac{(1.0546 \times 10^{-34})(9.81)}{2\pi(2.9979 \times 10^8)(1.381 \times 10^{-23})} = \frac{1.034 \times 10^{-33}}{2.601 \times 10^{-14}} = 3.98 \times 10^{-20}\ \mathrm{K}.$$

Four hundredths of a zepto-kelvin: utterly negligible. (b) Setting $T = 1$ K and solving for $a$: $a = 2\pi c k_\mathrm{B} T/\hbar = 2.601 \times 10^{-14}/1.0546 \times 10^{-34} = 2.47 \times 10^{20}\,\mathrm{m/s^2}$, which is $a/g = 2.47 \times 10^{20}/9.81 = 2.5 \times 10^{19}$ Earth gravities. (c) The Unruh temperature is linear in acceleration with a fantastically small coefficient ($\hbar/2\pi c k_{\mathrm{B}} \approx 4\times10^{-21}$ K per m/s$^2$), so producing even a 1 K bath requires accelerations $\sim 10^{20}g$, far beyond any sustained laboratory capability. The effect is firmly predicted (Chapter 4 derives it from first principles) but its direct thermal signature is buried below every noise floor — which is why the theory of this book seeks the correspondence's consequences in *decoherence* (Part III) and *cosmology* (Part IV) rather than in a laboratory Unruh thermometer.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch3 Unruh temperature)`

**What to carry forward**

1. Gravity is geometry. The equivalence principle (free fall cancels gravity locally) forces gravitational redshift $\Delta\nu/\nu = -\Delta\Phi/c^2$ — gravity is the geometry of time — and promotes the flat metric $\eta_{\mu\nu}$ to a field $g_{\mu\nu}(x)$ whose geodesics (3.10) are free fall (extremal ageing) and whose weak-field $g_{00} = -(1+2\Phi/c^2)$ recovers Newton.
2. Curvature is the uncancellable part. The Riemann tensor (3.13) (commutator of covariant derivatives) is the true, frame-independent field; its contractions give $G_{\mu\nu}$, which satisfies the *identity* $\nabla^\mu G_{\mu\nu} = 0$ (proved — discharging Chapter 1's note) and, via Lovelock's uniqueness, forces Einstein's equations $G_{\mu\nu} + \Lambda g_{\mu\nu} = (8\pi G/c^4)T_{\mu\nu}$, which reduce to Poisson's $\nabla^2\Phi = 4\pi G\rho$.
3. Every horizon carries a surface gravity, a temperature, and (Part III) an entropy. Rindler ($\kappa = a$, flat space, hides the wedge $x < ct$), Schwarzschild ($r_s = 2GM/c^2$: 2.95 km for the Sun, $1.48\times10^{-27}$ m for a kilogram), and de Sitter (Hubble horizon $c/H$, exponential expansion) share one structure. The null Raychaudhuri equation (3.39) — $d\theta/d\lambda = -\theta^2/(d-2) - \sigma^2 + \omega^2 - R_{\mu\nu}k^\mu k^\nu$ — shows ordinary matter *focuses* light, and is the workhorse that Chapter 10 uses to derive Einstein's equations from entanglement equilibrium.

# 4 Quantum Fields and the Unruh Effect

**Synopsis.** A quantum field is nothing more exotic than an infinite collection of harmonic oscillators, one per mode, and everything this book needs from field theory we build from the single oscillator the reader already knows. Three tools come out of that construction, and all three are load-bearing downstream. *Coherent states* are the field states that look most classical; their overlap is a Gaussian we compute exactly, and it is — to the letter — the visibility law of Chapter 2 with the field playing the environment. *Bogoliubov transformations* relate two observers who disagree about which modes count as "positive frequency," and the disagreement is measured by a number of particles. Put the two together on the Rindler horizon of Chapter 3 and the Minkowski vacuum turns thermal: an accelerated observer reads a temperature $T = \hbar a/2\pi c k_{\mathrm{B}}$. That factor of $2\pi$ is not decoration — it is the periodicity of Euclidean time, and it becomes the $\beta_{\mathrm{mod}} = 2\pi$ at the heart of the whole theory. We close by writing the gravitational field of a mass as a coherent state of gravitons: the precise object Chapter 11 decoheres.

**You will be able to.** quantize a free scalar field as a bath of oscillators and identify its vacuum and particles; construct coherent states, prove $a\left|\alpha\right\rangle = \alpha\left|\alpha\right\rangle$ and the overlap $|\left\langle\alpha|\beta\right\rangle|^2 = e^{-|\alpha-\beta|^2}$, and connect it to gravitational decoherence; carry out a Bogoliubov transformation and read off the particle content $\langle 0_a|N_b|0_a\rangle = |\beta|^2$; derive the Unruh temperature end to end and name the origin of its $2\pi$.

**Prerequisites.** Chapters 1–3. The harmonic-oscillator ladder operators $a, a^\dagger$ with $[a, a^\dagger] = 1$; Fourier transforms; special relativity. *No prior quantum field theory is assumed* — we build it here from the oscillator.

## 4.1 From one oscillator to a field

The reader knows the quantum harmonic oscillator: a particle in a potential $\frac{1}{2}m\omega^2x^2$, with Hamiltonian $H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2x^2$, solved once and for all by the ladder operators

$$a = \sqrt{\tfrac{m\omega}{2\hbar}}\Big(x + \tfrac{i}{m\omega}p\Big), \qquad a^\dagger = \sqrt{\tfrac{m\omega}{2\hbar}}\Big(x - \tfrac{i}{m\omega}p\Big), \qquad [a, a^\dagger] = \mathbb{1}, \tag{4.1}$$

in terms of which $H = \hbar\omega(a^\dagger a + \frac{1}{2})$. The number operator $N = a^\dagger a$ has eigenstates $\left|n\right\rangle$ with $N\left|n\right\rangle = n\left|n\right\rangle$; $a$ lowers the rung ($a\left|n\right\rangle = \sqrt{n}\left|n-1\right\rangle$), $a^\dagger$ raises it ($a^\dagger\left|n\right\rangle = \sqrt{n+1}\left|n+1\right\rangle$), and the ladder bottoms out at the ground state $\left|0\right\rangle$ with $a\left|0\right\rangle = 0$. This is the entire content of the oscillator, and it is the entire content of free quantum field theory too — repeated once for every mode. Let us make that precise.

### 4.1.1 A field is a bath of oscillators

Take the simplest relativistic field: a single real scalar $\phi(\vec{x}, t)$, the field-theory stand-in for "a number attached to every point of space," obeying the wave equation that a massless relativistic field must,

$$\Big(-\frac{1}{c^2}\partial_t^2 + \nabla^2\Big)\phi = 0. \tag{4.2}$$

(This is the massless Klein–Gordon equation; a mass term adds $-m^2c^2/\hbar^2$ inside the bracket and changes nothing in the logic below.) Operationally, $\phi$ is what a detector at $\vec{x}$ would read, and (4.2) says the readings propagate as waves at speed $c$. To quantize, we do the one thing we know how to do — diagonalize into oscillators — by Fourier-analysing the field into plane-wave modes. Write

$$\phi(\vec{x}, t) = \int \frac{d^3k}{(2\pi)^3}\,\frac{1}{\sqrt{2\omega_k}}\Big[a_{\vec{k}}\, e^{i(\vec{k}\cdot\vec{x}-\omega_k t)} + a_{\vec{k}}^\dagger\, e^{-i(\vec{k}\cdot\vec{x}-\omega_k t)}\Big], \qquad \omega_k = c|\vec{k}|. \tag{4.3}$$

Each wave-vector $\vec{k}$ labels one *mode* — a standing pattern $e^{i\vec{k}\cdot\vec{x}}$ oscillating in time at frequency $\omega_k = c|\vec{k}|$, which is exactly (4.2) demanding that spatial curvature $(-k^2)$ balance temporal curvature $(-\omega_k^2/c^2)$. The coefficient of each mode is not a number but an operator, $a_{\vec{k}}$ and its conjugate $a_{\vec{k}}^\dagger$, and the whole quantization amounts to declaring that these operators are ladder operators, one independent pair per mode:

$$[\, a_{\vec{k}},\, a_{\vec{k}'}^\dagger\,] = (2\pi)^3\delta^3(\vec{k}-\vec{k}')\,\mathbb{1}, \qquad [\, a_{\vec{k}},\, a_{\vec{k}'}\,] = 0. \tag{4.4}$$

The Dirac delta on the right is the continuum bookkeeping for "different modes are independent oscillators" — in a box of finite volume $V$ it would be an ordinary Kronecker delta, $[a_{\vec{k}}, a_{\vec{k}'}^\dagger] = \delta_{\vec{k}\vec{k}'}$, and the reader who prefers a discrete sum $\sum_{\vec{k}}$ over box modes loses nothing. Either way the message is one line: *a free field is a countable (or continuous) family of independent harmonic oscillators, indexed by wave-vector.* Everything from Chapter's-worth of oscillator algebra applies mode by mode.

### 4.1.2 The vacuum and its particles

Since each mode is an oscillator, the field Hamiltonian is a sum of oscillator Hamiltonians,

$$H = \int \frac{d^3k}{(2\pi)^3}\,\hbar\omega_k\,\Big(a_{\vec{k}}^\dagger a_{\vec{k}} + \tfrac{1}{2}\Big), \tag{4.5}$$

and its ground state is the state in which *every* mode is in its own oscillator ground state:

$$a_{\vec{k}}\,|0\rangle = 0 \quad \text{for all } \vec{k}. \tag{4.6}$$

This is the *vacuum* $|0\rangle$ — operationally, "no detector clicks," the state of lowest energy. It is emphatically not "nothing": every mode still carries its zero-point energy $\frac{1}{2}\hbar\omega_k$, and the field still fluctuates, $\langle 0|\phi^2|0\rangle \neq 0$. Those fluctuations are the substance of this chapter — the Unruh effect is what happens when one observer's "no clicks" is another observer's warm bath.

A *particle* is one rung up on one mode. Acting with $a_{\vec{k}}^\dagger$ on the vacuum,

$$|1_{\vec{k}}\rangle = a_{\vec{k}}^\dagger\,|0\rangle\,, \qquad H\,|1_{\vec{k}}\rangle = (E_0 + \hbar\omega_k)\,|1_{\vec{k}}\rangle\,, \tag{4.7}$$

adds one quantum of energy $\hbar\omega_k$ and momentum $\hbar\vec{k}$ to the field — which we call "a particle of momentum $\hbar\vec{k}$." Multiple $a^\dagger$'s build multi-particle states, and the number operator of a mode, $N_{\vec{k}} = a^\dagger_{\vec{k}} a_{\vec{k}}$, counts how many quanta occupy it. That is the whole dictionary: *modes are oscillators, excitations are particles, the vacuum is all oscillators in their ground state.* We never need more field theory than this; the physics of Part III lives entirely in which state the oscillators are in.

**Q 4.1** ***"You said "particle" means one rung up on a mode. But a mode $e^{i\vec{k}\cdot\vec{x}}$ fills all of space. In what sense is that a localized particle I could detect*** **here?"**

It is not localized — a single-mode state $|1_{\vec{k}}\rangle$ is a plane wave, spread over all space, with perfectly definite momentum and completely indefinite position, exactly as the uncertainty principle demands. A particle you can point at is a *wave packet*: a superposition $\int d^3k\, f(\vec{k})\, a^\dagger_{\vec{k}}\,|0\rangle$ of many modes, narrow in $\vec{k}$ so it has near-definite momentum, and correspondingly spread over a finite region so it has a location. The plane-wave modes are the convenient basis — the oscillators — not the physical particles, in the same way that $\sin(kx)$ is a convenient basis for a plucked string even though a real pluck is a localized bump built from many sin's. Nothing in what follows needs localized packets; the Unruh calculation is cleanest in the plane-wave basis, and the gravitational field of Chapter 11 is sourced mode by mode. But the honest answer to "where is the particle" is: wherever the wave packet is peaked, and a single mode is not a packet.

## 4.2 Coherent states: the classical face of a field

We now build the single most important tool in this book. Ask: which state of an oscillator behaves most like a *classical* oscillation — a definite amplitude and phase, ringing along a classical trajectory? Not the number states $|n\rangle$: those have completely random phase (the expectation $\langle n|x|n\rangle = 0$ for every $n$, so on average they sit still). We want a state whose $\langle x(t)\rangle$ traces a genuine classical sinusoid. The answer is the *coherent state*, and its properties are the properties of the gravitational field of a mass.

### 4.2.1 Definition and the eigenvalue property

Work with one mode; the field generalization is a product over modes, done in §4.2.3. Define the *displacement operator*

$$D(\alpha) = \exp(\alpha\, a^\dagger - \alpha^* a), \qquad \alpha \in \mathbb{C}, \tag{4.8}$$

and the coherent state as the vacuum displaced by it,

$$|\alpha\rangle = D(\alpha)\,|0\rangle\,. \tag{4.9}$$

The name is literal: $D(\alpha)$ shifts the oscillator's position and momentum by a $c$-number amount fixed by $\alpha$. To see it, use the operator identity (a special case of Baker–Campbell–Hausdorff, valid because $[a, a^\dagger] = \mathbb{1}$ commutes with everything)

$$D^\dagger(\alpha)\, a\, D(\alpha) = a + \alpha. \tag{4.10}$$

This is the defining action: conjugating the annihilation operator by $D(\alpha)$ adds the number $\alpha$. From it the central property follows in two lines. Act with $a$ on $|\alpha\rangle$:

$$a\,|\alpha\rangle = a\,D(\alpha)\,|0\rangle = D(\alpha)\,\underbrace{\left[D^\dagger(\alpha) a D(\alpha)\right]}_{a+\alpha}|0\rangle = D(\alpha)\,(a+\alpha)\,|0\rangle = \alpha\,D(\alpha)\,|0\rangle = \alpha\,|\alpha\rangle\,, \quad (4.11)$$

using $a\,|0\rangle = 0$ in the penultimate step. So a coherent state is an *eigenstate of the annihilation operator*:

$$a\,|\alpha\rangle = \alpha\,|\alpha\rangle\,. \quad (4.12)$$

The physical meaning is immediate: since $x$ and $p$ are built from $a + a^\dagger$ and $a - a^\dagger$ (invert (4.1)), the state $|\alpha\rangle$ has definite *average* position and momentum fixed by the real and imaginary parts of $\alpha$, and evolving in time it traces the classical ellipse in phase space. The coherent state is the quantum field's best imitation of a classical wave.

**Q 4.2 *"Why should the classical-looking state be an eigenstate of $a$ — an operator that isn't even Hermitian, so $\alpha$ isn't a measurable eigenvalue?"***

Because $a$ is what a classical wave measurement effectively picks out. A laser beam, a radio field, the Newtonian gravitational field of a mass — when you measure a classical field you are measuring its *amplitude and phase together*, which is precisely the (complex) combination $a$ packages, not the (Hermitian, one-at-a-time) $x$ or $p$ separately. The eigenvalue $\alpha$ being complex is the statement that a classical wave has both an amplitude ($|\alpha|$) and a phase ($\arg\alpha$). And there is a deeper payoff we cash in shortly: because $a\,|\alpha\rangle = \alpha\,|\alpha\rangle$ with a single number, removing one quantum from a coherent state leaves it *unchanged* up to normalization — which is why a coherent field is robust, macroscopically stable, and why the gravitational field of a grain does not care that you occasionally probe it with a test mass. It is the most classical thing a quantum field can be, and §4.2.4 makes "most classical" quantitative.

### 4.2.2 The overlap formula — the formula the book runs on

Everything in Part III hinges on *how distinguishable two coherent states are*, so we compute their overlap $\langle\alpha|\beta\rangle$ exactly. The cleanest route expands $|\alpha\rangle$ in number states. Apply the eigenvalue equation (4.12) to the ansatz $|\alpha\rangle = \sum_n c_n\,|n\rangle$: since $a\,|n\rangle = \sqrt{n}\,|n-1\rangle$, matching coefficients gives the recursion $c_n = \alpha\,c_{n-1}/\sqrt{n}$, hence $c_n = c_0\,\alpha^n/\sqrt{n!}$. Normalizing $\sum_n |c_n|^2 = |c_0|^2 \sum_n |\alpha|^{2n}/n! = |c_0|^2 e^{|\alpha|^2} = 1$ fixes $c_0 = e^{-|\alpha|^2/2}$, so

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\,|n\rangle\,. \quad (4.13)$$

(Two facts fall out for free: the occupation is Poissonian, $|c_n|^2 = e^{-|\alpha|^2}|\alpha|^{2n}/n!$, so $\langle N\rangle = |\alpha|^2$ — a coherent state contains $|\alpha|^2$ quanta on average.) Now the overlap of two coherent states is a single sum:

$$\langle\alpha|\beta\rangle = e^{-|\alpha|^2/2}e^{-|\beta|^2/2}\sum_{m,n}\frac{(\alpha^*)^m\beta^n}{\sqrt{m!\,n!}}\underbrace{\langle m|n\rangle}_{\delta_{mn}} = e^{-|\alpha|^2/2-|\beta|^2/2}\sum_n\frac{(\alpha^*\beta)^n}{n!}$$
$$= \exp\Big(-\tfrac{1}{2}|\alpha|^2 - \tfrac{1}{2}|\beta|^2 + \alpha^*\beta\Big). \tag{4.14}$$

The last sum is just $e^{\alpha^*\beta}$. This is the coherent-state overlap — exact, no approximation. Its *modulus squared* is what a decoherence calculation needs, and it simplifies beautifully. Write $-\frac{1}{2}|\alpha|^2 - \frac{1}{2}|\beta|^2 + \alpha^*\beta$ and add its complex conjugate:

$$|\langle\alpha|\beta\rangle|^2 = \exp(-|\alpha|^2 - |\beta|^2 + \alpha^*\beta + \alpha\beta^*) = \exp(-|\alpha-\beta|^2), \tag{4.15}$$

since $|\alpha-\beta|^2 = |\alpha|^2 + |\beta|^2 - \alpha^*\beta - \alpha\beta^*$. Read it plainly: *two coherent states overlap by a Gaussian in their separation in the complex plane.* Nearby amplitudes ($|\alpha-\beta| \ll 1$) are nearly identical; well-separated ones ($|\alpha-\beta| \gg 1$) are effectively orthogonal. There is no sharp boundary — coherent states are never exactly orthogonal for finite separation — but the transition is a Gaussian cliff.

### 4.2.3 Multimode fields and the visibility law

A field has one oscillator per mode, so its coherent state is a product, $|\{\alpha_{\vec{k}}\}\rangle = \bigotimes_{\vec{k}}|\alpha_{\vec{k}}\rangle$, and because the modes are independent the overlap factorizes exactly as the environment records did in Chapter 2:

$$|\langle\{\alpha_{\vec{k}}\}\,|\,\{\beta_{\vec{k}}\}\rangle|^2 = \prod_{\vec{k}}|\langle\alpha_{\vec{k}}|\beta_{\vec{k}}\rangle|^2 = \exp\Big(-\sum_{\vec{k}}|\alpha_{\vec{k}}-\beta_{\vec{k}}|^2\Big) \xrightarrow{\text{continuum}} \exp\Big(-\int\frac{d^3k}{(2\pi)^3}\,|\alpha_{\vec{k}}-\beta_{\vec{k}}|^2\Big). \tag{4.16}$$

This equation is the reason the chapter exists. Set it beside the visibility law of Chapter 2, $\mathcal{V} = |\langle E_L|E_R\rangle|$ (the visibility law (2.12) there): the environment states $|E_L\rangle$, $|E_R\rangle$ that decohere a superposition are, when the environment is a field, two coherent states $|\{\alpha^L_{\vec{k}}\}\rangle$, $|\{\alpha^R_{\vec{k}}\}\rangle$ — and the decoherence factor is exactly (4.16). *A coherent state is the environment record.* The overlap is the fringe contrast the system can still show.

> **The one formula, twice.** Chapter 2 gave the visibility $\mathcal{V} = |\langle E_L|E_R\rangle|$ abstractly. This chapter identifies the environment states as coherent states of a field and hands you their overlap in closed form, $\mathcal{V} = \exp\big(-\frac{1}{2}\int\frac{d^3k}{(2\pi)^3}|\alpha^L_{\vec{k}}-\alpha^R_{\vec{k}}|^2\big)$. Chapter 11 does nothing but evaluate this integral with the field being the linearized *gravitational* field and $\alpha^{L,R}_{\vec{k}}$ the amplitudes it sources when the mass sits left or right. The mode-sum in the exponent is the gravitational self-energy $E_{\mathrm{G}}$ of the difference source. Every hard step remaining in the book is inside that integral.

### 4.2.4 Why coherent states are "the most classical"

Two precise senses justify the slogan, and both matter downstream.

*First: minimum uncertainty.* Compute the spreads of $x$ and $p$ in $|\alpha\rangle$. Using $a|\alpha\rangle = \alpha|\alpha\rangle$ and its conjugate $\langle\alpha|\, a^\dagger = \langle\alpha|\, \alpha^*$, one finds $\langle x\rangle = \sqrt{2\hbar/m\omega}\,\Re\alpha$, $\langle p\rangle = \sqrt{2\hbar m\omega}\,\Im\alpha$, and the variances

$$(\Delta x)^2 = \frac{\hbar}{2m\omega}, \qquad (\Delta p)^2 = \frac{\hbar m\omega}{2}, \qquad \Delta x\,\Delta p = \frac{\hbar}{2}. \tag{4.17}$$

The product saturates the Heisenberg bound: a coherent state is as sharply defined in phase space as quantum mechanics permits, a minimum-uncertainty blob that neither spreads nor squeezes as it orbits. (The variances do not depend on $\alpha$ — displacing the vacuum moves the blob without deforming it, which is exactly what (4.10) said.)

*Second: the expectation follows the classical field.* Because $a\,|\alpha\rangle = \alpha\,|\alpha\rangle$, the expectation value of the field operator (4.3) in a coherent state is $\langle\alpha|\phi(\vec{x},t)|\alpha\rangle = \phi_{\rm cl}(\vec{x},t)$, the *classical* field configuration with complex amplitude $\alpha_{\vec{k}}$ per mode. The quantum coherent state and the classical wave share the same $\langle\phi\rangle$; the coherent state merely adds the irreducible minimum fluctuation (4.17) around it. This is the exact sense in which "the gravitational field of a mass is a coherent state of gravitons" (§4.5): the classical Newtonian potential *is* $\langle\alpha|\hat{\Phi}|\alpha\rangle$ for the appropriate $\alpha$, and the quantum which-path information lives in the fluctuation the classical description throws away.

**Worked Example 4.1 — How far apart before two field states forget each other**

Two coherent states of a single mode are "distinguishable" once their overlap (4.15) falls to $1/e$, i.e. once $|\alpha-\beta| = 1$. For the gravitational preview: a mode's amplitude difference between the two branches of a mass superposition is (from §4.5) $\alpha^L_{\vec{k}} - \alpha^R_{\vec{k}} \propto GM$, so a *single* mode is nowhere near distinguishing the branches — $GM$ is minute. Distinguishability is built up by *summing over all modes* in (4.16). Anticipating Chapter 11, the mode sum evaluates to

$$\tfrac{1}{2}\int \frac{d^3k}{(2\pi)^3}|\alpha^L_{\vec{k}} - \alpha^R_{\vec{k}}|^2 \;=\; \frac{E_{\rm G}\,t}{\hbar} \quad \text{(schematically)}, \qquad E_{\rm G} = \frac{GM^2}{d},$$

so no single mode decoheres anything, but the collective overlap decays with a rate set by the gravitational self-energy $E_{\rm G}$ — the same $E_{\rm G}$ as the benchmark of Chapter 1 ($6.7\times10^{-26}$ J at one microgram, one millimetre). The coherent-state overlap is where the number $E_{\rm G}$ enters the decoherence rate.

▶ Run it: `TOOLS/verify-paper-numerics.py` (cross-paper benchmark X: $E_{\rm G} = GM^2/d$)

## 4.3 Bogoliubov transformations: whose vacuum?

The mode expansion (4.3) made a hidden choice: it split the field into $e^{-i\omega_k t}$ ("positive frequency," carried by $a$) and $e^{+i\omega_k t}$ ("negative frequency," carried by $a^\dagger$), and the vacuum (4.6) was defined relative to *that* split. But "positive frequency" means "oscillating as $e^{-i\omega t}$ with respect to *some* time coordinate," and different observers carry different clocks. An observer using time $t$ and an observer using time $\tilde{t}$ can split the same field differently — and then they disagree about what the vacuum is, and about how many particles are present. The mathematics of that disagreement is the Bogoliubov transformation, and it is the engine of every "particle creation by gravity" phenomenon in physics: Unruh radiation, Hawking radiation, and the cosmological particle production of Chapter 14.

### 4.3.1 Two mode-bases, one field

Suppose two observers expand the same field in two sets of modes. Observer $a$ uses annihilation operators $a$ (vacuum $|0_a\rangle$, defined by $a\,|0_a\rangle = 0$); observer $b$ uses $b$ (vacuum $|0_b\rangle$, $b\,|0_b\rangle = 0$). Because both bases span the same field, each observer's operators are linear combinations of the other's. For a single pair of modes the relation is

$$a = \alpha\, b + \beta\, b^\dagger, \tag{4.18}$$

where $\alpha, \beta$ are complex numbers (the *Bogoliubov coefficients* — unrelated to the coherent-state $\alpha$ of the previous section; the clash of letters is standard and we keep it, disambiguating by context). The defining feature, and the whole source of the physics, is the $b^\dagger$ term: observer $a$'s annihilation operator is part creation in observer $b$'s language. Preserving the commutator $[a, a^\dagger] = \mathbb{1}$ forces a normalization on the coefficients:

$$[a, a^\dagger] = |\alpha|^2 [b, b^\dagger] + |\beta|^2 [b^\dagger, b] = \left(|\alpha|^2 - |\beta|^2\right) \mathbb{1} \overset{!}{=} \mathbb{1} \quad \Longrightarrow \quad |\alpha|^2 - |\beta|^2 = 1. \tag{4.19}$$

The transformation is a hyperbolic rotation ($\cosh^2 - \sinh^2 = 1$), not a plane one — a *squeeze*. Whenever $\beta \neq 0$, the two observers genuinely disagree.

### 4.3.2 The particle count in the other observer's vacuum

Here is the payoff. Observer $a$ prepares the field in *her* vacuum $|0_a\rangle$ — "empty," by her definition. How many $b$-particles does observer $b$ find in it? Compute the expected $b$-occupation, using (4.18) to write $b$ in terms of $a$'s operators. Invert (4.18) (the inverse of a Bogoliubov transformation is another one): $b = \alpha^* a - \beta a^\dagger$. Then

$$\begin{aligned} \langle 0_a|\, N_b\, |0_a\rangle &= \langle 0_a|\, b^\dagger b\, |0_a\rangle = \langle 0_a|(\alpha a^\dagger - \beta^* a)(\alpha^* a - \beta a^\dagger)|0_a\rangle \\ &= |\beta|^2\ \langle 0_a|\, a\, a^\dagger\, |0_a\rangle = |\beta|^2, \end{aligned} \tag{4.20}$$

where every term with an $a$ acting to the right (or $a^\dagger$ to the left) on $|0_a\rangle$ dies, leaving only the $|\beta|^2 a a^\dagger$ piece, and $\langle 0_a|a a^\dagger|0_a\rangle = \langle 0_a|(a^\dagger a + \mathbb{1})|0_a\rangle = 1$. So:

$$\langle 0_a|\, N_b\, |0_a\rangle = |\beta|^2. \tag{4.21}$$

*One observer's vacuum contains $|\beta|^2$ particles by the other observer's count.* If $\beta = 0$ the two agree on the vacuum and on all particle numbers (the transformation is a mere phase). If $\beta \neq 0$, "empty" for one is "occupied" for the other, and the occupation is set entirely by how much the two mode-splittings mismatch. Everything that follows — the Unruh effect above all — is the computation of $\beta$ for two specific observers. There is nothing more to particle creation by geometry than this single number.

**Q 4.3** ***"This sounds like a paradox. Particles are supposed to be real — either they're there or they aren't. How can the* number *of particles depend on who's asking?"***

It depends on who is asking because "particle" is defined relative to a notion of positive frequency, and positive frequency is defined relative to a time coordinate — an observer's clock. There is no observer-independent split of a field into "particles" and "vacuum" in a general spacetime; the split you learned in flat-space, inertial quantum field theory is a *privileged choice* available only to inertial observers, who all agree because they share (up to Lorentz transformation) one notion of time. The moment observers disagree about time — one accelerating, one inertial; one before, one after a cosmological expansion — they disagree about particles, and (4.21) is the exact measure of the disagreement. What is observer-independent is the *state* of the field (a single vector in Hilbert space) and the response of *concrete detectors* — and a concrete detector carried by the accelerating observer really does click, as we are about to see. The particle concept is frame-dependent; the detector clicks are not. Both statements are true, and the Unruh effect is where they meet.

## 4.4 The Unruh effect

We now assemble the pieces — Rindler coordinates from Chapter 3, the mode expansion of §4.1, and the Bogoliubov count of §4.3 — into the chapter's headline result: an observer accelerating uniformly through the Minkowski vacuum finds it populated with a *thermal* bath of particles, at a temperature fixed by her acceleration. The vacuum, it turns out, is warm to whoever runs.

### 4.4.1 The accelerated observer and her horizon

Chapter 3 built the coordinates of a uniformly accelerated observer: the *Rindler wedge*. We recall only the two facts we need and refer to Chapter 3 for the construction. An observer with constant proper acceleration $a$ moves, in Minkowski coordinates $(ct, x)$, along the hyperbola $x^2 - c^2t^2 = (c^2/a)^2$. Introducing Rindler coordinates $(\eta, \xi)$ adapted to her by

$$ct = \xi \sinh\Big(\frac{a\eta}{c}\Big), \qquad x = \xi \cosh\Big(\frac{a\eta}{c}\Big), \tag{4.22}$$

the flat metric becomes $ds^2 = -(a\xi/c)^2\, d\eta^2 + d\xi^2$ (the $\eta$-lines are the accelerated worldlines; $\eta$ is proper time on the worldline $\xi = c^2/a$). Two features carry the physics. First, the Rindler time $\eta$ generates *boosts*: advancing $\eta$ is a Lorentz boost about the origin, so the accelerated observer's "time translation" is a boost, and her Hamiltonian is the boost generator. Second — and this is the crux — the lines $x = \pm ct$ are a *horizon*: the accelerated observer can never receive signals from beyond $x = ct$, and the region $x < c|t|$ is forever hidden from her. She has access to *half* the field.

That last fact is the entire mechanism, and we can see the answer coming before any calculation. The observer sees only half of the field's degrees of freedom; the modes she cannot see are entangled with the modes she can (the Minkowski vacuum is one big entangled state across the horizon); and by Chapter 2, *tracing out the inaccessible half of an entangled pure state leaves a mixed state*. The Unruh temperature is what that mixed state's temperature turns out to be.

### 4.4.2 The Minkowski vacuum across the horizon

Quantize the field two ways: in Minkowski modes (positive frequency with respect to inertial time $t$, operators $a_{\vec{k}}$, vacuum $|0_M\rangle$) and in Rindler modes (positive frequency with respect to the accelerated observer's time $\eta$, operators $b^R_\omega$ on the right wedge and $b^L_\omega$ on the hidden left wedge). The two are related by a Bogoliubov transformation, and the coefficient $\beta$ is fixed by one clean analytic fact: matching a Minkowski mode (smooth across the horizon) to the Rindler modes requires continuing $\eta$ into the *imaginary* direction, where sinh and cosh in (4.22) become periodic.

Rather than grind through the mode-matching integrals, we take the route that exposes the $2\pi$ most cleanly — the one previewed in the review track — and express the Minkowski vacuum directly in Rindler modes. The result of the mode-matching (derived in full in Chapter 3's appendix; we quote its output and verify its consequence) is that for each Rindler frequency $\omega$ the Minkowski vacuum is a *two-mode squeezed state* pairing the right-wedge mode with its left-wedge partner:

$$|0_M\rangle = \prod_\omega \frac{1}{\cosh r_\omega} \sum_{n=0}^{\infty} (\tanh r_\omega)^n \, |n\rangle_\omega^R \, |n\rangle_\omega^L \,, \qquad \tanh r_\omega = e^{-\pi\omega c/a}. \tag{4.23}$$

The squeezing parameter $r_\omega$ is set entirely by the ratio of the mode frequency to the acceleration; the exponent $\pi\omega c/a$ is where the horizon's imaginary-time periodicity enters. This state is exactly the *thermofield double* of Chapter 2's purification: the hidden left wedge is the purifying partner of the visible right wedge, and the whole Minkowski vacuum is a single entangled pure state straddling the horizon.

**Q 4.4** ***"Where did** $\tanh r_\omega = e^{-\pi\omega c/a}$ **come from? You said "imaginary-time periodicity" but that's a phrase, not a derivation."***

Fair — here is the mechanism in one paragraph, with the full mode integrals deferred to Chapter 3. A Minkowski positive-frequency mode is analytic (has no singularities) in the lower half of complex $t$; this is the precise statement of "positive frequency" — it is what $e^{-i\omega t}$ with $\omega > 0$ means, that the function decays when $t \to t - i\infty$. Now express that same mode in Rindler coordinates (4.22). Because $ct = \xi \sinh(a\eta/c)$, continuing $t$ into the lower half-plane corresponds to shifting the Rindler time $\eta \to \eta - i\pi c/a$: at that imaginary shift, sinh and cosh pick up signs that carry a right-wedge mode to a left-wedge mode. Demanding that the combined mode be analytic (Minkowski positive frequency) forces the right- and left-wedge amplitudes into a fixed ratio, and that ratio is $e^{-\pi\omega c/a}$ — the Boltzmann factor at half the eventual temperature. The number $\pi$ is half of $2\pi$; the $2\pi$ is the period of the imaginary Rindler time that returns you to where you started (a full boost by imaginary angle $2\pi$ is the identity). That periodicity in imaginary time is the *KMS condition* — the mathematical fingerprint of thermal equilibrium — and Chapter 7 makes it the definition of temperature. Keep the number $2\pi$ in view: it is about to become the temperature's denominator, and in Chapter 9 the whole theory.

### 4.4.3 Tracing out the horizon: a thermal state

The accelerated observer cannot touch the left wedge, so her state is the reduced density matrix, traced over the left modes. This is exactly the partial trace of Chapter 2 applied to (4.23). For one frequency, tracing the two-mode squeezed state over the $L$ factor (the cross terms die on $\langle n^L|m^L\rangle = \delta_{nm}$, just as in Worked Example 2.2) leaves

$$\rho_\omega^R = \frac{1}{\cosh^2 r_\omega} \sum_n (\tanh^2 r_\omega)^n \, |n\rangle^R \langle n|^R = (1 - e^{-2\pi\omega c/a}) \sum_n e^{-2\pi\omega c n/a} \, |n\rangle^R \langle n|^R \,, \tag{4.24}$$

using $\tanh^2 r_\omega = e^{-2\pi\omega c/a}$ and $1/\cosh^2 r_\omega = 1 - \tanh^2 r_\omega$. Compare this with the standard thermal (Boltzmann–Gibbs) density matrix of an oscillator at temperature $T$, $\rho \propto \sum_n e^{-n\hbar\omega/k_B T} |n\rangle \langle n|$. The two match term by term provided

$$\frac{\hbar\omega}{k_B T} = \frac{2\pi\omega c}{a} \qquad \Longrightarrow \qquad k_B T = \frac{\hbar a}{2\pi c}. \tag{4.25}$$

The accelerated observer's reduced state is *exactly thermal* — a genuine Boltzmann distribution over particle number — at a temperature proportional to her acceleration. Equivalently, the mean occupation of a Rindler mode of frequency $\omega$ is the Bose–Einstein form

$$\langle N_\omega^R \rangle = |\beta_\omega|^2 = \frac{e^{-2\pi\omega c/a}}{1 - e^{-2\pi\omega c/a}} = \frac{1}{e^{2\pi\omega c/a} - 1}, \tag{4.26}$$

which we recognize (via (4.21)) as the Bogoliubov count $|\beta_\omega|^2$: the Minkowski vacuum contains, in the accelerated observer's mode $\omega$, exactly a thermal number of quanta. The two routes — squeezed-state trace and Bogoliubov count — give the same $|\beta_\omega|^2$, as they must.

### 4.4.4 The Unruh temperature

Reading off the temperature from (4.25) and restoring $k_B$:

$$\boxed{T_U = \frac{\hbar\, a}{2\pi\, c\, k_B}} \tag{4.27}$$

*An observer accelerating uniformly through the Minkowski vacuum measures a temperature $T_U$ proportional to her acceleration.* Acceleration through "empty space" feels warm. The vacuum is not an absolute — what one inertial observer calls the ground state, an accelerated observer calls a thermal bath, and a real thermometer she carries would come to equilibrium at $T_U$. This is the Unruh effect (Fulling 1973, Davies 1975, Unruh 1976), and it is the first exact demonstration in this book that *quantum statistics and spacetime geometry are the same subject*: a purely geometric quantity (acceleration, hence a horizon) produces a purely thermal spectrum, with Planck's constant setting the scale.

**Worked Example 4.2 — How warm is acceleration?**

The prefactor in (4.27) is tiny:

$$\frac{\hbar}{2\pi c k_B} = \frac{1.054571817 \times 10^{-34}}{2\pi\,(2.99792458 \times 10^8)\,(1.380649 \times 10^{-23})} = 4.055 \times 10^{-21}\ \mathrm{K}\,/\,(\mathrm{m\,s^{-2}}),$$

so $T_U = 4.055 \times 10^{-21}\, a$ with $a$ in m/s$^2$. Three readings:

- **Earth gravity**, $a = 9.8$ m/s$^2$: $T_U = 3.97 \times 10^{-20}$ K. Standing on the Earth's surface, you accelerate at $g$ and bathe in an Unruh temperature forty billionths of a billionth of a kelvin above absolute zero — utterly unmeasurable, which is why nobody noticed the vacuum was warm for the first three centuries of physics.
- **One kelvin** of Unruh temperature requires $a = k_B T \cdot 2\pi c/\hbar = 2\pi(2.998 \times 10^8)(1.381 \times 10^{-23})/(1.055 \times 10^{-34}) = 2.47 \times 10^{20}$ m/s$^2$: an acceleration twenty orders of magnitude

beyond Earth gravity. This is why the Unruh effect, though exact, has never been directly measured — the accelerations are astronomical.

- **Intense-laser scale**, $a \approx 10^{26}$ m/s$^2$ (an electron in a focused petawatt laser field): $T_{\mathrm{U}} \approx 4.1 \times 10^5$ K, hot enough that proposals to detect Unruh radiation in the radiation from such electrons are taken seriously.

▶ RUN IT: `TOOLS/verify-paper-numerics.py` `(section N: ch4 Unruh temperature` $T_{\mathrm{U}} = \hbar a/2\pi c k_{\mathrm{B}}$`)`

### 4.4.5 Why $2\pi$, and where it goes

The number $2\pi$ in (4.27) is not a fudge, and it is worth pausing on because it is the same $2\pi$ that runs through the rest of this book. Trace its origin: it is the period, in imaginary time, of the accelerated observer's clock. A boost through imaginary rapidity $2\pi$ is the identity (just as a rotation through $2\pi$ is), so the accelerated observer's Euclidean time is a *circle* of circumference $2\pi/a$ (in the natural units of the boost). A quantum field on a Euclidean time circle of circumference $\beta$ is, by a theorem we will meet as the KMS condition (Chapter 7), a thermal field at temperature $1/\beta$ — so the circumference $2\pi$ *is* the inverse temperature in boost units, and restoring dimensions gives (4.27).

**The $2\pi$ is a through-line.** The same $2\pi$ appears at three later, load-bearing places, and it is always this one:

- **Chapter 7 (Bisognano–Wichmann).** The modular Hamiltonian of the Rindler wedge is exactly $2\pi$ times the boost generator, $K = 2\pi H_{\mathrm{boost}}$ — the rigorous, all-orders version of (4.23), with the $2\pi$ being this same imaginary-time period.
- **Chapter 9 (Axiom III).** The central equation of the theory, $K = 2\pi H_{\mathrm{phys}}$, is the promotion of Bisognano–Wichmann from the Rindler wedge to a general causal diamond. Its $2\pi$ is the Unruh $2\pi$.
- **Chapter 14 (the MOND scale).** The de Sitter horizon carries an Unruh-like temperature $T = \hbar H_0/2\pi k_{\mathrm{B}}$, and the acceleration scale $a_0 = cH_0/2\pi$ inherits the same denominator.

The Unruh calculation is where the reader should locate the number $2\pi$ that the whole theory turns on. It is the periodicity of imaginary time at a horizon — nothing more exotic, and nothing less fundamental.

**Q 4.5** ***"If the Unruh effect has never been measured, why should I trust it enough to build a theory on its $2\pi$?"***

Because it is not really an independent hypothesis — it is a theorem of the quantum field theory we already trust, forced by nothing but the analyticity of positive-frequency modes and the geometry of a horizon (both used above). Every step was ordinary flat-space quantum field theory plus Rindler coordinates; no new physics entered. In that sense the Unruh temperature is as secure as the vacuum fluctuations that give the Lamb shift (measured to many digits) or the Casimir force (measured), which spring from the same $\langle 0|\phi^2|0\rangle \neq 0$. What has not been done is the specific, brutally hard experiment of accelerating a detector at $10^{20}$ m/s$^2$ and watching it thermalize — but the effect's *ingredients* are all confirmed, and its close cousin, Hawking radiation, has been observed in laboratory analogues (fluid and optical horizons). The theory of this book does not rest on Unruh radiation being detected; it rests on the modular structure (4.23) that the Unruh calculation *derives*, and that structure (Bisognano–Wichmann) is a rigorous theorem, [PROVEN]. The Unruh effect is the physically transparent window onto a mathematically airtight fact.

## 4.5 Linearized gravity as a spin-2 field

One field remains to quantize, the one this whole book is about: the gravitational field itself, in the weak regime where it is a small ripple on flat spacetime. We build only as much as Chapter 11 needs — enough to state precisely that "the gravitational field of a mass is a coherent state of gravitons."

### 4.5.1 The graviton is a spin-2 quantum

Return to the metric split of Chapter 1 (1.7),

$$g_{\mu\nu} = \eta_{\mu\nu} + \sqrt{32\pi G}\, h_{\mu\nu}, \tag{4.28}$$

which writes the geometry as flat spacetime $\eta_{\mu\nu}$ plus a small fluctuation $h_{\mu\nu}$ (the factor $\sqrt{32\pi G}$ is a normalization that makes $h$ a canonically normalized field, so its kinetic term looks like every other field's). Feeding (4.28) into Einstein's equations and keeping only terms linear in $h$ gives a wave equation for $h_{\mu\nu}$ identical in structure to (4.2) — the field $h_{\mu\nu}$ propagates at $c$ and is sourced by matter,

$$\Box\, h_{\mu\nu} = -\sqrt{8\pi G}\,\left(T_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}T\right), \qquad \Box = -\tfrac{1}{c^2}\partial_t^2 + \nabla^2, \tag{4.29}$$

with $T_{\mu\nu}$ the stress–energy tensor of Chapter 1 and $T = \eta^{\mu\nu}T_{\mu\nu}$ its trace. The field $h_{\mu\nu}$ carries two tensor indices where the scalar $\phi$ carried none, and after removing the gauge and constraint components (the analogue of choosing Coulomb gauge in electromagnetism) it has *two* physical polarizations, which transform under rotations as *spin 2*. Quantize it exactly as we quantized the scalar — one oscillator per mode and polarization — and the quanta are *gravitons*: massless, spin-2, the gravitational cousins of the photon.

### 4.5.2 The Newtonian field is a coherent state of gravitons

Now the point. A static mass $M$ at position $\vec{x}_A$ is a classical source $T_{00} = Mc^2\delta^3(\vec{x} - \vec{x}_A)$ (the other components vanish for a mass at rest), and the linearized equation (4.29) it sources is precisely a field driven by a classical current. A quantum field driven by a classical source lands in a *coherent state* — this is a general theorem (Glauber 1963; the driven oscillator's ground state is displaced), and it is the same statement as §4.2.4's "$\langle\hat{\Phi}\rangle = \Phi_{\rm cl}$." The upshot is that the gravitational field of the mass is

$$|g_A\rangle = D(\alpha_A)\,|0\rangle\,, \qquad \alpha_A(\vec{k}) = -\frac{4\pi GM}{k^2}\frac{e^{-i\vec{k}\cdot\vec{x}_A}}{\sqrt{2\hbar\omega_k}}, \tag{4.30}$$

a coherent state of gravitons whose amplitude $\alpha_A(\vec{k})$ is, mode by mode, the Fourier transform of the Newtonian potential — the factor $1/k^2$ is the Fourier transform of the Poisson constraint $\nabla^2\Phi = 4\pi G\rho$, the $e^{-i\vec{k}\cdot\vec{x}_A}$ places the mass at $\vec{x}_A$, and $M$ sets the overall strength. The classical Newtonian potential is recovered as $\langle g_A|\hat{\Phi}|g_A\rangle = -GM/|\vec{x} - \vec{x}_A|$, exactly the classicality of §4.2.4.

**The statement Chapter 11 rests on.** The gravitational field of a mass at $\vec{x}_A$ is the coherent state (4.30). Two branches of a superposition — mass at $\vec{x}_L$ versus $\vec{x}_R$ — source two coherent states $|g_L\rangle$, $|g_R\rangle$, and their overlap is the multimode Gaussian (4.16), with the mode-sum in the exponent equal to the gravitational self-energy $E_G = GM^2/d$ of the difference source. Decoherence of the superposition *is* this overlap. Every tool needed to compute it was built in this chapter.

**Scope.** Everything in §4.5 is [DERIVED] in the *linearized* regime $GM/c^2d \ll 1$: gravity treated as a weak spin-2 field on flat spacetime, valid far below the Planck scale (Chapter 1). Nonlinear graviton self-interaction and the full quantum theory of $h_{\mu\nu}$ are the non-renormalizable regime of Chapter 1 and play no role here — the decoherence of Chapter 11 lives entirely in the linear, coherent-state sector. The subtlety that gravity's constraint *forces* the coherent state to accompany the mass (unlike a gas molecule, which may or may not scatter) is the business of Chapter 11; here we establish only that the field state *is* coherent.

**Q 4.6** ***"You quantized $h_{\mu\nu}$ here, but Chapter 1 spent pages arguing that quantizing the metric fails. Which is it?"***

Both, and the reconciliation is exactly the effective-field-theory point of Chapter 1. Quantizing $h_{\mu\nu}$ as a *free* (or nearly free) spin-2 field on flat space works perfectly and is standard — it is what we just did, and it gives sensible gravitons, coherent states, and the Newtonian potential. What *fails* is the attempt to compute high-loop corrections, where the power counting of Chapter 1 ($[G] = \text{mass}^{-2}$, $L$-loop amplitudes $\sim (\Lambda/m_P)^{2L}$) makes the theory non-predictive above the Planck scale. The decoherence calculation of Chapter 11 lives at the *tree level* of this effective theory — the coherent state (4.30) is a classical-source (zero-loop) object — and is therefore firmly inside the regime where quantized linearized gravity is valid and uncontroversial. We use gravity as a quantum field theory exactly where it is trustworthy, and never where it is not. The theory of this book makes no claim about Planck-scale gravity; its wager is that the measurable quantum-gravitational effect lives here, in the coherent, linear, tree-level sector.

## Chapter FAQ

**Q: Is the vacuum energy $\sum \frac{1}{2}\hbar\omega_k$ in (4.5) real? It looks infinite.** The *absolute* zero-point energy is unmeasurable and conventionally dropped (we measure energy differences), and the sum's formal infinity is the first of the ultraviolet divergences that quantum field theory learns to handle. But zero-point *fluctuations* — $\langle 0|\phi^2|0\rangle \neq 0$, not the energy — are entirely real and measured (Lamb shift, Casimir force), and they are what the Unruh effect exploits. The lesson to carry: the vacuum's energy is a convention, but its correlations are physics, and the whole of Part III is about vacuum correlations across a horizon.

**Q: The coherent-state $\alpha$ and the Bogoliubov $\alpha$ are different things. Will this bite me later?** They are different (coherent-state $\alpha$ is a displacement amplitude, an eigenvalue of $a$; Bogoliubov $\alpha$ is a mode-mixing coefficient), and the clash is unfortunately standard. We disambiguate by context throughout: near $D(\alpha)$ and $|\alpha\rangle$ it is the coherent amplitude; near $a = \alpha b + \beta b^\dagger$ it is the Bogoliubov coefficient. When both appear in one calculation (they do not, in this book) one renames. No result depends on the letter.

**Q: Does the Unruh effect need a real, physical detector, or is it just a change of mathematical basis?** Both descriptions are correct and consistent. Mathematically it is a Bogoliubov change of basis, and one can insist "there are no real particles, only a different mode decomposition." Physically, a concrete two-level detector (an "Unruh–DeWitt detector") carried along the accelerated worldline genuinely gets excited at the thermal rate — it clicks, absorbing energy from its own acceleration via whatever agent drives it. The two statements agree because the detector *defines* a particle notion by its response, and that response is thermal. Frame-dependence of the particle concept and reality of the detector clicks are not in tension; the clicks are what (4.26) predicts.

**Q: Why a scalar field for most of the chapter, when gravity is spin 2?** Because every conceptual step — modes as oscillators, vacuum, particles, coherent states, Bogoliubov mixing, the Unruh thermality — is identical for any field; only the index structure and polarization count change. The scalar is the clean laboratory. Section 4.5 then transplants the whole apparatus to the spin-2 graviton with no new ideas, only more indices, and Chapter 11's detailed computation confirms that the scalar "proxy" captures the spin-2 answer exactly (the scalar-proxy result).

## Exercises

**Exercise 4.1.** (a) Using (4.13), verify directly that $\langle\alpha|\alpha\rangle = 1$ and that $\langle N\rangle = \langle\alpha|\, a^\dagger a \,|\alpha\rangle = |\alpha|^2$. (b) Two single-mode coherent states have $\alpha = 2$ and $\beta = 2.5$ (both real). Compute their overlap $|\langle\alpha|\beta\rangle|^2$ from (4.15), and the average photon numbers in each. (c) A field environment records which-path information in $N$ independent modes, each with branch amplitudes differing by $\delta = |\alpha_k^L - \alpha_k^R| = 0.05$. For what $N$ does the total visibility (4.16) fall to $1/e$? Relate your answer to the exponential decoherence law of Chapter 2.

**Exercise 4.2.** (a) Compute the Unruh temperature (4.27) for the acceleration at the surface of a proton-radius region, $a \sim 10^{31}$ m/s$^2$, and compare with the temperature at the centre of the Sun ($\sim 1.5 \times 10^7$ K). (b) The surface gravity of a solar-mass black hole is $\kappa = c^4/4GM_\odot \approx 1.5 \times 10^{13}$ m/s$^2$. Using the Unruh formula with $a \to \kappa$, compute the temperature and compare with the known Hawking temperature $T_\mathrm{H} = \hbar c^3/8\pi G M_\odot k_\mathrm{B} \approx 6 \times 10^{-8}$ K. What does the agreement tell you about the relationship between the Unruh and Hawking effects?

## Solutions

*Solution (Exercise 4.1).* (a) Normalization: $\langle\alpha|\alpha\rangle = e^{-|\alpha|^2}\sum_n |\alpha|^{2n}/n! = e^{-|\alpha|^2}e^{|\alpha|^2} = 1$. Number: $\langle\alpha|\, a^\dagger a \,|\alpha\rangle = \langle\alpha|\, \alpha^*\alpha \,|\alpha\rangle = |\alpha|^2$ using (4.12) on both sides. (b) With $\alpha = 2$, $\beta = 2.5$: $|\langle\alpha|\beta\rangle|^2 = e^{-(2-2.5)^2} = e^{-0.25} = 0.779$ — still 78% overlap despite differing amplitudes; single modes barely distinguish. Photon numbers $|\alpha|^2 = 4$ and $|\beta|^2 = 6.25$. (c) The total exponent is

$\sum_k |\alpha_k^L - \alpha_k^R|^2 = N\delta^2 = N(0.05)^2 = N/400$; setting this to 1 gives $N = 400$ modes. This is the exponential law of Chapter 2 (the exponential law (2.14)) with $\epsilon = \delta^2 = 2.5 \times 10^{-3}$ per mode: many nearly-uninformative modes accumulate into decoherence, exactly as many nearly-elastic collisions did. The coherent-state overlap makes "$\epsilon$ per record" concrete: it is the squared amplitude difference.

▶ RUN IT: `TOOLS/verify-paper-numerics.py (section N: ch4 coherent overlaps)`

*Solution (Exercise 4.2).* (a) $T_{\mathrm{U}} = 4.055 \times 10^{-21} \times 10^{31} = 4.1 \times 10^{10}$ K, some 2700 times hotter than the solar centre — at nuclear-scale accelerations the Unruh bath is genuinely hot, which is why analogue and high-energy proposals target this regime. (b) With $\kappa = 1.5 \times 10^{13}$ m/s$^2$: $T_{\mathrm{U}} = 4.055 \times 10^{-21} \times 1.5 \times 10^{13} = 6.1 \times 10^{-8}$ K, matching the Hawking temperature to the digits quoted. The agreement is not a coincidence: the Hawking temperature of a black hole is the Unruh temperature felt by a static observer just outside the horizon, whose proper acceleration (needed to hover) equals the surface gravity $\kappa$. Hawking radiation *is* the Unruh effect at a gravitational horizon — the same imaginary-time periodicity, the same $2\pi$, now around the Euclidean Schwarzschild time. Chapter 5 makes this the entry point to black-hole entropy and Axiom I.

▶ RUN IT: `TOOLS/verify-paper-numerics.py (section N: ch4 Unruh/Hawking scales)`

**What to carry forward**

1. A free quantum field is a bath of harmonic oscillators, one per mode (4.3): the vacuum is all oscillators in their ground state (4.6), particles are excitations, and no new machinery beyond the ladder operators is needed. Coherent states $|\alpha\rangle = D(\alpha)|0\rangle$ are the most classical field states (minimum uncertainty; $\langle\hat{\phi}\rangle = \phi_{\mathrm{cl}}$), and their overlap is the Gaussian $|\langle\{\alpha_k\}|\{\beta_k\}\rangle|^2 = \exp(-\sum_k |\alpha_k - \beta_k|^2)$ — which is Chapter 2's visibility law with the field as environment, and the exact object Chapter 11 computes for gravity.

2. A Bogoliubov transformation $a = \alpha b + \beta b^\dagger$ relates observers who disagree about positive frequency; one observer's vacuum contains $\langle 0_a|N_b|0_a\rangle = |\beta|^2$ particles by the other's count. This single number is all of particle creation by geometry.

3. An accelerated observer traces out the region behind her horizon and finds the Minkowski vacuum thermal at $T_{\mathrm{U}} = \hbar a/2\pi c k_{\mathrm{B}}$ ($4 \times 10^{-20}$ K at Earth gravity; 1 K needs $2.5 \times 10^{20}$ m/s$^2$). The $2\pi$ is the period of imaginary time at the horizon (the KMS condition), and it is the same $2\pi$ that becomes $K = 2\pi H_{\mathrm{phys}}$ in Chapter 9 and $a_0 = cH_0/2\pi$ in Chapter 14. The gravitational field of a mass is a coherent state of gravitons (4.30), [DERIVED] in the linearized, tree-level regime.

# 5 Gravity Meets Information

**Synopsis.** For four chapters the two rulebooks have kept to their own courtrooms. Here they collide, and the collision leaves a single, startling residue: a black hole has a *temperature* and an *entropy*, and the entropy is proportional not to the hole's volume but to the area of its horizon. We derive the Hawking temperature by standing an accelerated observer (Chapter 4) just outside a Schwarzschild horizon (Chapter 3); we let the first law of thermodynamics force the Bekenstein–Hawking area law $S_{\text{area}} = k_{\text{B}} A/4\ell_{\text{P}}^2$; we motivate the Bekenstein bound and the holographic bound that follow. Then we assemble the object this whole book turns on: the *generalized entropy* $S_{\text{gen}} = A/4\ell_{\text{P}}^2 + S_{\text{ext}}$, the horizon area plus the ordinary von Neumann entropy of matter outside it (Chapter 2). Its monotonic growth — the Generalized Second Law — ties geometry and information into one ledger. That expression, taken as a principle, is Axiom I of the theory (Chapter 9).

**You will be able to.** derive the Hawking temperature from Unruh plus redshift and evaluate it for a solar-mass hole; obtain $S_{\text{area}} = k_{\text{B}} A/4\ell_{\text{P}}^2$ from $dMc^2 = T\,dS$ and compute the entropy of a solar-mass hole; state and motivate the Bekenstein and holographic bounds; write down $S_{\text{gen}}$, run the box-into-the-hole argument for the Generalized Second Law, and say precisely which parts of $S_{\text{gen}}$ are physical and which are scheme-dependent bookkeeping.

**Prerequisites.** Chapters 1–4: von Neumann and relative entropy (Chapter 2); horizons, surface gravity, the Schwarzschild radius (Chapter 3); the Unruh temperature $T = \hbar a/2\pi c k_{\text{B}}$ (Chapter 4). First law of thermodynamics, $dE = T\,dS$.

## 5.1 The thermodynamic clue

Up to this point "entropy" and "geometry" have lived in different chapters. Chapter 2 built entropy as a bookkeeping of missing information; Chapters 3–4 built geometry, horizons, and the one place where quantum fields already noticed geometry — the Unruh effect, in which an observer accelerating through the vacuum finds it warm. This chapter is where the two meet, and they meet at an object that is unavoidably both geometric and thermal: a black hole.

The clue is old and, on first encounter, absurd. A black hole is defined by Chapter 3 as a region from which nothing — not even light — escapes. It is the coldest, deadest, most information-hostile object imaginable: everything that falls in is lost to the outside forever. Yet we are about to find that it has a definite temperature, radiates, and carries more entropy than anything else of its size. The route to that conclusion runs straight through the toolkit of the last two chapters, and we can walk it without any new postulate.

### 5.1.1 An accelerated observer at the horizon

Recall the operational content of the Unruh effect (Chapter 4). An observer with proper acceleration $a$, moving through the Minkowski vacuum, does not see empty space: her particle detector clicks as though immersed in a thermal bath at the temperature

$$T_{\mathrm{U}} = \frac{\hbar a}{2\pi c\, k_{\mathrm{B}}}. \tag{5.1}$$

The physical meaning, established in Chapter 4: acceleration erects a horizon (the Rindler horizon) behind the observer, she loses access to the modes beyond it, and tracing them out leaves her local state thermal — exactly the partial-trace-produces-mixedness mechanism of Chapter 2, now with spacetime doing the tracing.

Now bring this to a black hole. Chapter 3 gave the Schwarzschild geometry of a non-rotating mass $M$, with its horizon at the Schwarzschild radius

$$r_s = \frac{2GM}{c^2}, \tag{5.2}$$

and it defined the horizon's *surface gravity* $\kappa$ — operationally, the acceleration (measured far away, in the sense made precise below) needed to hold a test mass just outside the horizon against the pull of the hole. For Schwarzschild, Chapter 3 computes

$$\kappa = \frac{c^4}{4GM}. \tag{5.3}$$

This is the crucial import: a horizon, whether raised by acceleration (Rindler) or by mass (Schwarzschild), is a place where some observer's modes get traced out. The Unruh mechanism should not care *why* the horizon is there. So let us apply (5.1) at the Schwarzschild horizon, with the local acceleration playing the role of $a$, and see what temperature the hole acquires.

**Q 5.1 *"Why should the Unruh formula, derived for an observer accelerating through flat space, apply to a black-hole horizon at all?"***

Because near the horizon the two situations become the same geometry. Zoom in on any patch of a Schwarzschild horizon: over a region small compared with $r_s$, the curvature is negligible and the metric looks flat, while the horizon itself looks like a Rindler horizon — the boundary an accelerated observer sees. This is the equivalence principle of Chapter 3 doing its job: locally, "held static against a black hole" and "accelerating through empty space" are indistinguishable. So a static observer just outside the horizon is, in her own neighbourhood, exactly an Unruh observer, and (5.1) applies to her with $a$ equal to her local proper acceleration. The only new ingredient a curved spacetime adds is that her local temperature must be *redshifted* out to a distant observer — and that redshift, worked below, is what converts the (divergent) local acceleration at the horizon into the finite $\kappa$ of (5.3). The full, self-contained derivation of Hawking radiation from quantum field theory on the collapsing-star background is Hawking's 1975 calculation; we are recovering its temperature by the shortcut the toolkit allows, and we flag the shortcut as such. [DERIVED] (from Unruh + redshift; the underlying Hawking result is imported).

### 5.1.2 The Hawking temperature by redshift

A static observer hovering at radius $r$ just outside the horizon has a local proper acceleration $a_{\rm loc}(r)$ that grows without bound as $r \to r_s$: it costs ever more thrust to avoid falling in. By (5.1) she measures a *local* Unruh temperature $T_{\rm loc} = \hbar a_{\rm loc}/2\pi c k_{\rm B}$. Both her acceleration and her temperature diverge at the horizon. What a distant observer records is not this local value but its gravitationally redshifted image: a photon of local frequency $\nu_{\rm loc}$ climbing out of the potential well arrives at infinity with a lower frequency $\nu_\infty = \chi(r)\,\nu_{\rm loc}$, where the redshift factor $\chi(r)$ (the "lapse," built in Chapter 3) vanishes at the horizon. Temperature, being an energy scale, redshifts the same way:

$$T_\infty \;=\; \chi(r)\,T_{\rm loc}(r) \;=\; \chi(r)\,\frac{\hbar\,a_{\rm loc}(r)}{2\pi c\,k_{\rm B}}. \tag{5.4}$$

The divergences fight: $a_{\rm loc} \to \infty$ and $\chi \to 0$ as $r \to r_s$. Their product is finite, and it is precisely the surface gravity — this is what $\kappa$ *is*, geometrically: the horizon limit of $\chi\,a_{\rm loc}$,

$$\lim_{r\to r_s} \chi(r)\,a_{\rm loc}(r) \;=\; \kappa \;=\; \frac{c^4}{4GM}. \tag{5.5}$$

(We import (5.5) from Chapter 3, where the lapse and the local acceleration are computed explicitly; the statement that their product tends to the surface gravity is the definition of $\kappa$ for a static horizon.) Substituting (5.5) into (5.4) gives the temperature the black hole shows to the rest of the universe:

$$T_{\rm H} \;=\; \frac{\hbar\,\kappa}{2\pi c\,k_{\rm B}} \;=\; \frac{\hbar c^3}{8\pi GM\,k_{\rm B}}. \tag{5.6}$$

This is the *Hawking temperature*. In words: a black hole is not black. It glows, faintly, at a temperature fixed entirely by its mass — the heavier the hole, the colder it is, because $T_{\rm H} \propto 1/M$. A black hole is a thermodynamic object, with a genuine temperature an outside thermometer would (in principle) read.

**Worked Example 5.1 — The temperature of a solar-mass black hole**

Put a Sun's worth of mass, $M = M_\odot = 1.989 \times 10^{30}$ kg, into (5.6), using $G = 6.67430 \times 10^{-11}\,\mathrm{m^3kg^{-1}s^{-2}}$, $\hbar = 1.054571817 \times 10^{-34}\,\mathrm{J\,s}$, $c = 299792458\,\mathrm{m/s}$, $k_{\rm B} = 1.380649 \times 10^{-23}\,\mathrm{J/K}$. First the surface gravity (5.3):

$$\kappa = \frac{c^4}{4GM} = \frac{(2.99792458 \times 10^8)^4}{4\,(6.67430 \times 10^{-11})(1.989 \times 10^{30})} = 1.52 \times 10^{13}\ \mathrm{m/s^2}.$$

Then (5.6):

$$T_{\rm H} = \frac{\hbar\kappa}{2\pi c\,k_{\rm B}} = \frac{(1.054571817 \times 10^{-34})(1.52 \times 10^{13})}{2\pi\,(2.99792458 \times 10^8)(1.380649 \times 10^{-23})} = 6.2 \times 10^{-8}\ \mathrm{K}.$$

About sixty billionths of a kelvin — ten million times colder than the cosmic microwave background at 2.7 K. A solar-mass black hole is, thermodynamically, a frozen object: it absorbs the ambient radiation of the universe far faster than it emits, and it will not begin to evaporate until the cosmos cools below $T_{\mathrm{H}}$. The temperature is real but, for astrophysical holes, minuscule. [DERIVED]

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch5 Hawking temperature)`

**For the curious — Only the smallest holes are hot**

Because $T_{\mathrm{H}} \propto 1/M$, the Hawking temperature is astronomically small for astronomical holes and enormous for tiny ones. A hole of asteroid mass $\sim 10^{12}$ kg would glow near $10^{11}$ K; a hole of the Planck mass $m_{\mathrm{P}} \approx 2 \times 10^{-8}$ kg (Chapter 1) would have a temperature of order the Planck temperature, where this semiclassical derivation breaks down entirely. Radiation carries energy away, so a black hole *loses* mass, gets hotter, radiates faster — a runaway ending in a final flash. Dimensional analysis of the emitted power (a blackbody of area $A \sim r_s^2$ at temperature $T_{\mathrm{H}} \propto 1/M$) gives an evaporation time $t_{\mathrm{evap}} \sim G^2 M^3/\hbar c^4$; the bare combination gives $\sim 10^{63}$ years for a solar-mass hole, and the full Stefan–Boltzmann prefactor $(5120\pi \approx 1.6 \times 10^4)$ raises it to the standard $\sim 10^{67}$ years — either way unimaginably longer than the current age of the universe ($\sim 10^{10}$ yr). Evaporation matters for primordial and microscopic holes, not for the ones astronomers see — but it is the setting for the Page curve of §5.6, and it is where the deepest puzzle of the subject lives.

## 5.2 Entropy from the first law

A temperature is not a curiosity; it is a thermodynamic obligation. If a black hole has a temperature $T_{\mathrm{H}}$, then the first law of thermodynamics — $dE = T\,dS$, the same first law that governs steam engines — forces it to have an entropy too. We now let the first law do exactly that.

### 5.2.1 The area law, forced

The energy of a black hole is its rest energy, $E = Mc^2$. Throw a little mass $dM$ into the hole and its energy rises by $dE = c^2\,dM$. The first law says this must equal $T_{\mathrm{H}}\,dS_{\mathrm{BH}}$, where $S_{\mathrm{BH}}$ is whatever entropy the hole carries:

$$c^2\,dM \;=\; T_{\mathrm{H}}\,dS_{\mathrm{BH}} \qquad \Longrightarrow \qquad dS_{\mathrm{BH}} \;=\; \frac{c^2\,dM}{T_{\mathrm{H}}}. \tag{5.7}$$

Everything on the right is known. Substitute the Hawking temperature (5.6), $T_{\mathrm{H}} = \hbar c^3/(8\pi G M k_{\mathrm{B}})$:

$$dS_{\mathrm{BH}} \;=\; c^2\,dM \cdot \frac{8\pi G M\,k_{\mathrm{B}}}{\hbar c^3} \;=\; \frac{8\pi G\,k_{\mathrm{B}}}{\hbar c}\,M\,dM. \tag{5.8}$$

The right-hand side is an exact differential — $M\,dM = \frac{1}{2}\,d(M^2)$ — so we integrate from zero to the present mass with no arbitrary constant beyond the natural choice $S_{\mathrm{BH}} = 0$ at $M = 0$:

$$S_{\mathrm{BH}} \;=\; \frac{8\pi G\,k_{\mathrm{B}}}{\hbar c}\int_0^M M'\,dM' \;=\; \frac{4\pi G\,k_{\mathrm{B}}}{\hbar c}\,M^2. \tag{5.9}$$

This is already the answer, but it hides its meaning in the variable $M$. Trade $M$ for the horizon area. From the Schwarzschild radius (5.2), the horizon is a sphere of area

$$A = 4\pi r_s^2 = 4\pi\Big(\frac{2GM}{c^2}\Big)^2 = \frac{16\pi G^2 M^2}{c^4} \qquad\Longrightarrow\qquad M^2 = \frac{c^4 A}{16\pi G^2}. \tag{5.10}$$

Insert (5.10) into (5.9) and watch the mass disappear:

$$S_{\rm BH} = \frac{4\pi G\, k_{\rm B}}{\hbar c}\cdot\frac{c^4 A}{16\pi G^2} = \frac{k_{\rm B} c^3}{4G\hbar}\,A. \tag{5.11}$$

Now recognise the combination of constants. The Planck length (Chapter 1) is $\ell_{\rm P} = \sqrt{\hbar G/c^3}$, so $\ell_{\rm P}^2 = \hbar G/c^3$ and $c^3/(G\hbar) = 1/\ell_{\rm P}^2$. Equation (5.11) collapses to the form that is one of the most quoted in physics:

$$S_{\rm BH} = \frac{k_{\rm B}\, c^3}{4G\hbar}\,A = \frac{k_{\rm B} A}{4\ell_{\rm P}^2} \equiv S_{\rm area}. \tag{5.12}$$

This is the *Bekenstein–Hawking entropy*, and its physical reading is where the strangeness lives. The entropy of a black hole is one quarter of its horizon area, measured in Planck units — one nat (before the $k_{\rm B}$) for every four Planck squares of horizon. [DERIVED] (given the Hawking temperature; this is Theorem I.1 of the framework, DERIVED-THEOREMS.md).

**Q 5.2** ***"Every ordinary system I know has an entropy that scales with its volume — a gas, a solid, a star. Why on earth does a black hole's scale with area?"***

This is the profound part, and it is worth pausing on it rather than rushing past. For an ordinary body, entropy counts microstates, and microstates fill the interior: double the volume of an ideal gas at fixed density and you double its entropy. Extensivity — entropy proportional to volume — is so universal we take it for granted. A black hole flatly violates it. Its entropy (5.12) grows as $A \propto r_s^2$, not as the volume $\sim r_s^3$; a hole twice as wide has four times the entropy, not eight. The information content of the deepest gravitational object is written on its *surface*, as if the interior degrees of freedom had been projected onto the boundary. Every honest account of quantum gravity has had to reckon with this fact, and the two bounds of the next sections generalize it from black holes to *everything*. Chapter 2's closing remark — that when an area shows up as the answer to "how much can be hidden behind a horizon" it is the entanglement ledger being balanced — is about to be cashed in.

**Worked Example 5.2 — The entropy of a solar-mass black hole**

Take the solar-mass hole of Worked Example 5.1. Its Schwarzschild radius (5.2) is

$$r_s = \frac{2GM_\odot}{c^2} = \frac{2\,(6.67430\times10^{-11})(1.989\times10^{30})}{(2.99792458\times10^8)^2} = 2.95\times10^3\ \text{m}$$

— about three kilometres. Its horizon area is $A = 4\pi r_s^2 = 1.10\times10^8\ \text{m}^2$. The dimensionless entropy (in units of $k_B$) follows from (5.12) with $\ell_P^2 = 2.612 \times 10^{-70}$ m$^2$:

$$\frac{S_{BH}}{k_B} = \frac{A}{4\ell_P^2} = \frac{1.10 \times 10^8}{4\,(2.612 \times 10^{-70})} = 1.05 \times 10^{77}.$$

Restoring $k_B$ gives $S_{BH} = k_B \times 1.05 \times 10^{77} = 1.45 \times 10^{54}$ J/K. For contrast, the ordinary thermodynamic entropy of the *Sun* — summed over all $\sim 10^{57}$ of its particles at their thermal state — is only about $10^{58}\,k_B$. Collapse the Sun into a black hole of the same mass and its entropy jumps by nineteen orders of magnitude, from $\sim 10^{58}$ to $\sim 10^{77}$. The black hole is, for its mass and size, the *maximum-entropy object* in nature: nothing you can build out of the same matter, in the same region, hides more information. That last claim is not rhetoric — it is the content of the Bekenstein bound, which we derive next. [DERIVED]

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch5 Bekenstein-Hawking entropy)`

## 5.3 The Bekenstein bound

The solar entropy comparison suggests a universal ceiling: pour any matter into a region and there is a most entropy it can hold, and the black hole saturates it. Bekenstein made this quantitative before Hawking's temperature was known, by a thought experiment — a *Geroch process* — that uses nothing but the first law and the Bekenstein–Hawking area relation. We follow it, honestly labelling it a motivation rather than an airtight theorem.

### 5.3.1 Lowering a box to the horizon

Take a system — a box of gas, a hard drive, anything — of total energy $E$ (so rest-plus-internal energy; in its rest frame $E = mc^2$), linear size $R$, and entropy $S$. Lower it slowly on a string toward a large black hole and drop it across the horizon. Track the two entropies.

*Exterior entropy falls.* When the box crosses the horizon, its entropy $S$ leaves the outside world. The observers outside have lost $S$ from their ledger.

*Horizon area rises.* The box carried energy, so the hole's mass increases and its horizon grows, adding $\Delta S_{BH} = k_B\,\Delta A/4\ell_P^2$ to the hole's entropy by (5.12). The key move of the Geroch process is to *minimize* the energy actually delivered to the hole: by lowering the box quasi-statically on the string, one extracts work from its descent, and the energy that finally crosses the horizon can be reduced to the value it has when the box's centre sits a proper distance $\sim R$ (its own size) above the horizon — you cannot lower it further without part of it already being inside. At that point the redshifted energy delivered is of order

$$\Delta E_{\text{delivered}} \;\sim\; \frac{E\,\kappa R}{c^2} \quad \text{(schematically)}, \tag{5.13}$$

and the resulting area increase, through $dM = \Delta E/c^2$ and (5.10), works out to a horizon-entropy gain of order

$$\Delta S_{BH} \;\sim\; \frac{2\pi\,k_B\,R\,E}{\hbar c}. \tag{5.14}$$

Now demand the one principle we are not willing to give up: the Generalized Second Law of §5.5.2, that the *total* entropy — exterior plus horizon — must not decrease. The exterior lost $S$; the horizon gained (5.14); for the sum to stay non-negative we need the gain to cover the loss:

$$S \leq \frac{2\pi\, k_{\mathrm{B}}\, R\, E}{\hbar c}. \tag{5.15}$$

This is the *Bekenstein bound*. Its meaning is a universal information capacity: any physical system of size $R$ and energy $E$ can store at most $2\pi k_{\mathrm{B}} RE/\hbar c$ of entropy — at most that many nats of hidden information — *no matter what it is made of.* A more complicated hard drive, a denser gas, a cleverer encoding: none can exceed (5.15). The bound is saturated by a black hole (feed $E = Mc^2$ and $R = r_s$ into (5.15) and you recover, up to the $O(1)$ care the schematic lowering elides, the area law), which is why the solar-mass hole of Worked Example 5.2 was the maximum-entropy object of its size.

[MOTIVATED] — the Geroch argument is a physical motivation, not a rigorous theorem: the schematic steps (5.13)–(5.14) elide $O(1)$ factors and assume the box can be lowered to within $\sim R$ of the horizon without complication (buoyancy in the Unruh bath, finite-size and self-gravity corrections all enter a careful treatment). A rigorous, quantum field-theoretic version of (5.15) exists — Casini (2008) proves it as a consequence of the positivity of relative entropy (Chapter 2), with the energy $E$ read as a modular energy — and we flag it here because it is a first sighting of the pattern that organizes all of Part III: *a physical inequality restated as positivity of a relative entropy.* We import that rigorous version; the Geroch process is the picture that makes it plausible.

**Q 5.3 *"The bound* (5.15) *has no $G$ in it. How can it be a* gravitational *statement?"***

Sharp observation — and the answer is the first hint of something deep. Newton's constant cancelled between the Hawking temperature (which carries $1/G$) and the area law (which carries $G$), leaving (5.15) a relation among $\hbar$, $c$, $k_{\mathrm{B}}$, and the system's own $R, E$ alone. The bound was *derived* using gravity — we needed a black hole to lower the box into — but the result refers to no gravitational quantity. This is a recurring signature of the subject: gravity acts as the scaffolding that reveals an information-theoretic law, then withdraws, leaving a statement about information capacity that stands on its own. Casini's relative-entropy proof makes this explicit — it never mentions a black hole. The theory of this book takes that signature seriously: it reads gravity and quantum information as two faces of one structure (Chapter 8), and bounds like (5.15), gravitational in origin yet information-theoretic in form, are exactly what such a structure should produce.

## 5.4 The holographic bound

Bekenstein bounds the entropy by $RE$; but $E$ itself is bounded, because too much energy in too small a region *is* a black hole. Push that observation and the capacity of a region turns out to be set by its area alone — the most economical and most radical statement of the whole circle of ideas.

### 5.4.1 From energy to area

Consider a spatial region of size $R$, bounded by a surface of area $A \sim R^2$. How much entropy can it hold? Use the Bekenstein bound (5.15), but ask how large $E$ may be. If the region contains energy $E$ exceeding that of a black hole just filling it — roughly $E_{\max} \sim c^4 R/G$, the energy whose Schwarzschild radius equals $R$ — then it has already collapsed to a black hole, and the largest black hole that fits in the region has the horizon area $A$ of the region itself. For such a maximal hole the area law (5.12) gives directly

$$S_{\max}(\text{region}) \;=\; \frac{k_B A}{4\ell_P^2}. \tag{5.16}$$

This is the *holographic bound* (Bousso's covariant form makes the "area" precise as a light-sheet cross-section; Theorem I.2 of the framework derives it from the area law plus the Generalized Second Law along a light sheet). Its statement is stark: the maximum entropy in any region is its *bounding area* in Planck units, not its volume. A region cannot hold more information than would be written by tiling its *surface* with Planck-scale pixels, one nat per four of them.

The consequence, drawn out by 't Hooft and Susskind, is the *holographic principle*: the number of degrees of freedom in any region of space is bounded by its surface area, so a full description of the physics inside a volume can be encoded on its boundary — the world's information content is set by surfaces, not volumes, exactly as a hologram stores a three-dimensional image on a two-dimensional film. [MOTIVATED] — holography is a principle, extracted from the bounds above and given spectacular concrete form by the AdS/CFT correspondence (Maldacena 1997), but it is not a theorem proved from first principles for our universe. We adopt it as an organizing principle and import its rigorous realizations; the honest label stays warm.

**For the curious — How radical the area scaling is**

It is worth feeling the numbers. Ordinary matter is nowhere near the holographic bound. A region the size of a proton ($R \sim 10^{-15}$ m) has a bounding area of $\sim 10^{-29}$ m$^2$ and a holographic capacity of $\sim 10^{40}$ nats — vastly more than the handful of bits a proton actually carries. A region the size of the observable universe has a holographic capacity of $\sim 10^{122}$ nats, against the $\sim 10^{90}$ or so of actual particle entropy. Everyday physics lives deep below the ceiling; the ceiling is reached only by black holes, which is why they are the maximum-entropy objects and why they are the Rosetta stone of the subject. The gap between $R^3$ (naive counting) and $R^2$ (the true bound) is where quantum gravity hides its degree-of-freedom count — and it is the reason $S_{\text{gen}}$, built next, weighs an *area* against a bulk entropy rather than two volumes.

## 5.5 Generalized entropy and the Generalized Second Law

We now assemble the object this book is built on. The area law (5.12) gives a horizon an entropy; Chapter 2 gives ordinary matter a von Neumann entropy. A region with a horizon has both. The generalized entropy simply adds them.

### 5.5.1 The definition

**Definition 5.1** (Generalized entropy)**.** For a horizon (or, more generally, any codimension-two surface that splits space into an accessible "exterior" and an inaccessible "interior"), the generalized entropy is

$$S_{\text{gen}} \;=\; \frac{A}{4\ell_P^2} \;+\; S_{\text{ext}}, \qquad S_{\text{ext}} = -\operatorname{Tr}(\rho_{\text{ext}} \ln \rho_{\text{ext}}), \tag{5.17}$$

the horizon area in Planck units *plus* the ordinary von Neumann entropy (Chapter 2, in nats) of the quantum matter fields in the exterior region.

(We write $S_{\text{gen}}$ in nats here, dropping the explicit $k_{\text{B}}$; restore it, as in (5.12), whenever entropy meets a laboratory temperature.) The physical meaning: $S_{\text{gen}}$ is the *total* entropy an outside observer must assign — the entropy of the geometry (the area term) together with the entropy of everything she can see outside (the matter term). It is one ledger covering both, and it is the exact expression that becomes Axiom I of the theory in Chapter 9. Everything in Part III — the derivation of Einstein's equations from entanglement equilibrium (Chapter 10), the whole correspondence — rests on treating (5.17) as a fundamental object.

### 5.5.2 The Generalized Second Law

The ordinary second law says the entropy of a closed system never decreases. It fails, apparently, for a system with a horizon: drop an entropic box into a black hole and the *exterior* entropy $S_{\text{ext}}$ falls. The resolution — and the reason $S_{\text{gen}}$ is the right object — is that the horizon area rises to compensate. The *Generalized Second Law* (GSL) restores monotonicity to the combined ledger:

$$\frac{dS_{\text{gen}}}{dt} \;=\; \frac{d}{dt}\Big(\frac{A}{4\ell_{\text{P}}^2} + S_{\text{ext}}\Big) \;\geq\; 0. \tag{5.18}$$

Neither term alone is monotonic — the area can grow while exterior entropy shrinks, or vice versa — but their sum never decreases. The GSL is the statement that geometry and information trade off on *one* conserved-or-growing account, and it is what makes the identification of area with entropy physically compelling rather than a formal coincidence. [ASSUMED] — the GSL is not derived here; it is established physics (Bekenstein 1973; proved in wide generality by Wall 2012 for free and interacting fields), and we take it, as the framework does, as the irreducible input that gives the area law its thermodynamic teeth.

**Worked Example 5.3 — Dropping a box of entropy into a black hole**

Let us run the trade-off explicitly and check that (5.18) holds — this is the box argument of §5.3.1 seen from the $S_{\text{gen}}$ side. A box of energy $E$, size $R$, and entropy $S$ (saturating the Bekenstein bound, $S = 2\pi k_{\text{B}} RE/\hbar c$, so it is the hardest case) falls across the horizon.
*Before the crossing.* The box is outside, contributing its $S$ to $S_{\text{ext}}$; the horizon has area $A$. So $S_{\text{gen}}^{\text{before}} = A/4\ell_{\text{P}}^2 + S_{\text{ext}}^0 + S/k_{\text{B}}$, writing the box's contribution in nats.
*After the crossing.* The box is gone from the exterior: $S_{\text{ext}}$ drops back by $S/k_{\text{B}}$. But the box's energy $E$ has been added to the hole, so by $dM = E/c^2$ and (5.9) the horizon entropy rises by

$$\Delta\Big(\frac{A}{4\ell_{\text{P}}^2}\Big) = \frac{8\pi G}{\hbar c}\,M\,dM \;\gtrsim\; \frac{2\pi R\,E}{\hbar c} \;=\; \frac{S}{k_{\text{B}}},$$

where the inequality is exactly the Geroch estimate (5.14): the minimum area gain from delivering energy $E$ to a horizon, when the box is dropped from a proper distance $\sim R$, is at least the box's own entropy. So the net change is

$$\Delta S_{\text{gen}} = \underbrace{\Delta\Big(\frac{A}{4\ell_{\text{P}}^2}\Big)}_{\geq S/k_{\text{B}}} - \underbrace{\frac{S}{k_{\text{B}}}}_{\text{lost from } S_{\text{ext}}} \geq 0.$$

The horizon gains *at least* as much entropy as the exterior loses — the GSL holds, and it holds *because* the Bekenstein bound (5.15) caps how much entropy the box could have carried in the first place. The two results are the same statement read in opposite directions: the GSL implies the Bekenstein bound (§5.3.1), and the Bekenstein bound guarantees the GSL (here). [MOTIVATED] (schematic $O(1)$ factors as in §5.3.1).

▶ RUN IT: `TOOLS/verify-paper-numerics.py  (section N: ch5 GSL box-into-hole)`

### 5.5.3 What is physical and what is bookkeeping

One honesty is non-negotiable, and Chapter 2 prepared us for it. The split of $S_{\text{gen}}$ in (5.17) into an "area term" and a "matter term" is *not* physical on its own. The exterior entropy $S_{\text{ext}}$ of a quantum field is ultraviolet-divergent: as we saw in Chapter 2, a region is entangled with its complement at every length scale down to zero distance, so $S_{\text{ext}}$ diverges as the short-distance cutoff is removed. That divergence is not a disaster — it is absorbed, order by order, into a renormalization of Newton's constant $G$, which sits in the area term $A/4\ell_{\text{P}}^2 = Ac^3/4G\hbar$. In other words, moving the cutoff shifts entropy between the two terms: the area coefficient and the matter entropy are each *scheme-dependent*, and only their sum is fixed. The lesson is the same one Chapter 2 drew for relative entropy:

> **Only the sum, its differences, and its monotonicity are physical.** The value of $S_{\text{gen}}$ (after renormalization), the *change* $\Delta S_{\text{gen}}$ between two configurations, and the inequality $dS_{\text{gen}}/dt \geq 0$ are cutoff-independent and carry physical content. The individual area and matter terms are not separately meaningful — their split depends on the renormalization scheme. This is precisely why Chapter 2 insisted that relative entropy, a *difference*, survives in field theory where $S$ itself does not, and it is why Axiom II (Chapter 9) is written in relative entropy.

A second honesty, flagged so no reader mistakes what has been shown. The area law $S_{\text{area}} = A/4\ell_{\text{P}}^2$ itself is *not derived* in this chapter, nor anywhere in the framework, from something more primitive. We obtained it in §5.2.1 by *assuming* the Hawking temperature (itself imported via the Unruh–redshift shortcut) and running the first law. In the theory of Part III the area law is taken as the irreducible geometric *input* — the ingredient that Axiom I supplies and that everything else is built from. [ASSUMED] — the proportionality of horizon entropy to area is the one piece of gravitational input the axioms cannot manufacture; it is where geometry enters the information-theoretic structure, and the framework is honest that it is postulated, not proved. (What *is* derived, in Chapter 10, is the reverse: given the area law as $S_{\text{gen}}$, Einstein's equations follow from demanding entanglement equilibrium.)

**Q 5.4** ***"If the area law is just assumed and the split into two terms is scheme-dependent, what has this chapter actually established?"***

Three things, none of them small. *First*, that a black hole has a definite temperature (5.6) and therefore, by the first law, a definite entropy (5.12) — this is thermodynamics forced by the existence of $T_{\mathrm{H}}$, not an assumption. *Second*, that this entropy scales with area, not volume, and that the same area scaling bounds the information capacity of *any* region (§§5.3–5.4) — a structural fact about how much can be hidden behind a surface. *Third*, and most important for what follows, that geometry (area) and matter information (von Neumann entropy) are two entries in *one* monotonic ledger $S_{\text{gen}}$ (5.17). What we have *not* done is derive the area law from deeper axioms, and we say so plainly. The framework's wager, cashed in Chapter 10, is that $S_{\text{gen}}$ is the *right* fundamental object to postulate — that once you grant it, the whole of gravitational dynamics follows. This chapter earns the right to make that postulate by showing $S_{\text{gen}}$ is forced on us from three independent directions (thermodynamics, the bounds, the GSL); Chapter 9 elevates it to Axiom I; Chapter 10 shows what it buys.

## 5.6 An honest preview: the Page curve and quantum extremal surfaces

One thread remains loose, and it is the one that has driven the subject for the last decade. It belongs here because it uses every object of this chapter, but its resolution belongs to Chapters 6–7 and Chapter 13, so we state it qualitatively and flag it honestly.

Let a black hole evaporate (the "Only the smallest holes are hot" aside of §5.1). As it radiates, its horizon area shrinks, so its Bekenstein–Hawking entropy (5.12) falls monotonically to zero. But the radiation it emits carries entropy *out*, and Hawking's original calculation says that entropy only grows. Track the *fine-grained* (von Neumann) entropy of the radiation. If it grew without bound while the hole's own entropy fell to zero, the final radiation would carry more entropy than the hole ever had — and the pure initial state would have become mixed, destroying unitarity (the information-conservation of Chapter 2). This is the black-hole information puzzle.

The resolution, sharpened in 2019–2020, is the *Page curve*: the fine-grained entropy of the radiation does not grow forever. It rises, turns over at the "Page time" when the hole has emitted about half its entropy, and falls back to zero as the hole finishes evaporating — consistent with unitary evolution of a pure state. The turnover is enforced by a computational prescription that is exactly a $S_{\text{gen}}$-extremization: the fine-grained entropy of a region is obtained by finding the *quantum extremal surface* (QES) that extremizes the generalized entropy (5.17) — area plus exterior von Neumann entropy — and, past the Page time, the extremum jumps to a surface just behind the horizon that includes an "island" of the interior in the radiation's entropy account. The bookkeeping is $S_{\text{gen}}$, the very object of §5.5, promoted to a variational principle.

[MOTIVATED]/imported — the Page curve and the QES/island prescription are results we import from the recent literature (Penington; Almheiri–Engelhardt– Marolf–Maxfield, 2019), not derivations of this book. We flag them for two reasons. *First*, they show that $S_{\text{gen}}$ is not a heuristic but the operational object that computes fine-grained entropies in gravity — further evidence that Axiom I is postulating the right thing. *Second*, they expose a gap this chapter cannot fill: the QES prescription requires a *finite*, well-defined von Neumann entropy for a subregion, and Chapter 2 warned us that $S_{\text{ext}}$ is divergent. Making sense of a finite entropy for an observer with access to only part of spacetime is exactly the problem that forces the algebraic machinery of Chapters 6–7 (von Neumann algebras, the crossed product, the Type II observer algebra in which entropy becomes finite and well-defined), and it is the setting of the one open construction R1 (Chapter 13). The reader who wants to know why the next two chapters leave physics for functional analysis has the answer here: a finite $S_{\text{ext}}$, hence a well-posed $S_{\text{gen}}$, demands it.

## Chapter FAQ

**Q: The Hawking temperature came from Unruh applied "locally" at the horizon. Isn't that hand-waving?** It is a shortcut, and we labelled it so. The rigorous derivation is Hawking's: quantize a field on the spacetime of a collapsing star and compute the outgoing particle flux at late times; the result is a thermal spectrum at exactly (5.6). Our route — local Unruh temperature (5.1) times the horizon redshift (5.4) — reproduces the temperature because the surface gravity $\kappa$ is precisely the finite horizon-limit of (redshift)$\times$(local acceleration), (5.5). It gets the physics and the number right by leaning on the equivalence principle; it does not replace the full field-theory calculation, which we import.

**Q: Why one quarter? Where does the $1/4$ in $A/4\ell_P^2$ come from?** From the first law, once the Hawking temperature is fixed. The integration in (5.9)–(5.12) produces the $1/4$ automatically: it is $8\pi \to 4$ from $\int M\,dM = \frac{1}{2}M^2$, then the constants of $T_{\text{H}}$ and $A(M)$. There is no freedom to adjust it — fix $T_{\text{H}}$ and the coefficient is forced. Bekenstein's original 1973 argument fixed the proportionality of entropy to area but not the coefficient; Hawking's temperature (1975) pinned it to exactly $1/4$.

**Q: Does the area law hold for horizons other than black holes — Rindler, cosmological?** Yes, and this generality is the whole point. The derivation used only that a horizon has a surface gravity (hence a temperature) and obeys the first law — nothing specific to black holes. A Rindler horizon (Chapter 4) and a de Sitter cosmological horizon (Chapter 14) carry entropy $A/4\ell_P^2$ by the same argument. That universality is what lets Chapter 10 apply $S_{\text{gen}}$ to a *local* Rindler horizon at every point of spacetime and extract Einstein's equations, and what lets Chapter 14 read a MOND acceleration scale off the Hubble horizon.

**Q: Is the holographic principle really established, or is it a hope?** It is a well-motivated *principle* with one rigorous, fully worked realization (AdS/CFT) and no counterexample, but it is not a theorem for our universe. That is why we carry it [MOTIVATED]. The bounds of §§5.3– 5.4 are the honest, cutoff-aware content; "the world is a hologram" is the interpretive extrapolation, powerful and probably right but not proved. The framework of this book uses holography as a guiding intuition, not as a load-bearing axiom.

## Exercises

**Exercise 5.1.** Suppose the Earth ($M_\oplus = 5.972 \times 10^{24}$ kg) were compressed into a black hole. (a) Compute its Schwarzschild radius $r_s$ and comment on its size. (b) Compute its Hawking temperature from (5.6). (c) Compute its Bekenstein–Hawking entropy in units of $k_B$ from (5.12). Compare $S_{BH}/k_B$ with the value $1.05 \times 10^{77}$ for the solar-mass hole and check that the ratio scales as $M^2$.

**Exercise 5.2.** (a) Apply the Bekenstein bound (5.15) to a human being: take $R = 1$ m and $E = Mc^2$ with $M = 70$ kg. How many nats of information can a person maximally store? Compare with a rough estimate of the information actually in a human body ($\sim 10^{27}$ atoms, a few bits each). (b) At what radius $R$, holding $E = Mc^2$ fixed, would the person saturate the bound — i.e. become a black hole? Show this radius is $r_s$ up to an $O(1)$ factor, confirming that a black hole saturates the Bekenstein bound.

## Solutions

*Solution (Exercise 5.1).* (a) $r_s = 2GM_\oplus/c^2 = 2(6.67430 \times 10^{-11})(5.972 \times 10^{24})/(2.99792458 \times 10^8)^2 = 8.87 \times 10^{-3}$ m — about nine millimetres, the size of a small marble. Earth's entire mass would fit inside a horizon you could hold in your hand. (b) $T_H = \hbar c^3/(8\pi GM_\oplus k_B) = 6.2 \times 10^{-8}\,\mathrm{K} \times (M_\odot/M_\oplus) = 6.2 \times 10^{-8} \times (1.989 \times 10^{30}/5.972 \times 10^{24}) = 2.0 \times 10^{-2}$ K — about twenty millikelvin, since $T_H \propto 1/M$ and Earth is $\sim 3 \times 10^5$ times lighter than the Sun, so it is correspondingly hotter (but still far below the CMB). (c) $A = 4\pi r_s^2 = 4\pi(8.87 \times 10^{-3})^2 = 9.88 \times 10^{-4}\,\mathrm{m}^2$, and with $\ell_P^2 = 2.612 \times 10^{-70}\,\mathrm{m}^2$, $S_{BH}/k_B = A/4\ell_P^2 = 9.46 \times 10^{65}$. The ratio to the solar value is $9.46 \times 10^{65}/1.05 \times 10^{77} = 9.0 \times 10^{-12} = (M_\oplus/M_\odot)^2$, confirming $S_{BH} \propto M^2$ from (5.9). Heavier holes hold vastly more information.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch5 Earth-hole values)`

*Solution (Exercise 5.2).* (a) $E = Mc^2 = 70 \times (2.99792458 \times 10^8)^2 = 6.29 \times 10^{18}$ J. The bound (5.15) gives

$$\frac{S}{k_B} = \frac{2\pi RE}{\hbar c} = \frac{2\pi\,(1)\,(6.29 \times 10^{18})}{(1.054571817 \times 10^{-34})(2.99792458 \times 10^8)} = 1.25 \times 10^{45}\ \text{nats}.$$

A person can store at most $\sim 10^{45}$ nats. The information actually present — $\sim 10^{27}$ atoms times a handful of bits of state each, $\sim 10^{28}$ nats — is seventeen orders of magnitude below the ceiling. Ordinary matter is nowhere near saturating the bound, exactly as the holographic-capacity estimates of §5.4 anticipated. (b) Saturation of (5.15) at fixed $E = Mc^2$ occurs when the horizon area law is reached; setting $S/k_B = 2\pi RE/\hbar c$ equal to the Bekenstein–Hawking value $A/4\ell_P^2$ with $A = 4\pi R^2$ and $E = Mc^2$ and solving for $R$ returns $R = 2GM/c^2 = r_s$ up to the $O(1)$ factor the schematic lowering left undetermined. For $M = 70$ kg, $r_s = 2GM/c^2 = 1.04 \times 10^{-25}$ m — a hundred-billionth of a proton radius. Compressing a person to that size makes them a black hole that saturates the Bekenstein bound; the black hole is the maximum-entropy object, as claimed in Worked Example 5.2.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch5 Bekenstein-human values)`

**What to carry forward**

1. A black hole is a thermodynamic object. Unruh (5.1) applied at the Schwarzschild horizon, redshifted out (5.4), gives the Hawking temperature $T_H = \hbar c^3/8\pi GMk_B$ ($\approx 6 \times 10^{-8}$ K for a solar mass); the first law then forces the Bekenstein–Hawking

entropy $S_{\text{area}} = k_B A/4\ell_P^2$ ($\approx 10^{77} k_B$ for a solar mass). Entropy scales with *area*, not volume — the radical break.

2. That area scaling is universal. The Bekenstein bound $S \leq 2\pi k_B RE/\hbar c$ [MOTIVATED] caps any system's information; the holographic bound $S_{\max} = k_B A/4\ell_P^2$ [MOTIVATED] caps any region's by its bounding surface — the world's information is set by surfaces, not volumes.

3. Geometry and matter information share one ledger: $S_{\text{gen}} = A/4\ell_P^2 + S_{\text{ext}}$ (5.17), monotonic under the GSL $dS_{\text{gen}}/dt \geq 0$ [ASSUMED]. Only the sum, its differences, and its monotonicity are physical; the split is scheme-dependent (Chapter 2's relative-entropy lesson), and the area law itself is [ASSUMED] input. This $S_{\text{gen}}$ is Axiom I (Chapter 9); the Page curve and QES prescription [MOTIVATED] show it computes real fine-grained entropies and demand the algebraic machinery of Chapters 6–7.

# 6 Algebras of Observables

**Synopsis.** Two chapters ago the honest description of an observer was her density matrix; one chapter ago that description broke, because the von Neumann entropy of a region bounded by a horizon came out infinite. Both troubles point the same way: stop treating the global Hilbert space as primary, and put the *algebra of observables the observer can actually reach* first. We build that idea from the floor, entirely on finite matrices: what an operator algebra is, what its commutant is, what makes it a *factor* (an algebra with no leftover classical labels). We construct the Hilbert space from the algebra and a state — the GNS construction — explicitly on $2 \times 2$ matrices, and watch the purification of Chapter 2 reappear as a theorem. Then we classify: type I is ordinary quantum mechanics; type II has a trace but no smallest projection, so "dimension" becomes continuous; type III has *no trace at all*, hence no density matrix and no finite entropy — only relative entropy survives. The chapter ends on the fact the whole theory must digest: the observables in a region of relativistic quantum field theory form a type III algebra, which is exactly why Chapter 5's entropy diverged, and exactly what Chapter 7's crossed product will repair.

**You will be able to.** compute commutants and centers of matrix subalgebras by hand; decide whether a given algebra is a factor; run the GNS construction on a $2 \times 2$ state and identify its cyclic vector with a purification; state the type I/II/III classification through its trace and projection lattice, and explain from a biased infinite tensor product why a type III factor can carry no trace at all.

**Prerequisites.** Chapters 2 (density matrices, partial trace, entropy, relative entropy, purification) and 5 (the horizon-entropy divergence that motivates all of this). Linear algebra: Hermitian matrices, eigenvalues, block matrices, the trace. No prior contact with operator algebras is assumed — every abstraction is built here from matrices you can multiply on paper.

## 6.1 Why the algebra, and not the state

Two threads from earlier chapters are about to be tied together, and it is worth naming them before we touch a single new definition, because the whole apparatus of this chapter exists only to serve them.

*Thread one, from Chapter 2.* When an observer holds only part of the world, her honest description is a reduced density matrix, and we built it by asking one operational question: what statistics does she get for the observables *she can measure*? Everything about her situation was fixed not by the global state vector — she has no access to that — but by the set of operators she can apply, namely $A \otimes \mathbb{1}_B$ for her subsystem $A$. The state entered only through its restriction to that set. We noted there, in passing, that "what an observer can know is determined by the algebra of observables she can reach." This chapter cashes that remark.

*Thread two, from Chapter 5.* When we asked how much information can hide behind a horizon, the entanglement entropy of the region came out *infinite*: every length scale down to zero contributes, because a quantum field entangles a region with its complement at arbitrarily short distances. The density matrix $\rho_A$ of a region does not exist as a respectable object; $S(\rho_A) = -\mathrm{Tr}\,\rho_A \ln \rho_A$ diverges and takes the whole density-matrix formalism down with it. Yet the physics of a horizon is perfectly finite — a black hole has a temperature and an entropy we can write down. Something in our *bookkeeping*, not in the world, has failed.

Put the two threads side by side and a strategy suggests itself. In the first, the algebra of reachable observables was the thing that actually determined the physics, with the global state relegated to a supporting role. In the second, the global-state object (the density matrix) is precisely what broke. So let us promote the algebra to the leading role and see whether it survives where the density matrix did not. It will — and more than that, the *kind* of algebra a region carries will turn out to be the exact diagnosis of why the density matrix failed.

**Q 6.1** ***"This sounds like a lot of new mathematics to fix a bookkeeping problem. Why not just put a short-distance cutoff on the entropy and move on?"***

Because the cutoff hides the physics we are chasing rather than revealing it. Chapter 5 already showed the cost: a cutoff makes the entanglement entropy finite but imports an arbitrary length, and the famous "area term" $A/4\ell_{\mathrm{P}}^2$ then depends on where you drew the line. Worse, it disguises a *structural* fact as a mere large number. The claim of this chapter is that the divergence is not an accident to be regulated away but a signature: it tells you that the region's observables form an algebra of a specific type (type III) that *provably* admits no density matrix and no finite entropy, for reasons as clean as the ones that forbid a largest integer. Once you see the structure, the right repair is forced — it is the crossed product of Chapter 7, which enlarges the algebra to one that *does* carry a trace — and the area law comes out finite and meaningful. A cutoff would have papered over exactly the object the theory needs. We spend this chapter learning to read that object, and we learn it on finite matrices, where nothing is subtle, before letting the dimension run to infinity where the surprises live.

Here, then, is the plan of the chapter, stated as a goal we can check we have reached. *We will classify the possible algebras of observables, and identify which kind a region of spacetime carries.* The classification has exactly three families (with sub-families), and every one of them appears already in finite or barely-infinite dimensions, so we can meet each concretely before trusting any general statement.

## 6.2 Algebras, concretely

### 6.2.1 What an algebra of observables is

Start operationally. An observer has a collection of measurable quantities. Each is represented, as in ordinary quantum mechanics, by an operator on a Hilbert space. Two demands on this collection are physical, not mathematical. First, if she can measure $A$ and she can measure $B$, she can measure the observable built by applying one apparatus after another, or by adding readings, or by scaling a dial: the collection should be closed under products, sums, and scalar multiples. Second, every physical observable is Hermitian, and the adjoint operation $A \mapsto A^\dagger$ (which turns bras into kets and back) should not lead outside the collection. A set of operators closed under these operations is what we will call an algebra of observables. We state it once, precisely, and then work only with examples.

**Definition 6.1** ($*$-algebra of operators)**.** Let $\mathcal{H}$ be a Hilbert space and $B(\mathcal{H})$ the set of all bounded operators on it. A *$*$-algebra* $\mathcal{M} \subseteq B(\mathcal{H})$ is a collection of operators that is closed under (i) addition and scalar multiplication (a complex vector space), (ii) operator multiplication (if $A, B \in \mathcal{M}$ then $AB \in \mathcal{M}$), and (iii) the adjoint (if $A \in \mathcal{M}$ then $A^\dagger \in \mathcal{M}$). We take $\mathcal{M}$ to contain the identity $\mathbb{1}$. A *von Neumann algebra* is a $*$-algebra that is also closed in a topological sense we make precise in §6.3.2; in finite dimensions every $*$-algebra containing $\mathbb{1}$ is already a von Neumann algebra, so for now the distinction costs us nothing.

The physical reading of each clause: (i) says a measurable quantity can be rescaled and combined; (ii) says apparatuses compose; (iii) says the Hermitian observables — the actually-measurable ones — are the self-adjoint part of a structure that, for algebraic convenience, we let include their complex combinations too. The observables proper are the $A$ with $A^\dagger = A$; the $*$-algebra is the smallest self-consistent world containing them.

**Worked Example 6.1 — The basic example: all $n \times n$ matrices**

Take $\mathcal{H} = \mathbb{C}^n$, so operators are $n \times n$ complex matrices, and let $\mathcal{M} = M_n(\mathbb{C})$ be *all* of them. Check the three clauses in one breath: matrices add, scale, multiply, and transpose-conjugate, all staying $n \times n$; the identity matrix is present. So $M_n(\mathbb{C})$ is a $*$-algebra — the largest one on $\mathbb{C}^n$, since there are no matrices left out. This is the algebra of *ordinary, complete* quantum mechanics on an $n$-level system: an observer who can measure every Hermitian $n \times n$ matrix has access to $M_n(\mathbb{C})$, and nothing is hidden from her. Every qubit observable of Worked Example 2.1 lives in $M_2(\mathbb{C})$; a spin-$j$ particle lives in $M_{2j+1}(\mathbb{C})$. Hold this as the reference point: as we now build smaller algebras and, later, stranger ones, the question will always be "how does this differ from the full matrix algebra?"

### 6.2.2 Subalgebras: observables that cannot tell blocks apart

An observer with *partial* access holds not all of $M_n(\mathbb{C})$ but a subalgebra. The most transparent example, and the one every later structure is built from, is the block-diagonal one.

**Worked Example 6.2 — Block-diagonal observables**

Split $\mathbb{C}^4$ into two blocks, a first pair of coordinates and a second pair, and consider the operators that never mix them:

$$\mathcal{M} = \left\{ \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} : A, B \in M_2(\mathbb{C}) \right\} \subset M_4(\mathbb{C}). \tag{6.1}$$

Two such matrices add and multiply block by block ($\mathrm{diag}(A, B)\,\mathrm{diag}(A', B') = \mathrm{diag}(AA', BB')$), the adjoint acts block by block ($\mathrm{diag}(A, B)^\dagger = \mathrm{diag}(A^\dagger, B^\dagger)$), and $\mathbb{1} = \mathrm{diag}(\mathbb{1}_2, \mathbb{1}_2)$ is present. So (6.1) is a $*$-algebra. Physically it is the algebra of an observer who can perform any measurement *within* block 1 and any *within* block 2, but owns no apparatus that connects the two — no operator with a nonzero entry in the off-diagonal corners. The two blocks are, for her, separate worlds she cannot bridge.

The label "block 1 versus block 2" is a piece of information the observer of Worked Example 6.2 *can* read but cannot *change*. Every operator she has is block-diagonal, so it commutes with the projector

$$P_1 \;=\; \begin{pmatrix} \mathbb{1}_2 & 0 \\ 0 & 0 \end{pmatrix}, \tag{6.2}$$

which asks "are we in block 1?" She can measure $P_1$ (it is $\mathrm{diag}(\mathbb{1}_2, 0) \in \mathcal{M}$), it takes only the values 0 and 1, and no measurement she owns disturbs it. An observable like this — one that lies in the algebra, is not a multiple of the identity, and commutes with *everything* in the algebra — is a *classical* degree of freedom sitting inside a quantum theory. In quantum mechanics such an observable is called a *superselection rule*: it labels sectors between which no observer-accessible operator produces interference. Block 1 versus block 2 is exactly a superselection label. The presence or absence of such labels is what the whole type-classification will ultimately turn on, so we now build the tool that detects them.

## 6.3 The commutant, the double commutant, and factors

### 6.3.1 The commutant

Given an algebra $\mathcal{M}$ of observables an observer can reach, which operators does *nobody in her world* disturb? Precisely those that commute with everything she has.

**Definition 6.2** (Commutant)**.** The *commutant* of $\mathcal{M} \subseteq B(\mathcal{H})$ is

$$\mathcal{M}' \;=\; \{\, X \in B(\mathcal{H}) \;:\; XA = AX \text{ for all } A \in \mathcal{M} \,\}. \tag{6.3}$$

It is the set of all operators (on the same $\mathcal{H}$) that commute with every element of $\mathcal{M}$.

The commutant is itself always a $*$-algebra containing $\mathbb{1}$: if $X$ and $Y$ each commute with all of $\mathcal{M}$, so do $X + Y$, $XY$, $cX$, and (taking adjoints of $XA = AX$, using that $\mathcal{M}$ is closed under $\dagger$) $X^\dagger$. The physical meaning is the reason the concept exists: in the bipartite setup of Chapter 2, where the observer's algebra was $\mathcal{M} = M_{d_A}(\mathbb{C}) \otimes \mathbb{1}_B$ (all operators on her half), the commutant is $\mathcal{M}' = \mathbb{1}_A \otimes M_{d_B}(\mathbb{C})$ — the operators on the *other* half, the ones her distant partner controls. Commutant means "the observables complementary to mine." This is why the object will matter for a region of spacetime: $\mathcal{M}'$ will be the algebra of the *complementary* region, across the horizon.

**Worked Example 6.3 — Commutant of a block algebra**

Compute the commutant of (6.1) inside $M_4(\mathbb{C})$. We want all $X$ with $X\,\mathrm{diag}(A, B) = \mathrm{diag}(A, B)\,X$ for *every* $A, B \in M_2(\mathbb{C})$. Write $X$ in the same $2 \times 2$ block form,

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}, \qquad X_{ij} \in M_2(\mathbb{C}),$$

and impose the commutation. Multiplying out,

$$X\,\mathrm{diag}(A,B) = \begin{pmatrix} X_{11}A & X_{12}B \\ X_{21}A & X_{22}B \end{pmatrix}, \qquad \mathrm{diag}(A,B)\,X = \begin{pmatrix} AX_{11} & AX_{12} \\ BX_{21} & BX_{22} \end{pmatrix}.$$

Equating blocks: $X_{11}A = AX_{11}$ for all $A$, so $X_{11}$ commutes with every $2 \times 2$ matrix and is therefore a multiple of the identity, $X_{11} = \lambda \mathbb{1}_2$ (only scalars commute with all of $M_2$ — a fact we prove as Lemma 6.3 below). Likewise $X_{22} = \mu \mathbb{1}_2$. The off-diagonal conditions are $X_{12}B = AX_{12}$ for *all* $A, B$ independently; taking $A = \mathbb{1}, B = 0$ forces $X_{12} = 0$, and similarly $X_{21} = 0$. Hence

$$\mathcal{M}' \;=\; \left\{ \begin{pmatrix} \lambda \mathbb{1}_2 & 0 \\ 0 & \mu \mathbb{1}_2 \end{pmatrix} \;:\; \lambda, \mu \in \mathbb{C} \right\}. \tag{6.4}$$

The commutant is two-dimensional: it consists exactly of the operators that are constant on each block. In words, the only things nobody in the block-observer's world can disturb are "how much you are in block 1" and "how much in block 2" — the superselection label itself, and nothing else.

We used a small fact twice; let us prove it once, because it is the workhorse of every commutant computation in the chapter.

**Lemma 6.3** (Only scalars commute with everything)**.** *If $X \in M_n(\mathbb{C})$ satisfies $XA = AX$ for all $A \in M_n(\mathbb{C})$, then $X = \lambda \mathbb{1}$ for some $\lambda \in \mathbb{C}$. Equivalently, $M_n(\mathbb{C})' = \mathbb{C}\,\mathbb{1}$.*

*Derivation.* Let $E_{ij}$ be the matrix with a 1 in row $i$, column $j$, and zeros elsewhere; these span $M_n$, so it suffices that $X$ commute with each $E_{ij}$. Compute the two products entry by entry: $(XE_{ij})_{kl} = X_{ki}\delta_{jl}$ places column $i$ of $X$ into column $j$; $(E_{ij}X)_{kl} = \delta_{ki}X_{jl}$ places row $j$ of $X$ into row $i$. Setting $XE_{ij} = E_{ij}X$ and reading off entries: the diagonal choice $i = j$ gives $X_{ki} = 0$ for $k \neq i$ (off-diagonal entries of $X$ vanish) and, comparing $i = j$ with $i' = j'$, gives $X_{ii} = X_{jj}$ (all diagonal entries equal). So $X$ is a common value $\lambda$ times the identity. The reading: an operator that is invisible to every measurement — it commutes with all of them — carries no information at all beyond an overall scale.

### 6.3.2 The double commutant

Take the commutant of the commutant. This looks like it should spin off into an infinite tower — $\mathcal{M}', \mathcal{M}'', \mathcal{M}''', \ldots$ — but it does not, and the reason it stops is the single most important structural fact about these algebras.

**Theorem 6.4** (Double commutant, finite-dimensional form)**.** *Let $\mathcal{M} \subseteq M_n(\mathbb{C})$ be a $*$-algebra containing $\mathbb{1}$. Then $\mathcal{M}'' = \mathcal{M}$: taking the commutant twice returns the original algebra.*

*Derivation.* That $\mathcal{M} \subseteq \mathcal{M}''$ is immediate and needs no hypotheses: every $A \in \mathcal{M}$ commutes with everything in $\mathcal{M}'$ (that is what $\mathcal{M}'$ was defined to do), so $A$ lies in the commutant of $\mathcal{M}'$, which is $\mathcal{M}''$. The content is the reverse inclusion, $\mathcal{M}'' \subseteq \mathcal{M}$, and here we use the $*$-structure and the identity.

The clean route is through projections. A finite-dimensional $*$-algebra $\mathcal{M}$ with $\mathbb{1}$ has a structure theorem we can take as known from linear algebra (it is the statement that a set of matrices closed under products and adjoints is simultaneously block-diagonalizable): there is a basis of $\mathbb{C}^n$ in which

$$\mathcal{M} \;\cong\; \bigoplus_\alpha \big(M_{n_\alpha}(\mathbb{C}) \otimes \mathbb{1}_{m_\alpha}\big), \tag{6.5}$$

a direct sum of "full matrix algebra $M_{n_\alpha}$ acting on the first factor, doing nothing on a multiplicity space of dimension $m_\alpha$." (The block algebra (6.1) is the case of two summands, each $M_2 \otimes \mathbb{1}_1$.) Now compute the commutant of (6.5) summand by summand using Lemma 6.3 on each $M_{n_\alpha}$ factor: the operators commuting with $M_{n_\alpha} \otimes \mathbb{1}_{m_\alpha}$ are exactly $\mathbb{1}_{n_\alpha} \otimes M_{m_\alpha}(\mathbb{C})$ (scalars on the first factor, anything on the multiplicity space), so

$$\mathcal{M}' \;\cong\; \bigoplus_\alpha \big(\mathbb{1}_{n_\alpha} \otimes M_{m_\alpha}(\mathbb{C})\big). \tag{6.6}$$

The commutant has the *same block structure with the roles of the two tensor factors swapped.* Taking the commutant a second time swaps them back: $\mathcal{M}'' \cong \bigoplus_\alpha (M_{n_\alpha} \otimes \mathbb{1}_{m_\alpha}) = \mathcal{M}$. The swap is an involution, and two swaps are the identity.

The double-commutant theorem is why von Neumann algebras are the right objects. It says: the algebras that are their own double commutant are precisely those singled out by a topological closure (in infinite dimensions, closure in the weak operator topology; in finite dimensions, no closure is needed and *every* $*$-algebra with $\mathbb{1}$ qualifies). We will use it as a working definition — *a von Neumann algebra is one satisfying* $\mathcal{M}'' = \mathcal{M}$ — and its physical force is this: an algebra of observables is completely pinned down by its commutant, that is, by which complementary observables it fails to include. The observer's world and its complement determine each other exactly.

### 6.3.3 The center, and the definition of a factor

We can now isolate the classical labels of §6.2.2 in one formula. An observable is classical-within-$\mathcal{M}$ if it lies in $\mathcal{M}$ and commutes with all of $\mathcal{M}$ — that is, if it lies in both $\mathcal{M}$ and $\mathcal{M}'$.

**Definition 6.5** (Center; factor)**.** The *center* of $\mathcal{M}$ is

$$\mathcal{Z}(\mathcal{M}) \;=\; \mathcal{M} \cap \mathcal{M}', \tag{6.7}$$

the operators in $\mathcal{M}$ that commute with everything in $\mathcal{M}$. It always contains the multiples of the identity, $\mathbb{C}\,\mathbb{1}$. When it contains *nothing more* — when $\mathcal{Z}(\mathcal{M}) = \mathbb{C}\,\mathbb{1}$ — the algebra is called a *factor*.

A factor is an algebra with *no classical observables*: no superselection labels, no central quantity that every measurement leaves untouched, nothing that quietly commutes with the entire theory except the trivial constant. Read the two running examples through this definition:

- The full matrix algebra $M_n(\mathbb{C})$ is a *factor*: by Lemma 6.3 its commutant is $\mathbb{C}\,\mathbb{1}$, so the center $\mathcal{Z} = \mathcal{M} \cap \mathcal{M}' = \mathbb{C}\,\mathbb{1}$ is trivial. Ordinary complete quantum mechanics has no built-in classical labels — as it should.
- The block algebra (6.1) is *not* a factor. Its commutant (6.4) consists of the block scalars $\mathrm{diag}(\lambda\mathbb{1}_2, \mu\mathbb{1}_2)$, and the ones that also lie in $\mathcal{M}$ are all of them (they are block-diagonal). So

$\mathcal{Z} = \{\mathrm{diag}(\lambda \mathbb{1}_2, \mu \mathbb{1}_2)\}$ is *two*-dimensional. That extra dimension is the block label $P_1$ of (6.2) — the one classical bit "which block" — shown here as a nontrivial center.

**Q 6.2** ***"Why should I care whether an algebra is a factor? It sounds like a technicality about which matrices commute."***

Because a factor is the honest arena of a single quantum system, and the whole type-classification — the payoff of the chapter — is a classification of *factors*. Here is the physical stake. A nontrivial center is a classical observable hiding inside your quantum description: a quantity you can read but never superpose across, a label that secretly splits your one system into independent sectors. If a region of spacetime carried a nontrivial center, it would mean the region came pre-equipped with classical data — and then "quantum gravity in that region" would already be smuggling in classical structure through the back door, which is exactly the sort of thing this book is trying not to do. The claim we are building toward is that a region of spacetime carries a *factor* (no classical labels), and moreover a factor of a very particular, trace-free type. To even state that claim we need the word "factor" and we need the trace, which is next. A factor is where you go when you want to ask "what kind of quantum system is this, stripped of any classical bookkeeping?" — and the answer, for a region of spacetime, is the surprise the chapter is organized around.

## 6.4 Building the Hilbert space from the algebra: GNS

So far the Hilbert space came first and the algebra lived on it. If the algebra is to be primary, we must be able to run this backwards: given an algebra and a state on it — with no Hilbert space in hand — *construct* the Hilbert space. This is the Gel'fand–Naimark–Segal (GNS) construction, and it is the technical heart of "the algebra comes first." We build it once, in the abstract, and then do it fully explicitly on $2 \times 2$ matrices so it is concrete rather than incantatory.

### 6.4.1 States on an algebra

First, what is a "state" when there is no Hilbert space to hold a vector or a density matrix? Operationally, a state is the only thing an observer ever actually reports: an assignment of an expectation value to each observable.

**Definition 6.6** (State on an algebra)**.** A *state* on a $*$-algebra $\mathcal{M}$ is a linear map $\omega : \mathcal{M} \to \mathbb{C}$ that is (i) *positive*, $\omega(A^\dagger A) \geq 0$ for all $A$ (variances are non-negative), and (ii) *normalized*, $\omega(\mathbb{1}) = 1$ (probabilities sum to one).

In ordinary quantum mechanics every state has the form $\omega(A) = \mathrm{Tr}(\rho A)$ for a density matrix $\rho$, and Definition 6.6 is just the list of properties that $\mathrm{Tr}(\rho\,\cdot)$ enjoys, stripped of the density matrix itself. The point of the abstraction — and the reason it is not idle — is that in type III, where no $\rho$ exists, the state $\omega$ survives as a perfectly good linear functional. The state is more robust than the density matrix that usually represents it.

### 6.4.2 The construction

Given $\omega$ on $\mathcal{M}$, we build a Hilbert space $\mathcal{H}_\omega$, a representation of the algebra as operators on it, and a distinguished vector $|\Omega\rangle$ that reproduces $\omega$ as an ordinary expectation value. The idea is disarmingly direct: *use the algebra itself as the raw material of the Hilbert space*, and let $\omega$ supply the inner product.

*Derivation. Step 1 — vectors from operators.* Take the elements of $\mathcal{M}$ and declare each one to be a candidate vector; write $|A\rangle$ for the vector associated to the operator $A \in \mathcal{M}$. This is a complex vector space already (operators add and scale).

*Step 2 — inner product from the state.* Define

$$\langle A|B\rangle \;\equiv\; \omega(A^\dagger B). \tag{6.8}$$

This is linear in $B$, antilinear in $A$, and non-negative on the diagonal, $\langle A|A\rangle = \omega(A^\dagger A) \geq 0$, exactly by positivity of $\omega$. It is the inner product $\omega$ was born to provide.

*Step 3 — quotient out the null directions.* The form (6.8) may be degenerate: there can be nonzero operators $N$ with $\omega(N^\dagger N) = 0$ ("null" directions the state cannot see). These form a subspace, and we simply discard them — identify $|A\rangle$ and $|A+N\rangle$ for every null $N$. On the quotient, (6.8) is a genuine (positive-definite) inner product; complete it if the dimension is infinite. Call the result $\mathcal{H}_\omega$.

*Step 4 — the algebra acts by left multiplication.* Let $\mathcal{M}$ act on its own vectors by multiplying from the left,

$$\pi(A)\,|B\rangle \;\equiv\; |AB\rangle\,. \tag{6.9}$$

This respects products, $\pi(A)\pi(A')\,|B\rangle = |AA'B\rangle = \pi(AA')\,|B\rangle$, and one checks $\pi(A^\dagger) = \pi(A)^\dagger$ against (6.8), so $\pi$ is a genuine $*$-representation of the algebra as operators on $\mathcal{H}_\omega$.

*Step 5 — the cyclic vector.* Single out the vector coming from the identity,

$$|\Omega\rangle \;\equiv\; |\mathbb{1}\rangle\,. \tag{6.10}$$

It reproduces the state as an ordinary quantum expectation value,

$$\langle\Omega|\,\pi(A)\,|\Omega\rangle \;=\; \langle\mathbb{1}|A\rangle \;=\; \omega(\mathbb{1}^\dagger A) \;=\; \omega(A), \tag{6.11}$$

and it is *cyclic*: acting on it with the algebra, $\pi(A)\,|\Omega\rangle = |A\rangle$, sweeps out every vector, so $|\Omega\rangle$ is a single vector from which the whole space is generated.

> The GNS construction turns any state $\omega$ on any $*$-algebra $\mathcal{M}$ into a concrete Hilbert space $\mathcal{H}_\omega$, a representation $\pi$, and a cyclic vector $|\Omega\rangle$ with $\langle\Omega|\,\pi(A)\,|\Omega\rangle = \omega(A)$. *The abstract state becomes a bona fide vector-state expectation value* — without a Hilbert space ever having been assumed at the start.

### 6.4.3 GNS on a $2\times 2$ state, explicitly

Abstraction discharged; now the concrete case, done in full, so the machinery above is a computation you have watched happen.

**Worked Example 6.4 — GNS for a thermal qubit**

Let $\mathcal{M} = M_2(\mathbb{C})$ and let the state be the one given by a diagonal density matrix

$$\rho = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}, \qquad 0 < p < 1, \qquad \omega(A) = \mathrm{Tr}(\rho A).$$

(For a two-level system of energies 0 and $E$ in a heat bath this is the thermal state with $p = 1/(1+e^{-E/kT})$; the label "thermal" will matter in Chapter 7.)

*The Hilbert space.* Vectors are $2 \times 2$ matrices $|A\rangle$, and the inner product is $\langle A|B\rangle = \mathrm{Tr}(\rho\, A^\dagger B)$. Take the basis of matrix units $E_{11}, E_{12}, E_{21}, E_{22}$ and compute Gram entries. Since $E_{ij}^\dagger E_{kl} = \delta_{ik} E_{jl}$ and $\mathrm{Tr}(\rho E_{jl}) = \rho_{lj} = \rho_{jj}\delta_{jl}$,

$$\langle E_{ij}|E_{kl}\rangle = \delta_{ik}\, \rho_{jj}\, \delta_{jl}.$$

Nothing is null here (both $\rho_{jj} = p, 1-p$ are positive), so the space is four-dimensional — *note that already: a two-level system has produced a four-dimensional GNS space*, because we doubled the algebra into vectors. The normalized basis is $|E_{ij}\rangle / \sqrt{\rho_{jj}}$.

*The cyclic vector.* $|\Omega\rangle = |\mathbb{1}\rangle = |E_{11}\rangle + |E_{22}\rangle$, with $\langle\Omega|\Omega\rangle = \mathrm{Tr}(\rho\,\mathbb{1}) = 1$. Check reproduction on, say, $A = E_{11}$ (the projector onto level 1):

$$\langle\Omega|\, \pi(E_{11})\, |\Omega\rangle = \langle\mathbb{1}|E_{11}\rangle = \mathrm{Tr}(\rho E_{11}) = p = \omega(E_{11}),$$

the probability of level 1, as it must be.

*The representation.* $\pi(A)$ acts by left multiplication on the four- dimensional space; in the ordered basis $(E_{11}, E_{21}, E_{12}, E_{22})$ it is literally $A \oplus A$ — two copies of $A$ — because left multiplication by $A$ acts identically on the two columns. So the GNS representation of $M_2(\mathbb{C})$ in a faithful state is *two copies* of the defining representation: $\pi(\mathcal{M}) \cong M_2 \otimes \mathbb{1}_2$ acting on $\mathbb{C}^2 \otimes \mathbb{C}^2$.

The last line of Worked Example 6.4 is not a curiosity; it is the whole point, and it connects straight back to Chapter 2. The GNS space of a qubit in a faithful state is $\mathbb{C}^2 \otimes \mathbb{C}^2$, the algebra acts as $M_2 \otimes \mathbb{1}_2$ on the first factor, and its commutant is $\mathbb{1}_2 \otimes M_2$ on the second. The cyclic vector $|\Omega\rangle$, written in this tensor form, is

$$|\Omega\rangle \;=\; \sqrt{p}\; |1\rangle\, |1\rangle + \sqrt{1-p}\; |2\rangle\, |2\rangle\,, \tag{6.12}$$

which is exactly a *purification* of $\rho$ in the sense of Chapter 2, §2.3: a pure state on a doubled space whose reduced density matrix on the first factor is $\rho = \mathrm{diag}(p, 1-p)$. Tracing out the second factor of (6.12) returns $\rho$, by the Schmidt-decomposition computation we did there. The GNS vector *is* the purification we built by hand in Chapter 2 — now derived, not posited, and derived for any algebra and any state, including ones with no density matrix at all.

**Q 6.3** ***"So GNS is just purification with extra vocabulary? Chapter 2 already doubled a mixed state into a pure one."***

For a finite matrix algebra, yes — and that equivalence is precisely why you should trust the abstraction: it reduces, in the case you can check by hand, to something you already proved.

But two things are genuinely new and both are load-bearing for Part III. First, GNS needs *no ambient Hilbert space*: it manufactures $\mathcal{H}_\omega$ from the algebra and the state alone, so it still works when there is no $\rho$ to purify — in type III (§6.5.3) the density matrix does not exist, yet $\omega$ does, and GNS hands you a Hilbert space and a cyclic vector regardless. Second, the doubling is not arbitrary bookkeeping: the commutant $\mathcal{M}'$ that appears on the second factor is a real algebra of real complementary observables, and the cyclic vector $|\Omega\rangle$ is *cyclic and separating* for $\mathcal{M}$ (acting with $\mathcal{M}$ sweeps out the space, and no nonzero element of $\mathcal{M}$ annihilates it). That pair of properties — which (6.12) exhibits concretely whenever $0 < p < 1$ — is the exact input Chapter 7's Tomita–Takesaki theory demands to manufacture the modular flow. The purification you met in Chapter 2 was the seed of the engine that drives the entire theory; GNS is what lets that engine turn even where density matrices have ceased to exist.

## 6.5 The type classification, by example

We now have every tool. The classification of factors — Murray and von Neumann's, from the 1930s — sorts them by two linked questions: *which projections does the algebra have*, and *does it carry a trace*. We will not axiomatize; we will exhibit one factor of each type and read the answers off it. Keep two finite-dimensional facts about $M_n(\mathbb{C})$ as the reference, because every "type" is a way of departing from them.

A *projection* is a self-adjoint idempotent, $P^\dagger = P = P^2$ — an observable answering a yes/no question. In $M_n(\mathbb{C})$ the projections are exactly the orthogonal projectors onto subspaces, and there is a *trace* Tr that assigns to each projection the dimension of its range, an integer in $\{0, 1, \ldots, n\}$. Two features of this will each fail, one at a time, as we descend the classification:

- *Minimal projections exist*: the rank-one projectors $|\psi\rangle\langle\psi|$ are smallest — no nonzero projection is strictly below them. These are the "atoms," the pure states.
- *The trace takes discrete values*: $\mathrm{Tr}(P) \in \{0, 1, \ldots, n\}$, and normalized to $\tau = \mathrm{Tr}/n$ it takes values in $\{0, \frac{1}{n}, \ldots, 1\}$ — a ladder, not a continuum.

### 6.5.1 Type I: ordinary quantum mechanics

**Type I.** A factor with *minimal (rank-one) projections* and a trace that *counts dimension*. The finite ones are exactly the full matrix algebras $M_n(\mathbb{C})$ (called type $\mathrm{I}_n$); the infinite one is $B(\mathcal{H})$ for a separable infinite-dimensional $\mathcal{H}$ (type $\mathrm{I}_\infty$). *This is the home of standard quantum mechanics.*

There is nothing to build — Worked Example 6.1 already gave the example, and every quantum-mechanics course lives here. A type I factor has pure states (the minimal projections), a density matrix for every state, a von Neumann entropy $-\mathrm{Tr}\,\rho\ln\rho$, and a trace whose values on projections are dimensions. Everything from Chapter 2 is the type I story. The reason the chapter cannot stop here is that a *region of a quantum field is not type I*, and to see why the next two types must exist we follow the one construction that manufactures them: gluing together infinitely many copies of $M_2(\mathbb{C})$.

### 6.5.2 Type II: continuous dimension

Type II is the possibility that a factor has a trace but *no minimal projection*: the "atoms" dissolve, and the trace — which still measures the size of a projection — takes a *continuum* of values. A projection can have trace $1/3$, or $1/\pi$, without there being any smaller projection to build it from. We construct such a factor as a limit of matrix algebras.

**Worked Example 6.5 — The hyperfinite $\mathrm{II}_1$ factor as a limit of matrices**

Consider a chain of qubits and the algebras of the first $N$ of them,

$$\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots, \qquad \mathcal{A}_N = \underbrace{M_2(\mathbb{C}) \otimes \cdots \otimes M_2(\mathbb{C})}_{N \text{ factors}} = M_{2^N}(\mathbb{C}),$$

each sitting inside the next by $A \mapsto A \otimes \mathbb{1}_2$ (an observable of the first $N$ qubits is, in particular, an observable of the first $N+1$, doing nothing to the new one). On each $\mathcal{A}_N$ use the *normalized* trace

$$\tau_N(A) = \frac{1}{2^N}\,\mathrm{Tr}(A), \qquad \tau_N(\mathbb{1}) = 1, \tag{6.13}$$

which counts the *fraction* of the dimension a projection occupies. The key compatibility: the embedding $A \mapsto A \otimes \mathbb{1}_2$ doubles both the trace and the dimension, so $\tau_{N+1}(A \otimes \mathbb{1}_2) = \tau_N(A)$ — *the normalized trace is consistent along the chain*, giving one trace $\tau$ on the union $\mathcal{A}_\infty = \bigcup_N \mathcal{A}_N$. Completing $\mathcal{A}_\infty$ (in the Hilbert space GNS builds from $\tau$) yields a von Neumann algebra $\mathcal{R}$: the *hyperfinite type $II_1$ factor.*

Now watch the two type I features break. Consider the projection $P_1 = \mathrm{diag}(1,0)$ on the first qubit, with $\tau(P_1) = 1/2$. Split it using the *second* qubit: $P_1 \otimes \mathrm{diag}(1,0)$ is a smaller projection below $P_1$, with $\tau = 1/4$. Split again with the third qubit: $\tau = 1/8$. Nothing stops this — there is always another qubit — so *there is no minimal projection*: every projection has one strictly below it. And by taking projections onto cleverly chosen subspaces one can realize *every* value $\tau(P) \in [0,1]$, not just the dyadic fractions: the trace of a projection ranges over a *continuum*. A projection of trace exactly $1/3$ exists in $\mathcal{R}$, even though $1/3$ is not a fraction of the form $k/2^N$ — it is approached by projections of trace $\lfloor 2^N/3 \rfloor / 2^N$ as $N \to \infty$, and $\mathcal{R}$ contains the limit.

**Type $\mathrm{II}_1$.** A factor with a trace $\tau$ ($\tau(\mathbb{1}) = 1$) but *no minimal projection*: the trace of a projection sweeps the whole interval $[0,1]$. "Dimension" has become continuous. The hyperfinite $\mathcal{R}$ of Worked Example 6.5 is *the* canonical example — the completion of the infinite chain of matrix algebras under the normalized trace.

Two remarks make type II less exotic than it first sounds. First, it is a *factor*: one checks that the union of the $\mathcal{A}_N$ has trivial center, because any central element would have to commute with every qubit observable and Lemma 6.3 then forces it to be scalar. No classical labels survive the limit. Second, the trace $\tau$ still does everything a trace should — it is linear, positive, and cyclic ($\tau(AB) = \tau(BA)$) — so type II *does* support density matrices ($\hat\rho$ with $\omega(A) = \tau(\hat\rho A)$) and a finite entropy $-\tau(\hat\rho \ln \hat\rho)$. Type II is the mild surprise: continuous dimension, but life goes on. It is exactly the type that Chapter 7's crossed product will produce as the *repair* of type III, restoring a trace and hence a rate. Which is our cue to meet the patient.

### 6.5.3 Type III: no trace, no density matrix, no entropy

Type III is the possibility that a factor has *no trace at all* — not a continuous one, not any. With no trace there is no density matrix (nothing to put inside $\tau(\hat{\rho}\,\cdot)$), no von Neumann entropy, no purity; the only survivor from Chapter 2's toolkit is *relative* entropy, which we flagged there as the one member of the family that lives without a trace. This is the type a region of spacetime carries, so we must see concretely why a trace can be *impossible*, not merely absent. The construction that shows it is Powers' (1967): the same infinite chain of qubits as Worked Example 6.5, but built with a *biased* state instead of the trace.

**Worked Example 6.6 — The Powers factor: a biased infinite tensor product**

Take again the chain $\mathcal{A}_N = M_2(\mathbb{C})^{\otimes N}$, but on each qubit fix the same non-maximally-mixed state

$$\rho_\lambda = \begin{pmatrix} \lambda & 0 \\ 0 & 1-\lambda \end{pmatrix}, \qquad 0 < \lambda < \tfrac{1}{2},$$

and give the chain the product state $\omega_\lambda = \bigotimes_k \rho_\lambda$ (each qubit independently in $\rho_\lambda$). The single number that will decide everything is the *eigenvalue ratio*

$$\frac{\lambda}{1-\lambda} \in (0,1), \tag{6.14}$$

the odds of finding a qubit in its lower- versus upper-weighted level. Run GNS on $\omega_\lambda$ over the whole chain; the resulting von Neumann algebra $\mathcal{R}_\lambda$ is the *Powers factor*, and for $\lambda \neq \frac{1}{2}$ it is type $\mathrm{III}_\lambda$ — *no trace.* (We label the factor by the eigenvalue $\lambda$; the standard Connes invariant is the eigenvalue *ratio* $\lambda/(1-\lambda)$, a relabelling of the same one-parameter family.)

*Why the trace cannot exist.* Suppose, for contradiction, a trace $\tau$ did exist on $\mathcal{R}_\lambda$ — a linear functional with the cyclic property $\tau(AB) = \tau(BA)$. A trace is by definition blind to unitary rotations: $\tau(UAU^\dagger) = \tau(A)$ for any unitary $U$, since $\tau(UAU^\dagger) = \tau(AU^\dagger U) = \tau(A)$ by cyclicity. Now here is the obstruction. On qubit $k$ consider the flip $U = \sigma_x^{(k)}$, which swaps the two levels. It rotates $\rho_\lambda$ into $\sigma_x \rho_\lambda \sigma_x = \mathrm{diag}(1-\lambda, \lambda)$ — the *reversed* weighting. A trace, being flip-invariant, must assign the projector $P_+^{(k)} = \mathrm{diag}(1,0)$ and its flip $P_-^{(k)} = \mathrm{diag}(0,1)$ the *same* value, $\tau(P_+^{(k)}) = \tau(P_-^{(k)})$. But the *state* $\omega_\lambda$ weighs them unequally, $\omega_\lambda(P_+^{(k)}) = \lambda \neq 1-\lambda = \omega_\lambda(P_-^{(k)})$. In finite dimensions this mismatch is harmless — the trace and the state are simply different functionals, related by a density matrix $\rho_\lambda$ that soaks up the bias, $\omega_\lambda = \tau(\rho_\lambda\,\cdot)$. The infinite chain is what makes it fatal: the modular flow of $\omega_\lambda$ (Chapter 7) rescales each qubit's two levels by the ratio $\lambda/(1-\lambda)$, and iterated over infinitely many qubits this rescaling has *no fixed normalization* a trace could sit at. Concretely, the density matrix "$\rho_\lambda^{\otimes\infty}$" that would relate a trace to $\omega_\lambda$ has eigenvalues $\lambda^n(1-\lambda)^{N-n}$ that, as $N \to \infty$, cluster at 0 with no largest — there is no bounded operator to be it, and correspondingly no trace for it to be measured against. The bias $\lambda/(1-\lambda) \neq 1$, tensored infinitely, *destroys the trace outright.*

**Q 6.4 *"Where exactly did infinity do the damage? Each finite chain $\mathcal{A}_N$ is just $M_{2^N}(\mathbb{C})$, which is type I and certainly has a trace."***

Precisely right, and pinpointing the failure is worth doing because it is the same mechanism that will make a spacetime region type III. Every finite chain $\mathcal{A}_N$ is type I with the ordinary trace, and the biased state $\omega_\lambda$ on it is a perfectly good density matrix $\rho_\lambda^{\otimes N}$. The trace and the state disagree about the flip, but a *finite* density matrix reconciles them. The catastrophe is in how the finite pictures are *glued*. In Worked Example 6.5 the gluing used the *normalized* trace, which was consistent along the chain ($\tau_{N+1} = \tau_N$ on the smaller algebra) precisely because the maximally mixed state $\mathrm{diag}(\frac{1}{2}, \frac{1}{2})$ is flip-symmetric — the trace survived to the limit. Here the biased state is *not* flip-symmetric, so the would-be normalized trace is *inconsistent* between $\mathcal{A}_N$ and $\mathcal{A}_{N+1}$: the new qubit contributes a factor $\lambda$ or $1-\lambda$ that will not cancel. There is no single trace the whole chain agrees on. The ratio (6.14) is the exact measure of the inconsistency: at $\lambda = \frac{1}{2}$ (ratio 1) it vanishes and you recover the type $\mathrm{II}_1$ factor $\mathcal{R}$; away from $\frac{1}{2}$ it is nonzero and the limit is type III. The subscript in "$\mathrm{III}_\lambda$" records the ratio. A region of a quantum field is type III for the identical reason: the vacuum, restricted to the region, is a thermal (biased, flip-asymmetric) state at every scale down to zero distance, and infinitely many scales of bias, glued together, admit no trace.

**Type III.** A factor with *no trace whatsoever*: no $\tau$ with $\tau(AB) = \tau(BA)$ and $\tau(\mathbb{1})$ finite exists. Consequently there is *no density matrix* representing any state ($\omega \neq \tau(\hat{\rho}\,\cdot)$ for any $\hat{\rho}$), *no von Neumann entropy* $-\mathrm{Tr}\,\rho \ln \rho$, and *no purity*. What survives is the state $\omega$ itself (a linear functional, Definition 6.6) and the *relative entropy* $S(\omega\|\varphi)$ comparing two states — which Araki defined without any density matrix, and which was flagged in Chapter 2, §2.5.3, as the one entropy that survives the passage to infinitely many degrees of freedom. The Powers factor $\mathcal{R}_\lambda$ $(0 < \lambda < \frac{1}{2})$ is the concrete instance.

Notice what has and has not been lost. The algebra is still there; states are still there; the comparison of states — relative entropy, distinguishability — is still there and still finite. What is gone is the *absolute* notion of "how much entropy this one state has," because that quantity was always built from the trace, and the trace is gone. This is not a defect of our tools; it is a true property of the physical system, and Chapter 2 prepared us for it by insisting that only *differences* and *comparisons* of entropy are physical. Type III is where that insistence becomes not good taste but necessity.

**For the curious — the three types in one table, and where each lives**

It is worth having the whole classification in view at once. Every von Neumann factor is exactly one of:

| Type | minimal projection? | trace? | where it lives |
|---|---|---|---|
| $\mathrm{I}_n,\ \mathrm{I}_\infty$ | yes (rank one) | yes, integer-valued | ordinary QM: $M_n(\mathbb{C}),\ B(\mathcal{H})$ |
| $\mathrm{II}_1,\ \mathrm{II}_\infty$ | no | yes, continuous $[0,1]$ or $[0,\infty)$ | crossed products; $\mathcal{R}$ |
| $\mathrm{III}_\lambda$ | no | none | regions in QFT; horizons |

The march down the list is a march of *loss*: type I loses nothing; type II gives up the atom (minimal projection) but keeps the trace; type III gives up the trace itself. Murray and von Neumann found types I and II in the 1930s; the existence of a continuum of *distinct* type III factors (indexed by $\lambda$) was Powers' 1967 result, and Connes' classification of them (1973) won the Fields Medal. For this book the entire weight rests on the last row — and the good news is that the crossed product of Chapter 7 moves a system *up* the table, from III back to II, restoring exactly the trace whose absence forbids a decoherence rate.

## 6.6 The punchline: a region of spacetime is type III

We can now state, precisely and honestly, the structural fact this whole chapter was built to reach — and connect it to the divergence that started us off in Chapter 5.

**Principle 6.7** (Local algebras are type $\mathrm{III}_1$)**.** *In relativistic quantum field theory, the algebra of observables localized in a region with a nontrivial causal structure — a Rindler wedge, or a causal diamond of the kind an experiment occupies — is a* hyperfinite type $\mathrm{III}_1$ factor*. It has no trace, no density matrix for the region, and no finite von Neumann entanglement entropy.*

[PROVEN] — this is not a claim of the present theory but an imported rigorous result of algebraic quantum field theory. That the wedge algebra is type $\mathrm{III}_1$ follows from the Bisognano–Wichmann theorem (which identifies the modular flow of the wedge with a Lorentz boost — the fact Chapter 7 builds on) together with the structural analyses of Buchholz, Fredenhagen, and others; the extension to diamonds and general regions is due to the same school. The reader need not take it on faith that the constructions exist — they are theorems — but their proofs use continuum field theory machinery beyond a finite-matrix chapter. What this chapter has supplied is the *meaning*: we built, in Worked Example 6.6, a type III factor from scratch and saw exactly why a biased state repeated at infinitely many scales admits no trace. Principle 6.7 says a region of spacetime is such a factor, because the vacuum restricted to the region is precisely a thermal (biased) state at every length scale down to zero — the same infinitely-repeated bias, now supplied by the physics of the vacuum rather than by our hand.

Here is the reward: *this explains the divergence of Chapter 5 completely.* We asked there for the von Neumann entanglement entropy of a region and got infinity, and we were tempted to blame a bad cutoff. The real reason is now visible and structural: the region's algebra is type III, and *a type III factor has no von Neumann entropy at all*, because it has no trace to build one from. The divergence was the formalism's way of reporting that we had asked a type-I question — "what is $-\mathrm{Tr}\,\rho\ln\rho$?" — of a type-III algebra, where neither $\rho$ nor Tr exists. It is the same category error as asking for the largest element of an open interval: the answer is not a large number but a diagnosis that the question was mis-posed.

**Q 6.5** ***"If a region has no entropy and no density matrix, how can a black hole have a perfectly definite entropy $A/4\ell_{\mathrm{P}}^2$, and how can this book compute a decoherence rate, which surely needs a $\rho$ and a $d/dt$?"***

This is the exact question that organizes Part III, and the honest answer has two halves. *First half*, what survives: even in type III, differences and comparisons survive. The relative entropy $S(\omega\|\varphi)$ is finite and rigorous (Araki), and the black-hole entropy that is physical is a *difference* — the generalized entropy of Chapter 5, which becomes Axiom I — not an absolute $-\mathrm{Tr}\,\rho\ln\rho$. So the finite, meaningful quantities of gravity are exactly the type-III-safe ones, and that is not a coincidence but the reason the theory is written in relative entropy from the start. *Second half*, the repair: a *rate* does need a trace and a density matrix, and type III has neither, so to get a rate the algebra must be *enlarged*. That enlargement is the crossed product of Chapter 7: adjoining to the region's algebra a single new observable — an observer's clock, conjugate to the modular flow — produces a larger algebra that is type II, which *does* have a trace, a density matrix, and hence an entropy and a rate. The type III obstruction is real, and the crossed product is its resolution; the price is that an observer with a clock must be part of the description. The residue of making that construction rigorous for the finite diamond sourced by a mass is the single open problem R1 of Chapter 13. So: no, a bare region has no entropy or rate; yes, a region-plus-observer does; and the whole machinery of getting from the first to the second is what the next chapter builds. The 1.6 nanoseconds on the cover lives in the type II algebra the crossed product manufactures, and this chapter has just identified the type III algebra it must be manufactured *from*.

That is the state of the account at the end of the chapter, and it is worth stating the honest scope plainly, in line, as the style of this book demands. Principle 6.7 is [PROVEN] — the type of a local algebra is settled mathematics. That the crossed product restores a type II algebra with a trace is also [PROVEN] (Takesaki's theorem, unconditional; Chapter 7). What remains [CONJECTURED] is not the existence of the repair but its *execution for the specific finite diamond sourced by a superposed mass*, where the modular flow is not an exact symmetry but a conformal-Killing one — the construction R1, isolated and scoped in Chapter 13. We have reached, precisely, the obstruction; the next chapter builds the tool that removes it, and Part III spends its length turning that tool on the physics.

## Chapter FAQ

**Q: Is a von Neumann algebra the same as a C*-algebra? I have seen both names.** They are close cousins, distinguished by which closure you demand. A C*-algebra is closed in the norm topology; a von Neumann algebra is closed in the weaker "weak operator" topology, equivalently (by the double-commutant theorem) it satisfies $\mathcal{M}'' = \mathcal{M}$. Von Neumann algebras are the larger, weaker-closed objects, and they are the right ones here because the weak closure is what guarantees enough projections to run the type classification — a C*-algebra need not have any nontrivial projections at all. In finite dimensions the two notions coincide, which is why nothing in our matrix examples felt the difference. The one-line rule: *von Neumann = generated by its projections and equal to its own double commutant.*

**Q: Why call it a "factor"? The word suggests multiplication.** The name is Murray and von Neumann's and it does mean what it sounds like. A general von Neumann algebra decomposes as a "direct integral" — a continuous version of the direct sum (6.5) — over its center, and the pieces of that decomposition, the ones with trivial center, are the irreducible building blocks. The whole algebra *factors* into them, as an integer factors into primes, and each block is a *factor*. Classifying factors (this chapter) is therefore the atom-level task; general algebras are assembled from factors over their classical center.

**Q: The GNS space in Worked Example 6.4 was four-dimensional for a two-level system. Isn't that wasteful — doubling the dimension?** It is doubling, and it is not waste; it is the point. The GNS space carries both the algebra $\mathcal{M}$ (on the first factor) and its commutant $\mathcal{M}'$ (on the second) as genuine operators, and the cyclic vector is a purification that makes the state pure on the doubled space. That doubled structure — the thermofield double, in physics language — is exactly what Chapter 7 needs: the modular conjugation $J$ maps $\mathcal{M}$ onto $\mathcal{M}'$, i.e. the first factor onto the second, and could not exist without the doubling. When the state is pure to begin with (a rank-one $\rho$), one of the doubled factors is trivial and no dimension is added — the doubling is proportional to how mixed the state is, i.e. to how much there is to purify.

**Q: If type III has no entropy, why does everyone talk about "entanglement entropy of a region" as if it were finite?** Loosely, and with a cutoff quietly installed. The phrase is shorthand for a regulated quantity — put a short-distance cutoff $\epsilon$, compute $-\mathrm{Tr}\,\rho_\epsilon \ln \rho_\epsilon$, and find it scales as (area)$/\epsilon^{d-2}$. That leading term is real and universal (it is where the "area law" comes from), but its coefficient depends on $\epsilon$ and so is not by itself physical, exactly as Chapter 5 warned. The type-III statement is the cutoff-free truth underneath: the absolute entropy does not exist, only differences (relative entropies, and entropy *changes* under a physical process) are finite and scheme-independent. When a calculation gives a finite "entanglement entropy," look for the difference it is secretly computing.

## Exercises

**Exercise 6.1.** Let $\mathcal{M} \subset M_3(\mathbb{C})$ be the algebra of all *diagonal* $3 \times 3$ matrices. (a) Show it is a $*$-algebra containing $\mathbb{1}$. (b) Compute its commutant $\mathcal{M}'$. (c) Compute the center $\mathcal{Z}(\mathcal{M})$ and decide whether $\mathcal{M}$ is a factor. (d) How many superselection sectors does $\mathcal{M}$ describe, and what is the physical picture of an observer holding exactly this algebra?

**Exercise 6.2.** Repeat the GNS construction of Worked Example 6.4 for the *maximally mixed* qubit state $\rho = \frac{1}{2}\mathbb{1}$ ($\omega(A) = \frac{1}{2}\mathrm{Tr}\,A$). (a) Write the inner product on the four matrix-unit vectors and confirm the space is four-dimensional. (b) Write the cyclic vector $|\Omega\rangle$ in the tensor form $\mathbb{C}^2 \otimes \mathbb{C}^2$ and identify which entangled state it is. (c) Verify $|\Omega\rangle$ is separating (no nonzero $A \in M_2$ has $\pi(A)\,|\Omega\rangle = 0$) and explain in one sentence why a *pure* state $\rho = |\psi\rangle\,\langle\psi|$ would *fail* to give a separating cyclic vector.

**Exercise 6.3.** In the hyperfinite $\mathrm{II}_1$ chain of Worked Example 6.5, exhibit explicitly a projection of normalized trace exactly $3/8$, using observables of the first three qubits only. (Hint: work in $\mathcal{A}_3 = M_8(\mathbb{C})$ and choose a rank-$k$ diagonal projector; which $k$ gives $\tau = 3/8$?) Then argue in one line why *no* projection of trace $3/8$ is minimal.

**Exercise 6.4.** For the Powers factor $\mathcal{R}_\lambda$ of Worked Example 6.6 with $\lambda = 1/3$: (a) State the eigenvalue ratio (6.14). (b) On the first qubit, write the two projectors $P_+, P_-$ and give both their state-values $\omega_\lambda(P_\pm)$ and the values a hypothetical flip-invariant trace would be forced to assign; identify the contradiction. (c) In one sentence, what value of $\lambda$ would make $\mathcal{R}_\lambda$ type $\mathrm{II}_1$ instead of type III, and why?

## Solutions

*Solution (Exercise 6.1).* (a) Diagonal matrices add, scale, multiply (entrywise), and conjugate-transpose (a real-then-conjugate of the diagonal, still diagonal) staying diagonal; $\mathbb{1}$ is diagonal. So $\mathcal{M}$ is a $*$-algebra with identity. (b) A matrix $X$ commutes with *all* diagonal matrices iff it commutes with each $E_{ii}$; the computation of Lemma 6.3 with $i = j$ forces every off-diagonal entry $X_{kl}$ $(k \neq l)$ to vanish, while diagonal $X$ automatically commute with diagonal matrices. Hence $\mathcal{M}' = \{\text{diagonal matrices}\} = \mathcal{M}$: this algebra is its own commutant (it is *maximal abelian*). (c) $\mathcal{Z} = \mathcal{M} \cap \mathcal{M}' = \mathcal{M}$, the full three-dimensional space of diagonal matrices — as large as it can be. Since $\mathcal{Z} \neq \mathbb{C}\mathbb{1}$, $\mathcal{M}$ is emphatically *not* a factor. (d) Three superselection sectors — one per diagonal slot. The observer holds only the three commuting projectors $E_{11}, E_{22}, E_{33}$ and their combinations: she can ask "which of the three levels?" and read a classical answer, but owns no operator connecting the levels, so she can never prepare or detect a superposition of them. This is a purely classical (probabilistic) system dressed as a quantum one — the opposite extreme from the full-matrix factor of Worked Example 6.1.

*Solution (Exercise 6.2).* (a) With $\rho = \frac{1}{2}\mathbb{1}$, $\langle E_{ij}|E_{kl}\rangle = \frac{1}{2}\delta_{ik}\delta_{jl}$: the four matrix units are orthogonal, each of norm-squared $\frac{1}{2}$, so the space is four-dimensional with orthonormal basis $\sqrt{2}\,|E_{ij}\rangle$. (b) $|\Omega\rangle = |\mathbb{1}\rangle = |E_{11}\rangle + |E_{22}\rangle$, and in tensor form (6.12) with $p = \frac{1}{2}$ it is $|\Omega\rangle = \frac{1}{\sqrt{2}}(|1\rangle|1\rangle + |2\rangle|2\rangle)$ — the Bell state $|\Phi^+\rangle$ of Chapter 2, Worked Example 2.2, whose reduced density matrix is exactly $\frac{1}{2}\mathbb{1}$. (c) $\pi(A)|\Omega\rangle = |A\rangle$, which is zero only if $A = 0$ (the map $A \mapsto |A\rangle$ is injective here since no direction is null), so $|\Omega\rangle$ is separating. A pure state $\rho = |\psi\rangle\langle\psi|$ would make the GNS inner product $\langle A|B\rangle = \langle\psi| A^\dagger B |\psi\rangle$ degenerate — any $A$ with $A|\psi\rangle = 0$ is null — so the cyclic vector, while still cyclic, would *not* be separating: purity kills the doubling, and with it the separating property Tomita–Takesaki needs. (Only faithful states, $0 < p < 1$, give a cyclic *and* separating vector — the input of Chapter 7.)

*Solution (Exercise 6.3).* Work in $\mathcal{A}_3 = M_8(\mathbb{C})$, where the normalized trace of a rank-$k$ diagonal projector is $\tau = k/8$. Choosing $k = 3$ gives $\tau = 3/8$ exactly: e.g. $P = \mathrm{diag}(1,1,1,0,0,0,0,0)$, a projection onto three of the eight basis states of the first three qubits. It is not minimal because the *fourth* qubit splits it: $P \otimes \mathrm{diag}(1,0)$ is a strictly smaller projection below it, with $\tau = 3/16$ — and the chain never ends, so no projection of any positive trace can be minimal. (This is the concrete face of "no minimal projection": there is always one more qubit to halve any projection you name.)

*Solution (Exercise 6.4).* (a) With $\lambda = 1/3$, the ratio is $\lambda/(1-\lambda) = (1/3)/(2/3) = 1/2$. So $\mathcal{R}_{1/3}$ is the type III$_\lambda$ factor with $\lambda = 1/3$ (some texts index by the ratio $1/2$; the physics is the same bias). (b) $P_+ = \mathrm{diag}(1,0)$, $P_- = \mathrm{diag}(0,1)$ on the first qubit. The state assigns $\omega_\lambda(P_+) = \lambda = 1/3$ and $\omega_\lambda(P_-) = 1 - \lambda = 2/3$ — unequal. A flip-invariant trace, obeying $\tau(P_+) = \tau(\sigma_x P_+ \sigma_x) = \tau(P_-)$, is *forced* to assign them the same value. In finite dimensions the density matrix $\rho_\lambda$ absorbs the difference; over the infinite chain no such bounded density matrix exists (its eigenvalues cluster at zero with no maximum), so the trace has nowhere to sit and cannot exist — that is the contradiction. (c) $\lambda = 1/2$ makes the ratio 1: the state becomes the flip-symmetric maximally mixed state, the bias vanishes, the normalized trace is consistent along the chain, and the limit is the type II$_1$ factor $\mathcal{R}$ of Worked Example 6.5 — trace restored. Any $\lambda \neq \frac{1}{2}$ leaves a nonzero bias and gives type III.

**What to carry forward**

1. The primary object is the *algebra* of observables an observer can reach, a von Neumann algebra $\mathcal{M}$ (a $*$-algebra with $\mathcal{M}'' = \mathcal{M}$). Its *commutant* $\mathcal{M}'$ is the complementary observables; its *center* $\mathcal{Z} = \mathcal{M} \cap \mathcal{M}'$ holds the classical (superselection) labels, and a *factor* — the arena of a single quantum system — is one with trivial center $\mathbb{C}\mathbb{1}$. All of this is visible on finite block matrices.

2. The *GNS construction* builds a Hilbert space, a representation, and a cyclic vector $|\Omega\rangle$ from an algebra and a state $\omega$ alone, with $\langle\Omega|\,\pi(A)\,|\Omega\rangle = \omega(A)$ — and for a faithful $2 \times 2$ state $|\Omega\rangle$ is exactly the *purification* of Chapter 2, cyclic and separating, the seed of Chapter 7's modular flow. GNS needs no ambient Hilbert space, so it works where no density matrix exists.

3. Factors come in three types, each met concretely: *type I* ($M_n(\mathbb{C})$, $B(\mathcal{H})$ — ordinary QM, minimal projections and an integer trace); *type II* (a trace but no minimal projection — continuous dimension, a projection of trace $1/3$, the hyperfinite $\mathcal{R}$); *type III* (*no trace at all*, hence no density matrix and no von Neumann entropy — only relative entropy survives; the Powers factor $\mathcal{R}_\lambda$ from a biased infinite tensor product, the ratio $\lambda/(1-\lambda) \neq 1$ killing the trace). A region of relativistic quantum field theory carries a hyperfinite *type III*$_1$ factor ([PROVEN]): that is *why* Chapter 5's entropy diverged, and the crossed product of Chapter 7 — adjoining an observer's clock to move III$\to$II — is the repair, with the finite-diamond execution the open problem R1 of Chapter 13.

# 7 Modular Theory, KMS, and the Observer's Clock

**Synopsis.** Give an algebra of observables a state, and — if the state is honest enough to be nowhere blind — the pair secretes a *clock* nobody put there. That is the Tomita–Takesaki theorem, and it is the mathematical climax of our toolkit. We build it from a two-by-two example where you can watch every operator by hand, discover that its modular flow is nothing but time evolution by the state's own Hamiltonian at the state's own temperature, and then state the general theorem it forces. The KMS condition turns "thermal" into a property of correlators alone, and the Bisognano–Wichmann theorem cashes the whole abstraction into a single geometric fact: the vacuum's clock on a Rindler wedge *is* the Lorentz boost, ticking at inverse temperature exactly $2\pi$. For a finite causal diamond the clock is again geometric, a local re-weighting of energy. Finally the crossed product gives the observer that clock as a physical degree of freedom, and in doing so repairs the type III catastrophe of Chapter 6: the enlarged algebra has a trace, hence an entropy, hence the generalized entropy of Axiom I. The modular Hamiltonian $K$ built here is the object the theory's central equation, $K = 2\pi H_{\text{phys}}$, is about.

**You will be able to.** construct the modular operator $\Delta$ and conjugation $J$ from the Tomita operator $S = J\Delta^{1/2}$ on a finite-dimensional example; show $\Delta = \rho \otimes \rho^{-1}$ for a Gibbs state and read off $K = -\ln \Delta$; state and verify the KMS condition; reproduce the Bisognano–Wichmann $\beta_{\text{mod}} = 2\pi$ from Euclidean rotation; write down the Hislop–Longo diamond weight $f(r) = (R^2 - r^2)/2R$; and explain how the crossed product turns a type $\text{III}_1$ algebra into a type II algebra with a trace.

**Prerequisites.** Chapter 6 (von Neumann algebras, factor types $\text{III}_1$, GNS, the cyclic–separating vector), heavily; Chapter 4 (the Unruh $2\pi$, the KMS preview); Chapter 3 (the Rindler wedge and boost); Chapter 2 (relative entropy, purification). Finite-linear algebra: tensor products, the polar decomposition, matrix logarithms.

## 7.1 The question: does a state carry a clock?

Chapter 6 handed us an uncomfortable object. The observables an observer can reach form a von Neumann algebra $M$; the state of the world is carried by a single vector $|\Omega\rangle$ in a Hilbert space; and the pair is *cyclic and separating* — $|\Omega\rangle$ is cyclic ($M\,|\Omega\rangle$ is dense: acting with the algebra reaches everywhere) and separating ($A\,|\Omega\rangle = 0$ forces $A = 0$: no observable is invisible to the state). The vacuum on a Rindler wedge is exactly such a pair, and so, we will find, is every physically sensible restriction of a field theory to a region. What Chapter 6 did *not* tell us is whether such a pair carries any dynamics of its own.

Here is the question, put operationally. An observer with algebra $M$ and state $|\Omega\rangle$ can measure the observables in $M$ and nothing else. Is there a canonical one-parameter family of transformations of $M$ — a flow $A \mapsto \sigma_t(A)$ — that the state itself selects, with no extra input? Not a flow we *choose* by writing down a Hamiltonian, but one the data $(M, |\Omega\rangle)$ already determine. If so, that flow deserves to be called the state's intrinsic notion of time.

The astonishing answer, due to Tomita and Takesaki, is *yes*: every cyclic and separating pair carries a unique such flow, canonically. We will not take this on faith. We will build the machine on a system small enough to hold in one hand — a single qubit doubled — watch the flow it produces turn out to be ordinary thermal time, and only then state the theorem in the generality the rest of the book needs.

**Q 7.1** ***"Why should an algebra plus a state know anything about time? Time is a coordinate on spacetime; these are abstract operators."***

That is precisely the surprise, and it is the seed of the entire theory. In ordinary mechanics you *supply* the dynamics: you write a Hamiltonian $H$ by hand and evolve by $e^{-iHt}$. Modular theory says that once you have committed to a state $|\Omega\rangle$ that is nowhere blind, the freedom to choose is gone — the state has already picked a flow, and (this is the load-bearing part) that flow is a *thermal* evolution whose "temperature" is set by the state. When the state is the vacuum seen by an accelerated observer, the flow will turn out to be the boost, i.e. literally a motion through spacetime (§7.5). The abstract clock and the geometric clock coincide. That coincidence, promoted to an equation, is what Chapter 9 makes Axiom III and what the whole of Part III is trying to prove.

## 7.2 Tomita–Takesaki from a two-by-two example

### 7.2.1 The doubled Hilbert space and the flip

Take the smallest nontrivial case: a single qubit, algebra $M = M_2(\mathbb{C})$, the two-by-two complex matrices. A von Neumann algebra needs a Hilbert space to act on and a cyclic–separating vector living in it, and a 2-dimensional space supplies neither (a 2-vector cannot be cyclic for a 4-dimensional algebra). The fix is the construction of Chapter 6: represent $M_2$ on the *doubled* space $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$, with $M$ acting on the left factor as $A \otimes \mathbb{1}$ and its commutant $M' = \{B'\}$ acting on the right as $\mathbb{1} \otimes B^{\mathsf{T}}$. This is the finite-dimensional face of the GNS construction, and it is nothing more exotic than purification (§2.3, Chapter 2): a mixed state of one qubit is the reduced state of a pure vector on two.

Fix a density matrix $\rho$ on the qubit, strictly positive ($\rho = \sum_k p_k |k\rangle\langle k|$ with every $p_k > 0$ — *this* is separating; a zero eigenvalue would leave a direction the state cannot see), and purify it into

$$|\Omega\rangle \;=\; \sum_k \sqrt{p_k}\; |k\rangle \otimes |k\rangle \;=\; (\rho^{1/2} \otimes \mathbb{1})\,|I\rangle\,, \qquad |I\rangle \equiv \sum_k |k\rangle \otimes |k\rangle\,. \tag{7.1}$$

The second form is a bookkeeping trick worth its weight: $|I\rangle$ is the "unnormalised maximally entangled" vector, the vectorised identity matrix, and it satisfies the one identity we will use again and again,

$$(A \otimes \mathbb{1})\,|I\rangle \;=\; (\mathbb{1} \otimes A^{\mathsf{T}})\,|I\rangle \qquad \text{for every matrix } A. \tag{7.2}$$

(Check it on basis elements: both sides send $|I\rangle$ to $\sum_k (A\,|k\rangle) \otimes |k\rangle = \sum_{k,l} A_{lk}\,|l\rangle \otimes |k\rangle$.) In words: pushing a matrix through the entangled vector from the left equals pushing its transpose through from the right. Every formula below is one application of (7.2) away from obvious.

### 7.2.2 The Tomita operator S

Now the definition that starts everything. Modular theory is built on one antilinear operator, the *Tomita operator* $S$, defined by its action on the dense set $M\,|\Omega\rangle$:

$$S\,(A\,|\Omega\rangle) \;=\; A^\dagger\,|\Omega\rangle\,, \qquad A \in M. \tag{7.3}$$

Read (7.3) slowly, because its meaning is more physical than it looks. The vector $A\,|\Omega\rangle$ is "the state with the observable $A$ applied." $S$ maps it to "the state with $A^\dagger$ applied." It is the operation "take the adjoint," transported from the algebra onto the Hilbert space by the state. It is antilinear ($S(cA\,|\Omega\rangle) = \bar{c}\,A^\dagger\,|\Omega\rangle$, because adjoint conjugates scalars), and it is well-defined precisely because $|\Omega\rangle$ is separating: if $A\,|\Omega\rangle = A'\,|\Omega\rangle$ then $(A - A')\,|\Omega\rangle = 0$ forces $A = A'$, so the value $A^\dagger\,|\Omega\rangle$ does not depend on which $A$ we used to name the input vector.

$S$ is generically not a symmetry — it does not preserve lengths, because $A^\dagger\,|\Omega\rangle$ need not have the same norm as $A\,|\Omega\rangle$. The failure to preserve lengths is measured by a positive operator, and *that operator is the whole point.* Every antilinear operator has a polar decomposition, exactly as every complex number is (phase)×(modulus):

$$S \;=\; J\,\Delta^{1/2}, \qquad \Delta \;\equiv\; S^\dagger S \;>\; 0, \tag{7.4}$$

with $\Delta$ a positive operator (the "modulus," called the *modular operator*) and $J$ an antiunitary (the "phase," called the *modular conjugation*, satisfying $J^2 = \mathbb{1}$ and $J^\dagger = J$). The modular Hamiltonian is the logarithm of the modulus,

$$K \;\equiv\; -\ln\Delta, \qquad \text{equivalently} \qquad \Delta \;=\; e^{-K}. \tag{7.5}$$

The sign is a convention chosen so that $K$ will play the role of a Hamiltonian (large $\Delta$ = low modular energy). We now compute all three objects explicitly for the qubit and watch them acquire physical meaning.

**Worked Example 7.1 — Tomita–Takesaki for one Gibbs qubit**

Take the state (7.1) built from a thermal density matrix $\rho = e^{-\beta h}/Z$ of a single-qubit Hamiltonian $h$ (so $Z = \operatorname{Tr} e^{-\beta h}$). We find $S$, $\Delta$, $J$, and $K$ from the definitions, using nothing but (7.2).

*Step 1 — act with $S$.* An observable is $A = A \otimes \mathbb{1}$; its vector is $A\,|\Omega\rangle = (A\rho^{1/2} \otimes \mathbb{1})\,|I\rangle$. By definition $S(A\,|\Omega\rangle) = A^\dagger\,|\Omega\rangle = (A^\dagger\rho^{1/2} \otimes \mathbb{1})\,|I\rangle$. Push both through $|I\rangle$ with (7.2) to compare them on the same footing; the cleanest route is to compute $\Delta = S^\dagger S$ directly.

*Step 2 — the modular operator.* A short computation (do it in the $|k\rangle \otimes |l\rangle$ basis, or quote the standard result) gives the modular operator as a left-right pair,

$$\boxed{\Delta \;=\; \rho \otimes \rho^{-1}} \qquad \Longrightarrow \qquad \Delta^{it}\,(A \otimes \mathbb{1})\,\Delta^{-it} \;=\; (\rho^{it}\,A\,\rho^{-it}) \otimes \mathbb{1}. \tag{7.6}$$

The right-hand identity is the sanity check: $\Delta^{it}$ conjugates a left-algebra element back into the left algebra, dressing it with $\rho^{it}(\cdot)\rho^{-it}$ and leaving the commutant factor untouched.

*Step 3 — the modular Hamiltonian.* Take minus the logarithm of (7.6), using $\ln(\rho \otimes \rho^{-1}) = \ln \rho \otimes \mathbb{1} - \mathbb{1} \otimes \ln \rho$ and $\ln \rho = -\beta h - \ln Z$:

$$K = -\ln \Delta = \beta(h \otimes \mathbb{1} - \mathbb{1} \otimes h). \tag{7.7}$$

The $\ln Z$ pieces cancel between the two factors — the modular Hamiltonian never sees the partition function.
*Step 4 — the conjugation.* From $S = J\Delta^{1/2}$ and (7.6) one finds $J$ acts as complex conjugation composed with the flip of the two factors; concretely $J(|k\rangle \otimes |l\rangle) = |l\rangle \otimes |k\rangle$ up to conjugation. Its job is structural, not dynamical, and §7.3 says what it does.

### 7.2.3 Reading the answer: modular flow is thermal time

Equation (7.7) is the seed of this entire book, so we dwell on it. The modular flow it generates is

$$\sigma_t(A) = \Delta^{it} A \Delta^{-it} = e^{iKt} A e^{-iKt} = e^{i\beta h t} A e^{-i\beta h t} \qquad (A \in M). \tag{7.8}$$

This is *Heisenberg time evolution by the physical Hamiltonian* $h$, run for a rescaled time $\beta t$. Strip away the doubling and read the physics:

> For a thermal state $\rho = e^{-\beta h}/Z$, the Tomita–Takesaki modular flow is ordinary time evolution generated by the system's own Hamiltonian $h$, with the modular time $t$ related to physical time $t_{\text{phys}}$ by $t_{\text{phys}} = \beta t$. The modular Hamiltonian is $K = \beta h$ (on the left factor). *Modular time is physical time, divided by temperature.*

This is the first appearance of the theory's spine. The state has handed us a clock, and the clock ticks in step with the real Hamiltonian: one modular tick $t$ corresponds to a physical interval $\beta t$, so a colder (larger-$\beta$) state stretches each modular tick over more physical time, while a hotter state packs the same modular interval into less of it. The proportionality constant between the two clocks is the inverse temperature. Hold this: in §7.5 the state will be the vacuum, "$h$" will be the boost generator, and the proportionality constant will turn out to be *exactly* $2\pi$, with no freedom to be anything else — the number the whole theory hangs on.

***Q 7.2 "You assumed the state was already thermal, $\rho = e^{-\beta h}/Z$. Isn't it circular to then discover a thermal flow?"***

It would be circular if we had *needed* the thermal form, but we did not. The construction (7.3)–(7.5) takes an *arbitrary* strictly-positive $\rho$ and returns $\Delta = \rho \otimes \rho^{-1}$, $K = -\ln \rho \otimes \mathbb{1} + \mathbb{1} \otimes \ln \rho$, whatever $\rho$ is. Any density matrix can be *written* as $e^{-\beta h}$ for the choice $h = -\frac{1}{\beta} \ln \rho$ (at any fixed reference $\beta$), so the content is not "thermal states flow thermally" but the reverse and stronger statement: *every* faithful state defines an $h$ — its modular Hamiltonian — with respect to which *it is* the Gibbs state, and its intrinsic flow is evolution by that $h$. The state does not have to be handed a temperature; it supplies its own. That is why the theorem is remarkable rather than tautological, and why it survives into the type III world of §7.3, where neither $\rho$ nor $h$ exists on its own but $\Delta$ and $K$ still do.

## 7.3 The theorem in general

The finite example did the pedagogy; now we state what Tomita and Takesaki proved for any von Neumann algebra, including the type III factors of Chapter 6 where no $\rho$ and no trace exist. Everything survives because the objects $S$, $\Delta$, $J$, $K$ were defined (7.3) without ever mentioning a density matrix — only the algebra, the vector, and the adjoint.

**Theorem 7.1** (Tomita–Takesaki). *Let $M$ be a von Neumann algebra with a cyclic and separating vector $|\Omega\rangle$. Let $S = J\Delta^{1/2}$ be the polar decomposition of the closure of the Tomita operator $S(A\,|\Omega\rangle) = A^\dagger\,|\Omega\rangle$. Then:*

1. (Modular flow preserves the algebra.) *$\Delta^{it}\,M\,\Delta^{-it} = M$ for every real $t$: the flow $\sigma_t(A) = \Delta^{it}A\Delta^{-it}$ is a one-parameter group of automorphisms of $M$.*
2. (Conjugation swaps algebra and commutant.) *$J\,M\,J = M'$, where $M'$ is the commutant. $J$ fixes $|\Omega\rangle$ ($J\,|\Omega\rangle = |\Omega\rangle$) and $\Delta\,|\Omega\rangle = |\Omega\rangle$.*

*The modular Hamiltonian $K = -\ln\Delta$ generates the flow, $\sigma_t(A) = e^{iKt}A\,e^{-iKt}$.* [PROVEN]

Two clauses, two meanings. Clause 1 is the miracle: a flow that *stays inside the algebra*. There is no reason a generic conjugation should map observables the observer can reach back to observables she can reach — but the modular flow does, exactly, for all time. This is what licenses calling it a dynamics *of the system*, not a leak into the environment. Clause 2 is the bookkeeping that makes it consistent: $J$ is the mirror exchanging the observer's algebra $M$ with the degrees of freedom she cannot touch, $M'$. In the qubit example $M'$ was the purifying second qubit, and $J$ was the flip; in the Rindler case $M'$ will be the opposite wedge, and $J$ the CPT-like reflection across the horizon.

**Principle 7.2** (The modular clock). *Every faithful state $|\Omega\rangle$ on every von Neumann algebra $M$ carries a canonical, state-determined flow $\sigma_t$ — its* thermal time*. There is no external clock and no chosen Hamiltonian; the state* is *the clock. In the type III algebras of quantum field theory, where $\rho$, $H$, and the trace all fail to exist, this modular flow is the only surviving notion of time internal to a region, and $K = -\ln\Delta$ the only surviving Hamiltonian.*

**Q 7.3 "If $K = -\ln\Delta$ exists even in type III where the Hamiltonian $H$ does not, are they the same thing or not?"**

They are not equal, and keeping them apart is the honest core of the whole program. In the finite case they were proportional, $K = \beta H$ (up to the doubling), with $\beta$ a number. In type III there is no $H$ at all: the region has no total energy operator, only the modular $K$, which always exists. The theory's central claim (Chapter 12, Chapter 13) is that for a gravitating region the modular Hamiltonian of the physical vacuum is again proportional to a physical generator, with the proportionality constant pinned to $2\pi$: $K = 2\pi H_{\text{phys}}$. In flat space this is a *theorem* — Bisognano–Wichmann, §7.5. For a dynamical gravitational diamond it is the theory's one [CONJECTURED] equation, proven in expectation values and every solvable model (including the ones in this chapter), open as a full operator equation. That $K$ exists where $H$ does not is exactly why the correspondence between them is nontrivial and worth a whole book.

## 7.4 The KMS condition: thermality without a bath

The modular flow was built to be a state's intrinsic time. We now name the property that makes it a *thermal* time — the Kubo–Martin–Schwinger (KMS) condition — and we do it in a way that survives into type III, because it never mentions $\rho$. The operational question is: given only correlators of observables, how would you *detect* that a state is thermal? The KMS answer: by an analyticity of two-point functions in imaginary time.

### 7.4.1 KMS from the Gibbs state

Start where we can compute. For a Gibbs state $\rho = e^{-\beta H}/Z$ with flow $\sigma_t(B) = e^{iHt}Be^{-iHt}$, consider two observables $A, B$ and the correlator $\langle A\,\sigma_t(B)\rangle = \mathrm{Tr}(\rho\, A\, e^{iHt}Be^{-iHt})$. The trace's cyclicity, together with the fact that $\rho = e^{-\beta H}$ is *itself* an imaginary-time evolution, ties the two orderings together. Slide $\rho$ through:

$$\begin{aligned}\langle \sigma_t(B)\, A\rangle &= \mathrm{Tr}\big(e^{-\beta H}e^{iHt}Be^{-iHt}A\big)/Z \\ &= \mathrm{Tr}\big(A\, e^{-\beta H}e^{iHt}Be^{-iHt}\big)/Z \qquad \text{(cyclicity)} \\ &= \mathrm{Tr}\big(e^{-\beta H}A\, e^{iH(t+i\beta)}Be^{-iH(t+i\beta)}\big)/Z \qquad (\text{insert } e^{\beta H}e^{-\beta H}) \\ &= \big\langle A\, \sigma_{t+i\beta}(B)\big\rangle. \end{aligned} \tag{7.9}$$

The third line moved one factor of $e^{-\beta H}$ from the far left to dress the flow, turning real time $t$ into complex time $t + i\beta$. The result is the KMS condition.

**Definition 7.3** (KMS condition)**.** A state $\omega$ is *KMS at inverse temperature* $\beta$ with respect to a flow $\sigma_t$ if for all observables $A, B$ the correlator $F_{AB}(t) = \omega(A\,\sigma_t(B))$ extends to an analytic function on the strip $0 < \mathrm{Im}\, t < \beta$, continuous on its closure, satisfying the boundary (periodicity) relation

$$\omega\big(A\,\sigma_{t+i\beta}(B)\big) \;=\; \omega\big(\sigma_t(B)\, A\big). \tag{7.10}$$

The physical reading of (7.10) is that thermal correlators are *periodic in imaginary time with period* $\beta$: an equilibrium state at temperature $T = 1/\beta$ is, in the Euclidean picture, a field theory on a circle of circumference $\beta$. Nothing in Definition 7.3 refers to a heat bath, a density matrix, or even a trace — it is a statement about correlators alone, and so it makes sense verbatim in the type III algebras where $\rho$ does not exist. That is its power.

$$\text{modular flow of } |\Omega\rangle \iff |\Omega\rangle \text{ is KMS at } \beta_{\mathrm{mod}} = 1 \text{ w.r.t. } \sigma_t. \tag{7.11}$$

The tie-back to Tomita–Takesaki is exact and it is the reason modular theory and thermodynamics are the same subject: the modular flow $\sigma_t$ is the *unique* flow with respect to which $|\Omega\rangle$ is KMS, and in modular time the period is $\beta_{\mathrm{mod}} = 1$ (because $K = -\ln \Delta$ absorbed the temperature into the definition of $t$). Restoring physical time reinstates the physical $\beta$, exactly as (7.8) traded $t$ for $\beta t$. To say a state is thermal, then, is not to say it came from a reservoir; it is to say its correlators are analytic and periodic in imaginary modular time. Thermality is intrinsic.

**Q 7.4 *"Definition 7.3 looks like a piece of complex-analysis bookkeeping. Why is periodicity in imaginary time the* definition *of hot?"***

Because temperature *is* imaginary-time periodicity — that is not a slogan but the content of the Euclidean path integral you met in Chapter 4. The thermal partition function $Z = \mathrm{Tr}\, e^{-\beta H}$ is the ordinary evolution operator $e^{-iHt}$ continued to imaginary time $t = -i\beta$ and traced, and the trace glues the imaginary-time interval into a circle of circumference $\beta$. Correlators computed on that circle are automatically periodic with period $\beta$ — which is exactly (7.10). So the KMS condition is not a proxy for thermality dressed up in analysis; it is thermality, written in the only language (correlators) that still works when the Hilbert-space objects $\rho$ and $H$ have dissolved. This is why, when Chapter 4 found the Unruh temperature by demanding that the Minkowski vacuum look periodic in the Rindler observer's imaginary time, it was really discovering that the vacuum is a KMS state — the subject of the next section.

## 7.5 Bisognano–Wichmann: the load-bearing $2\pi$

We now cash the entire abstraction into one geometric fact, and it is the flat- space, exact instance of the theory's central identity. Recall the Rindler wedge of Chapter 3: the region $x > |t|$ of Minkowski space accessible to an observer accelerating uniformly along $x$. Its boundary is a horizon; the observer's proper time runs along the orbits of the Lorentz boost in the $t$–$x$ plane,

$$B \;=\; \tfrac{1}{c}(x\,\partial_t + c^2 t\,\partial_x), \qquad \text{orbits: hyperbolae } x^2 - c^2t^2 = \text{const.} \tag{7.12}$$

The algebra $M = \mathcal{A}(\text{wedge})$ is the field observables in the wedge; the state is the Minkowski vacuum $|0\rangle$, which Chapter 6 identified as cyclic and separating for the wedge (a type $\mathrm{III}_1$ pair). Tomita and Takesaki guarantee this pair has a modular flow. Bisognano and Wichmann computed what it is.

**Theorem 7.4** (Bisognano–Wichmann)**.** *For the algebra of a Rindler wedge in the Minkowski vacuum $|0\rangle$, the modular flow is the Lorentz boost (7.12), and the modular Hamiltonian is*

$$\boxed{K \;=\; 2\pi\, K_{\text{boost}}} \qquad \textit{(units } c = 1\textit{)}, \tag{7.13}$$

*where $K_{\text{boost}}$ is the generator of (7.12) (the boost "energy"). Equivalently, the vacuum is KMS with respect to the boost flow at inverse temperature exactly $\beta_{\text{mod}} = 2\pi$. The modular conjugation $J$ is the CPT reflection sending the wedge to its opposite.* [PROVEN] *(a rigorous theorem of algebraic quantum field theory).*

Equation (7.13) is the ancestor of $K = 2\pi H_{\text{phys}}$. Here the "physical Hamiltonian" is the boost generator — the honest generator of the accelerated observer's proper time — and the modular Hamiltonian is exactly $2\pi$ times it, with *no free constant*. Let us see where the $2\pi$ comes from, because its inevitability is the whole reason the theory has a parameter-free number at its heart.

### 7.5.1 Why exactly $2\pi$: Euclidean rotation

The cleanest argument is the one Chapter 4 used for Unruh, now read through modular eyes. Continue Minkowski time to Euclidean time, $t \to -i\tau_E$. The boost (7.12), a hyperbolic "rotation" mixing $t$ and $x$, becomes an *ordinary rotation* in the Euclidean $(\tau_E, x)$ plane:

$$\text{boost by rapidity } \eta \xrightarrow{t \to -i\tau_E} \text{rotation by angle } \theta = \eta \text{ in the } (\tau_E, x) \text{ plane.} \tag{7.14}$$

Now demand that the Euclidean vacuum be a smooth, single-valued state on the plane. A rotation by angle $\theta$ is generated by the angular momentum $K_{\text{boost}}$ (analytically continued), and the plane closes up under rotation by $\theta = 2\pi$: going around once returns you to the start. Smoothness of the vacuum at the origin (the tip of the wedge, i.e. the horizon edge) forces the state to be *periodic under a $2\pi$ rotation*. But periodicity of the state under $e^{-\theta K_{\text{boost}}}$ with period $2\pi$ is exactly the KMS condition (7.10) at inverse temperature $2\pi$ in boost time:

$$e^{-2\pi K_{\text{boost}}} = \Delta, \qquad \text{i.e.} \qquad K = -\ln \Delta = 2\pi\, K_{\text{boost}}. \tag{7.15}$$

The $2\pi$ is the circumference of a full circle. It is not fitted to anything; it is the angle you turn through to come back to where you started. This is the deepest reason the number $2\pi$ pervades the theory — the Unruh temperature $T = \hbar a/2\pi c k_B$, the diamond temperature below, the MOND scale $a_0 = cH_0/2\pi$ of Chapter 14: all of them are one full Euclidean revolution, seen from different regions.

**Worked Example 7.2 — Recovering the Unruh temperature from the modular $2\pi$**

Bisognano–Wichmann says the vacuum is KMS at $\beta_{\text{mod}} = 2\pi$ in *boost* time (the dimensionless rapidity $\eta$). An observer with proper acceleration $a$ measures proper time $\tau$ along the orbit, related to rapidity by $\eta = a\tau/c$ (the defining property of uniform acceleration, Chapter 3). The KMS period in *proper* time is therefore

$$\beta_{\text{phys}} = \frac{c}{a}\,\beta_{\text{mod}} = \frac{2\pi c}{a}, \qquad k_B T = \frac{1}{\beta_{\text{phys}}}\hbar_{\text{-restored}} = \frac{\hbar a}{2\pi c}, \tag{7.16}$$

restoring $\hbar$ and $k_B$ by dimensions. This is the Unruh temperature of Chapter 4, now derived as a corollary of Tomita–Takesaki: the accelerated observer's "bath" is the vacuum's modular flow, and its temperature is fixed by the single geometric $2\pi$ of (7.15). No new input entered; the whole thermodynamics of horizons is one full turn of the Euclidean plane.

▶ RUN IT: `TOOLS/verify-paper-numerics.py` (section N: modular $\beta = 2\pi$; Unruh $T = \hbar a/2\pi c k_B$)

**Q 7.5 *"Bisognano–Wichmann is about flat space and a boost. The theory is about a gravitating mass in a laboratory. What does an eternally accelerating observer have to do with a microgram grain?"***

Everything, by the equivalence principle, and this is the bridge the whole book walks. The wedge is the flat-space, exactly-solvable member of a family; a finite causal diamond (§7.6) is the tabletop-sized member, and a gravitating region is the member with curvature turned on. In each, the claim is the same: the modular Hamiltonian of the physical vacuum is $2\pi$ times the generator of the observer's physical time. Bisognano–Wichmann *proves* it where the geometry is exact and the matter free; the diamond extends it to a finite region for conformal matter ([PROVEN], §7.6); and $K = 2\pi H_{\text{phys}}$ for the dynamical gravitational diamond is the [CONJECTURED] generalization that Chapter 13 works on. The grain enters because *its* two branches source two slightly different geometries, hence two slightly different modular Hamiltonians, and the phase between them winds at the rate $2\pi\,\Delta H_{\rm phys}/\hbar = E_{\rm G}/\hbar$ (Chapter 11). The eternal accelerator is not a detour; it is the one case where the number can be nailed down with complete rigor, and the rest of the theory is the attempt to carry that number into the laboratory.

## 7.6 The finite causal diamond

The Rindler wedge is infinite — it stretches to spatial infinity — and its modular Hamiltonian, the boost, is a global object. A laboratory sees a finite region. The natural finite analogue of a wedge is a *causal diamond* $\mathcal{D}_R$: the set of points that can both be reached from and return to a central worldline of proper duration $2R$, a double cone of radius $R$. Does its vacuum modular flow remain geometric? For *conformal* matter — a field theory invariant under local rescalings of the metric — the answer is yes, and it is a theorem of Hislop and Longo.

**Theorem 7.5** (Hislop–Longo diamond modular Hamiltonian)**.** *For the vacuum of a conformal field theory restricted to a causal diamond of radius $R$, the modular Hamiltonian is local — a weighted integral of the energy density $T_{00}$,*

$$K \;=\; 2\pi\int_{r<R} d^3x\, f(r)\, T_{00}(x), \qquad f(r) \;=\; \frac{R^2-r^2}{2R}, \tag{7.17}$$

*and generates the conformal-Killing flow $\zeta = \frac{\pi}{R}\big[\frac{1}{2}(R^2-t^2-r^2)\partial_t - tr\,\partial_r\big]$, which maps the diamond onto the static cylinder $\mathbb{R}\times H^3$ (hyperbolic space of curvature radius $R/2$) — the Casini–Huerta–Myers (CHM) map — where $\zeta$ is an exact time-translation Killing vector.* [PROVEN] *(for conformal matter).*

The structure of (7.17) is worth reading against Bisognano–Wichmann. There the modular Hamiltonian was $2\pi\times$(boost energy), an integral of $T_{00}$ weighted by the distance from the horizon. Here it is $2\pi\times$(a local energy), weighted by $f(r) = (R^2-r^2)/2R$ — the same $2\pi$, the same "$T_{00}$-weighted-by-position" shape, now for a bounded region. The weight $f(r)$ vanishes at the edge $r = R$ (the horizon of the diamond, where the modular flow freezes) and is largest at the center. Its central value,

$$f(0) \;=\; \frac{R}{2}, \tag{7.18}$$

sets the local temperature an observer at the center reads off the modular flow: inverting the modular-to-proper conversion gives the Tolman-blueshifted local temperature $T_{\rm loc}(r) = 1/2\pi f(r)$, so at the center

$$T_{\rm loc}(0) \;=\; \frac{1}{2\pi f(0)} \;=\; \frac{1}{\pi R}. \tag{7.19}$$

Both (7.18) and (7.19) are already checked numerically by the repository's harness (the graviton-modular-flow paper's "CHM weight at center $f(0) = R/2$" and "central temperature $1/(\pi R)$").

▶ Run it: `TOOLS/verify-paper-numerics.py` (section N: diamond $f(0) = R/2$, $T_{\rm loc}(0) = 1/\pi R$)

**Honest scope.** Theorem 7.5 requires the matter to be *conformal*. For a generic (non-conformal) field the diamond modular Hamiltonian is *not* the simple local expression (7.17) — the CHM map, which relies on Weyl invariance, is unavailable. The physical graviton is non-conformal, which is exactly the obstruction Chapter 13 confronts; its partial resolution

> (the graviton reduces to two conformally-behaved scalar towers) is [PROVEN AT CFT PROXY], not for the full field. We flag this here rather than bury it: the diamond result is exact for a CFT and a controlled approximation otherwise.

The diamond is the object Chapter 13 works with, for one reason: it is finite, so its modular Hamiltonian $K$ is the modular Hamiltonian *of a laboratory*, and the local temperature (7.19) is the temperature a tabletop observer would ascribe to the vacuum in her diamond. When a mass superposition sits near the center, its two branches shift $K$ by the same $E_G$, and — because both the modular gap $\Delta K = 2\pi f(0)\, E_G/\hbar$ and the observer's clock rate $d\tau/ds = 2\pi f(0)$ carry the same factor $f(0) = R/2$ — the arbitrary diamond size $R$ *cancels* from the physical phase rate, leaving $E_G/\hbar$ independent of $R$. That IR-safety (Exercise 7.3) is why a formula about an infinite region can predict a finite laboratory number.

## 7.7 The crossed product: giving the observer a clock

We arrive at the resolution of Chapter 6's catastrophe. A local algebra in quantum field theory is a type III$_1$ factor. Chapter 6 made the cost of this vivid: a type III$_1$ factor has *no trace*, and therefore no density matrix $\rho$, no von Neumann entropy $-\operatorname{Tr}(\rho \ln \rho)$, and — fatally for us — no way to define the *rate* of change of an entropy. The static overlap phase of gravitational decoherence (Chapter 2's visibility) lives happily in type III; the *rate* does not, because a rate is a $d/dt$ of a trace that is not there. To get a rate we must enlarge the algebra until a trace reappears. The tool is the crossed product, and the physics of the enlargement is: *give the observer a clock.*

### 7.7.1 The construction, told physically

The state already carries a flow, $\sigma_t$, its modular time (§7.3). The trouble is that in type III this flow has no generator inside the algebra — $K$ acts on the Hilbert space but is not an observable the region owns, precisely because there is no total energy operator in a type III factor. The crossed-product idea (Chandrasekaran–Longo–Penington– Witten, building on Takesaki) is to *adjoin the generator* as a new physical degree of freedom: introduce an observer with her own clock, a canonical pair $(X, P)$ where $P$ is the observer's energy and $X$ the reading of her clock, and *gauge* the modular flow by identifying "flow forward by modular time $t$" with "advance the observer's clock by $t$."

Concretely, the crossed product $M \rtimes_\sigma \mathbb{R}$ is the algebra generated by $M$ together with the clock, with the modular flow absorbed into a constraint that ties the observer's energy $P$ to the modular Hamiltonian:

$$\widehat{M} \;=\; M \rtimes_\sigma \mathbb{R} \;=\; \{\, M \text{ and the clock } (X, P)\,\} \Big/ \,(P = K + \text{observer term}). \tag{7.20}$$

The constraint is the operational content of "the observer's clock *is* the modular clock." What Takesaki proved is that this enlargement changes the type, decisively.

**Theorem 7.6** (Takesaki; crossed product by modular flow)**.** *The crossed product of a type III$_1$ factor by its own modular automorphism group is a type II$_\infty$ von Neumann algebra. It carries a faithful normal semifinite* trace*; hence density matrices $\hat\rho$ and a finite von Neumann entropy*

$S = -\operatorname{tr}(\hat\rho \ln \hat\rho)$ *exist on* $\widehat{M}$. [PROVEN] *(unconditional; needs no isometry, no geometry, no conformal assumption).*

The type dropped from $\mathrm{III}_1$ to $\mathrm{II}_\infty$, and with it a trace was born. Read the physics: the observer's clock supplied the "missing dimension" along which the modular flow acts, and modding out by that flow (the constraint) leaves a type II algebra that behaves, for entropy purposes, almost like an ordinary quantum system. A further restriction — bounding the observer's energy below, physically reasonable — descends $\mathrm{II}_\infty \to \mathrm{II}_1$, a factor with a *finite* trace and a genuinely normalizable maximal-entropy state.

$$\underbrace{M}_{\mathrm{III}_1:\ \text{no trace, no } S} \xrightarrow{\text{adjoin observer's clock}} \underbrace{\widehat{M} = M \rtimes_\sigma \mathbb{R}}_{\mathrm{II}_\infty:\ \text{trace, hence } S} . \tag{7.21}$$

### 7.7.2 The entropy is the generalized entropy

The payoff closes the arc that began in Chapter 5. The (renormalized) von Neumann entropy of a state on the type II crossed product $\widehat{M}$ is not an abstract new quantity: it *equals* the generalized entropy,

$$S(\hat\rho) = \frac{A}{4G\hbar} + S_{\text{ext}} + \text{const}, \tag{7.22}$$

where $A$ is the area of the region's edge, $S_{\text{ext}}$ the entanglement entropy of the matter outside, and the additive constant a scheme convention (only differences and monotonicity are physical). The area law that Chapter 5 *postulated* as Axiom I is, in this algebraic setting, a *theorem* about the trace on $\widehat{M}$. The type III divergence of the bare entanglement entropy — the reason Chapter 2 insisted only relative entropy survives — is exactly what the crossed-product renormalization absorbs into the finite $A/4G\hbar$.

> **What is rigorous and what is the frontier.** Takesaki's theorem (Theorem 7.6) — crossed product of $\mathrm{III}_1$ by modular flow is $\mathrm{II}_\infty$ with a trace — is [PROVEN], unconditionally. Its application to a *static* geometric flow (the de Sitter horizon, the CHM cylinder for conformal matter) is [PROVEN] at the CFT-proxy standing. Its application to the *dynamical gravitational diamond* — where the modular flow is conformal-Killing rather than an exact isometry, and the matter is the non-conformal graviton — is the open construction R1 (Chapter 13). The honest status: the machine that manufactures a trace is a theorem; installing it on a gravitating tabletop diamond is the theory's one remaining engineering problem.

**Q 7.6 *"Adjoining an "observer" to make the mathematics work sounds like a fudge. Where does the observer physically come from?"***

The observer is not decoration; she is forced. A type III region has no total energy operator of its own — ask "how much energy is in the wedge?" and the answer is a divergence, not a number, because the vacuum is infinitely entangled across the boundary at every scale (Chapter 2). The only way to speak of energy, entropy, or a rate for the region is *relationally*: pick a reference system whose energy you *can* track — a clock — and measure the region's quantities against it. That reference is the observer, and the constraint (7.20) says nothing more than "energy is conserved between the region and the clock." In gravity this is not optional: the total Hamiltonian of a closed universe *is* constrained to vanish (the Wheeler–DeWitt constraint, Chapter 11), so every physical quantity is automatically relational, measured against some subsystem playing the clock. The crossed product is the mathematics of that fact. The observer is the theory taking seriously that there is no view from outside a gravitating region.

### 7.7.3 The Modular Diamond Principle

We have now built, from finite matrices upward, every object the theory's foundation names. It is worth stating the foundation in one sentence, so the next chapters have a target.

**Principle 7.7** (Modular Diamond Principle)**.** *The physical content of a small causal diamond $\mathcal{D}$ is the observer-dressed modular structure of its vacuum: the gravitationally constrained algebra of $\mathcal{D}$ is the crossed product of its field algebra by its observer's modular flow (§7.7), its physical state sits at entanglement equilibrium (Chapter 10), and* its modular flow at $\beta_{\text{mod}} = 2\pi$ is $\mathcal{D}$'s physical time*.*

Three clauses — algebra, state, time — and each is an object of this chapter or the two that follow. The algebra clause is the crossed product (Theorem 7.6), whose entropy is $S_{\text{gen}}$ ((7.22)); the state clause is entanglement equilibrium, from which Chapter 10 recovers Einstein's equations; the time clause is the identity $K = 2\pi H_{\text{phys}}$. Chapter 9 promotes these to the three primitive axioms. The time clause carries the whole conjectural burden: it is [PROVEN] in flat space (Bisognano–Wichmann, (7.13)), [PROVEN] for the conformal diamond (Hislop–Longo, (7.17)), [PROVEN] in expectation values and every solvable model, and [CONJECTURED] as a full operator equation for the dynamical gravitational diamond — the one place experiment, not mathematics, will decide (Chapter 15).

## Chapter FAQ

**Q: Is the modular Hamiltonian $K$ an observable one could in principle measure?** For a general region, no — and that is the point of the crossed product. In a type III factor $K = -\ln \Delta$ acts on the Hilbert space but is not *affiliated* to the region's algebra: it is not built from observables the region owns, so no apparatus confined to the region measures it. Only after adjoining the observer's clock (§7.7) does a physical, measurable generator appear — the observer's energy $P$, tied to $K$ by the constraint (7.20). The modular Hamiltonian is measurable only relationally, against a clock. This is a feature: it is why the theory's central equation is a statement about the *relation* between $K$ and a physical generator, not about $K$ alone.

**Q: In the finite example $K = \beta h$ had a temperature $\beta$; in Bisognano–Wichmann it was exactly $2\pi$; in the diamond there is a local $T_{\text{loc}}(r)$. Which is the temperature?** All three, consistently. The dimensionless modular period is always $\beta_{\text{mod}} = 2\pi$ (or 1 in the units where the temperature is absorbed into the flow) — that is a theorem, the Euclidean $2\pi$ of §7.5.1. What varies is the conversion from modular time to *proper* time, and that conversion is a local, position-dependent redshift factor: it is constant for the uniformly accelerated Rindler observer, and it is the Tolman factor $2\pi f(r)$ for the diamond observer, giving $T_{\text{loc}}(r) = 1/2\pi f(r)$. One universal $2\pi$, dressed by local geometry into the temperature a given observer reads.

**Q: Why does the type have to drop all the way to II, rather than to the familiar type I of ordinary quantum mechanics?** Because a genuine field-theoretic region cannot be type I — type I would mean the region factorizes cleanly from its complement, with a tensor-product Hilbert space and a finite entanglement entropy, and the short-distance entanglement across the boundary forbids that (Chapter 6). The crossed product adds exactly one clock degree of freedom, the least possible, and lands on type II: it has a trace and a renormalized entropy (unlike III), but that entropy is defined only up to an additive constant and only differences are physical (unlike I). Type II is the honest middle: enough structure for a rate, not so much as to pretend the boundary entanglement went away.

**Q: Bisognano–Wichmann is exact and beautiful. Why isn't the whole theory then just proven?** Because Bisognano–Wichmann holds for a *static wedge in flat space with free matter*, and a laboratory grain gravitates. Turning on gravity does three things at once: the region becomes finite (handled — the diamond, §7.6), the modular flow becomes conformal-Killing rather than an exact isometry (handled for conformal matter, open for the graviton), and the matter that sources the geometry is the non-conformal graviton (the open residue R1). Each step is a theorem or a controlled model in this chapter; the composition, for the physical dynamical diamond, is the one thing not yet proven. Chapter 13 is the ledger of exactly what remains.

## Exercises

**Exercise 7.1.** Let $\rho = \mathrm{diag}(p, 1-p)$ with $0 < p < 1$ be a qubit state, purified as in (7.1). (a) Write down $\Delta = \rho \otimes \rho^{-1}$ explicitly as a $4 \times 4$ diagonal matrix and list its four eigenvalues. (b) Compute the modular Hamiltonian $K = -\ln \Delta$ and verify it has the form $\beta(h \otimes \mathbb{1} - \mathbb{1} \otimes h)$ for suitable $\beta, h$; identify the "modular energy gaps" $\omega_{ij} = k_i - k_j$ (the exponents appearing in $\Delta^{it}$). (c) Confirm $\Delta\,|\Omega\rangle = |\Omega\rangle$ (the state is modular-flow invariant) directly from (7.1).

**Exercise 7.2.** Take a single harmonic oscillator, $H = \omega_0\, a^\dagger a$, in a Gibbs state at inverse temperature $\beta$. For $A = a^\dagger$, $B = a$, compute both orderings $F(t) = \langle A\, \sigma_t(B)\rangle$ and $G(t) = \langle \sigma_t(B)\, A\rangle$ with $\sigma_t(a) = e^{iHt} a\, e^{-iHt}$, and verify the KMS relation (7.10) explicitly by showing $F(t + i\beta) = G(t)$. (Hint: $\sigma_t(a) = e^{-i\omega_0 t} a$; you will need $\langle aa^\dagger\rangle$ and $\langle a^\dagger a\rangle$ for the thermal oscillator.)

**Exercise 7.3.** For the Hislop–Longo diamond (7.17) with a two-branch mass superposition of energy gap $E_{\mathrm{G}}$ placed near the center, the modular gap between branches is $\Delta K = 2\pi f(0)\, E_{\mathrm{G}}/\hbar$ and the observer's proper-time clock advances at $d\tau/ds = 2\pi f(0)$. (a) Using $f(0) = R/2$, write both quantities in terms of $R$. (b) Show that the physical phase rate $d\Phi/d\tau = \Delta K/(d\tau/ds)$ is independent of the diamond radius $R$, and equals $E_{\mathrm{G}}/\hbar$. (c) Explain in one sentence why this "IR-safety" is what allows a formula built on an arbitrary-sized region to predict a definite laboratory decoherence rate.

## Solutions

*Solution (Exercise 7.1).* (a) With $\rho = \mathrm{diag}(p, 1-p)$ and $\rho^{-1} = \mathrm{diag}(1/p, 1/(1-p))$, the tensor product $\Delta = \rho \otimes \rho^{-1}$ is diagonal with entries $\{p \cdot \frac{1}{p},\ p \cdot \frac{1}{1-p},\ (1-p) \cdot \frac{1}{p},\ (1-p) \cdot \frac{1}{1-p}\} =$

$\{1,\ \frac{p}{1-p},\ \frac{1-p}{p},\ 1\}$. Note the reciprocal pair $\frac{p}{1-p}, \frac{1-p}{p}$ and the two unit eigenvalues — the spectrum is symmetric under $\Delta \to \Delta^{-1}$, as $J$ demands. (b) The Hamiltonian $h = -\frac{1}{\beta}\ln\rho$ has, at $\beta = 1$, $h = \mathrm{diag}(-\ln p, -\ln(1-p))$, and $K = -\ln\Delta = h \otimes \mathbb{1} - \mathbb{1} \otimes h$, i.e. $\beta = 1$. The nonzero modular gaps are $\omega = \pm \ln\frac{p}{1-p}$ (and 0, doubly). At $p = 1/2$ all gaps vanish: the maximally mixed state has trivial modular flow (its $\rho$ commutes with everything). (c) $\Delta\,|\Omega\rangle = (\rho \otimes \rho^{-1}) \sum_k \sqrt{p_k}\,|kk\rangle = \sum_k \sqrt{p_k}\,(\rho\,|k\rangle) \otimes (\rho^{-1}\,|k\rangle) = \sum_k \sqrt{p_k}\,p_k\,p_k^{-1}\,|kk\rangle = \sum_k \sqrt{p_k}\,|kk\rangle = |\Omega\rangle$: the $\rho$ and $\rho^{-1}$ eigenvalues cancel on the diagonal entangled vector, so $|\Omega\rangle$ is invariant, as Theorem 7.1 requires.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch7 modular spectrum)`

*Solution (Exercise 7.2).* With $\sigma_t(a) = e^{-i\omega_0 t}a$, the two orderings are

$$F(t) = \langle a^\dagger\,\sigma_t(a)\rangle = e^{-i\omega_0 t}\langle a^\dagger a\rangle = e^{-i\omega_0 t}\,\bar{n}, \qquad G(t) = \langle \sigma_t(a)\,a^\dagger\rangle = e^{-i\omega_0 t}\langle a a^\dagger\rangle = e^{-i\omega_0 t}(\bar{n}+1),$$

where $\bar{n} = 1/(e^{\beta\omega_0} - 1)$ is the thermal occupation and $[a, a^\dagger] = \mathbb{1}$ gave $\langle aa^\dagger\rangle = \bar{n} + 1$. Now continue $F$ upward across the strip, $t \to t + i\beta$, as the KMS relation (7.10) instructs: the flow factor becomes $e^{-i\omega_0(t+i\beta)} = e^{-i\omega_0 t}e^{\beta\omega_0}$, so

$$F(t + i\beta) = e^{-i\omega_0 t}\,e^{\beta\omega_0}\,\bar{n} = e^{-i\omega_0 t}\,(\bar{n}+1) = G(t),$$

using $e^{\beta\omega_0}\bar{n} = e^{\beta\omega_0}/(e^{\beta\omega_0} - 1) = \bar{n} + 1$. The boundary relation holds exactly, and the algebraic identity that made it work, $e^{-\beta\omega_0}(\bar{n}+1) = \bar{n}$, is *detailed balance*: emission (the $\bar{n}+1$ ordering) exceeds absorption (the $\bar{n}$ ordering) by precisely the Boltzmann factor $e^{\beta\omega_0}$, and the KMS strip's one-sided orientation ($0 < \mathrm{Im}\,t < \beta$, fixed by convergence of the continuation) is where that factor lives. The detailed-balance ratio between the two orderings *is* the thermal content of the state.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch7 KMS detailed balance)`

*Solution (Exercise 7.3).* (a) With $f(0) = (R^2 - 0)/2R = R/2$: the modular gap is $\Delta K = 2\pi f(0)\,E_G/\hbar = 2\pi \cdot \frac{R}{2} \cdot E_G/\hbar = \pi R\,E_G/\hbar$, and the clock rate is $d\tau/ds = 2\pi f(0) = \pi R$. (b) The physical phase rate is the ratio,

$$\frac{d\Phi}{d\tau} = \frac{\Delta K}{d\tau/ds} = \frac{\pi R\,E_G/\hbar}{\pi R} = \frac{E_G}{\hbar},$$

in which the factor $\pi R$ — carrying *all* the $R$-dependence — cancels top and bottom. The phase rate is independent of $R$ (verified in the harness across $\sim 10^6$ decades in $R$). (c) The diamond radius $R$ is an arbitrary infrared regulator, an artifact of choosing how large a region to draw around the laboratory; its cancellation means the predicted decoherence rate $E_G/\hbar$ is a property of the grain alone, not of the bookkeeping region — which is why an object defined on a region of any size can predict the definite 1.6 ns benchmark of Chapter 1.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch7 diamond IR-safety)`

**What to carry forward**

1. **Every faithful state carries a clock.** From the Tomita operator $S = J\Delta^{1/2}$ a von Neumann algebra with cyclic–separating vector $|\Omega\rangle$ produces a canonical modular flow $\sigma_t = \Delta^{it}(\cdot)\Delta^{-it}$ with modular Hamiltonian $K = -\ln\Delta$. On a Gibbs state $\Delta = \rho \otimes \rho^{-1}$ and $K = \beta\,h$: *modular time is physical time divided by temperature.* [PROVEN]

2. **Thermality is intrinsic, and the number is $2\pi$.** The modular flow is the unique flow making $|\Omega\rangle$ KMS (imaginary-time-periodic correlators), no bath required. For the Rindler wedge in the Minkowski vacuum, Bisognano–Wichmann gives $K = 2\pi K_{\text{boost}}$ exactly, the $2\pi$ forced by one full Euclidean rotation — the flat-space instance of $K = 2\pi H_{\text{phys}}$, and the source of every $2\pi$ in the book. [PROVEN]

3. **The crossed product gives the observer a clock and repairs type III.** Adjoining the modular generator as a physical degree of freedom turns the traceless type $\text{III}_1$ algebra into a type II algebra with a trace (Takesaki), whose entropy is the generalized entropy $A/4G\hbar + S_{\text{ext}}$ — Axiom I as a theorem. Applied to the finite conformal diamond ($f(r) = (R^2 - r^2)/2R$, $f(0) = R/2$, $T_{\text{loc}}(0) = 1/\pi R$) it is [PROVEN]; applied to the dynamical gravitational diamond it is the open construction R1, and the time clause $K = 2\pi H_{\text{phys}}$ is the theory's central [CONJECTURED] equation (Chapter 13).

# Part III

# The Theory

# 8
# The Correspondence Stated

**Synopsis.** Part I found the impasse; Part II built the toolkit — density matrices, horizons, quantum fields, and, in Chapters 6–7, the one object the theory is actually written in: a single von Neumann algebra carrying a single state, with a commutant and a modular clock. This chapter spends that toolkit on one sentence. *Quantum mechanics and general relativity are two complementary restrictions of one state on one algebra.* We make that precise — a matter subalgebra, its commutant the geometry subalgebra, one global state restricting to each — and we retire the cylinder-and-shadows analogy of Chapter 1 by replacing it with a dictionary. We then *derive* the map the rest of Part III computes with: linearity plus branch-wise sourcing turns a matter superposition into a matter–geometry entangled state, and tracing out the geometry recovers the decoherence functional of Chapter 2, now with spacetime as the environment. Finally we are honest about what this frame is not: it does not quantize the metric, does not solve the measurement problem, does not replace either parent theory.

**You will be able to.** state the correspondence as a relation between a subalgebra and its commutant; derive the duality map $\sum_n c_n |\psi_n\rangle \mapsto \sum_n c_n |\psi_n\rangle \otimes |g_n\rangle$ from linearity and branch-wise sourcing, and read the reduced matter state off it; say precisely where QGC agrees with and departs from string theory, loop quantum gravity, and dynamical collapse; and name the four things the frame does not claim.

**Prerequisites.** Chapters 1 (the impasse and its branch-wise reading), 2 (partial trace, the visibility law), 6 (algebra, commutant, factor), and 7 (crossed product, the modular clock). No new mathematics is introduced.

## 8.1 The claim, in one sentence, made precise

Chapter 1 ended on a wager: that "quantum mechanics" and "general relativity" are not two theories awaiting a merger but two appearances of one underlying structure, the way a cylinder throws a circular shadow on one wall and a rectangular one on another. That was an analogy, and we promised on contact with mathematics to retire it. Part II built the mathematics. We can now state the same idea as a relation between objects we have actually constructed, with no shadows in it.

The underlying structure is the one built in Chapter 6: a single von Neumann algebra $\mathcal{M}$ of observables, acting on a Hilbert space $\mathcal{H}$, carrying a single physical state $\omega$ (an expectation functional $A \mapsto \omega(A)$, Definition 6.6). One algebra, one state — that is the whole world, as far as any observer is concerned. "Quantum mechanics" and "general relativity" are then not two structures but two *restrictions* of this one: two subalgebras to which $\omega$ is cut down.

To name them we need the operational content of "matter" and "geometry." Take $\mathcal{M}_{\mathrm{mat}} \subset \mathcal{M}$ to be the subalgebra generated by the matter observables an observer can act on — positions and momenta of grains, field amplitudes, the things a laboratory manipulates. Take $\mathcal{M}_{\text{geo}}$ to be the subalgebra of geometric observables — the gravitational field the matter sources, the distances and curvatures a surveyor measures with light rays and test masses (the operational readouts of Chapter 3). The central structural input, which Chapters 11–13 will earn in detail, is that these two are not independent: the geometry an observer can measure is exactly the geometry *sourced by* the matter she cannot reach, and vice versa. In the language of Chapter 6, geometry is the *commutant* of matter:

$$\mathcal{M}_{\text{geo}} \;=\; (\mathcal{M}_{\text{mat}})' \;\equiv\; \{\, B : [B, A] = 0 \text{ for all } A \in \mathcal{M}_{\text{mat}} \,\}, \tag{8.1}$$

the set of all observables that commute with every matter observable (Definition 6.2). One line of physical meaning: an observer who has seized every matter handle in $\mathcal{M}_{\text{mat}}$ still cannot, from inside $\mathcal{M}_{\text{mat}}$, read the geometric data in its commutant — geometry is precisely the information about the world that is invisible to the matter algebra, carried in correlations rather than in any matter observable. This is the same inside/outside split that gave the horizon an entropy in Chapter 5, now naming the two sides "matter" and "geometry."

The single state $\omega$ restricts to each subalgebra by simple forgetting:

$$\omega_{\text{QM}} \;\equiv\; \omega\big|_{\mathcal{M}_{\text{mat}}}, \qquad \omega_{\text{GR}} \;\equiv\; \omega\big|_{\mathcal{M}_{\text{geo}}}, \tag{8.2}$$

where $\omega|_{\mathcal{M}_{\text{mat}}}(A) = \omega(A)$ evaluated only on $A \in \mathcal{M}_{\text{mat}}$, and likewise for geometry. The restriction $\omega_{\text{QM}}$ is what Chapter 2 called the reduced state of matter — a density matrix if $\mathcal{M}_{\text{mat}}$ is a type I factor, a genuinely algebraic state otherwise. It is the object ordinary quantum mechanics computes with. The restriction $\omega_{\text{GR}}$ is the state of geometry, and it has one further property that makes it look like general relativity rather than like a second quantum system: it is *approximately classical*, meaning $\mathcal{M}_{\text{geo}}$ is approximately abelian in the state $\omega$,

$$\omega([B_1, B_2]) \;\approx\; 0 \qquad \text{for } B_1, B_2 \in \mathcal{M}_{\text{geo}}, \tag{8.3}$$

so that geometric observables have joint values with negligible quantum uncertainty. One line of meaning: a nearly abelian algebra in a given state is a nearly classical one — its observables commute, so they can be measured together and behave like c-numbers, which is exactly what a classical metric $g_{\mu\nu}(x)$ is. General relativity is what the world's one state looks like after you forget everything except the (nearly commuting) geometric observables. Chapter 2's decoherence is the reason (8.3) holds: the environment that makes the geometric branches decohere is why geometry looks classical, and §8.2 shows this is not a coincidence but the same computation.

**Principle 8.1** (Quantum–geometric correspondence)**.** *There is one von Neumann algebra $\mathcal{M}$ carrying one state $\omega$. Quantum mechanics is the restriction $\omega|_{\mathcal{M}_{\text{mat}}}$ of $\omega$ to a matter subalgebra $\mathcal{M}_{\text{mat}} \subset \mathcal{M}$; general relativity is the restriction $\omega|_{\mathcal{M}_{\text{geo}}}$ to the commutant $\mathcal{M}_{\text{geo}} = (\mathcal{M}_{\text{mat}})'$, which is approximately abelian in $\omega$ and hence approximately classical. Neither restriction is the whole state, and neither is prior: they are two complementary shadows cast by one $\omega$, and the physics of the seam is the physics of the correlation between them that both restrictions discard.*

The word to underline in Principle 8.1 is *restriction*. The move it makes is not to *quantize gravity* — we do not promote $g_{\mu\nu}$ to a new operator field and hunt for its quanta. It is to *recognize two shadows of one object* — to notice that the matter algebra and its commutant were there all along inside a single $\mathcal{M}$, and that "quantizing the metric" was the wrong question because geometry was never a separate thing to quantize. It is the commutant of matter. This is the Maldacena move made small: not a metaphor about cylinders but a dictionary between two restrictions, with a definite object $(\mathcal{M}, \omega)$ on which both sides are computed.

**The dictionary that replaces the cylinder.** Chapter 1 spoke of one object casting two shadows. Here is the same sentence with the analogy discharged:

| *One object* | *Quantum shadow* | *Geometric shadow* |
|---|---|---|
| Algebra $\mathcal{M}$ | subalgebra $\mathcal{M}_{\text{mat}}$ | commutant $\mathcal{M}_{\text{geo}} = (\mathcal{M}_{\text{mat}})'$ |
| State $\omega$ | $\omega\|_{\mathcal{M}_{\text{mat}}}$ | $\omega\|_{\mathcal{M}_{\text{geo}}}$ |
| Structure of state | density matrix / coherence | metric $g_{\mu\nu}$ (nearly abelian) |
| The seam | the correlation both restrictions discard | |

Every row is an object built in Part II, not an image. The "cylinder" was the promise; this table is the payment.

**Q 8.1** ***"You said geometry is the commutant of matter — but the commutant is a purely quantum, kinematical object. How can it*** **be** ***spacetime geometry rather than just some leftover operators?"***

The identification is a physical claim, and it is the load-bearing one of the whole book; (8.1) is where the theory puts its weight, not a definition we are free to make. What makes it more than bookkeeping is that the commutant is not an inert leftover: in Chapter 7 we saw that a state equips its algebra with a *modular flow*, an intrinsic clock (Principle 7.2), and the Bisognano–Wichmann theorem showed that for the vacuum on a wedge this abstract clock *is* a geometric boost — an isometry of spacetime. The commutant, in other words, already carries geometric data; modular theory is the bridge from "operators that commute with matter" to "the geometry matter lives in." The correspondence asserts that this bridge is exact enough to identify the commutant's structure with $g_{\mu\nu}$ in the semiclassical regime. That assertion is what Axiom III (Chapter 9) commits to and what the central equation $K = 2\pi H_{\text{phys}}$ of Chapter 7 makes quantitative — and it is [CONJECTURED] at the operator level, exactly as flagged there. The correspondence is stated here; it is earned, to the extent it can be, over Chapters 11–13.

## 8.2 The duality map, derived

Principle 8.1 is a statement about structure. To compute with it we need the concrete state $\omega$ produces when matter is put in superposition, and this we can *derive* — from two ingredients we already trust, and nothing else. The derivation is the formal version of the branch-wise picture Chapter 1 flagged as "where the theory starts."

*Ingredient one: linearity.* Quantum mechanics is linear (Chapter 1): if $|\psi_1\rangle$ and $|\psi_2\rangle$ are possible matter states, so is any superposition $\sum_n c_n |\psi_n\rangle$, and the dynamics acts on the sum term by term. Nothing in this book weakens this — QGC adds no nonlinear or stochastic term to the Schrödinger equation (a promise we redeem, and contrast with collapse models, in §8.4).

*Ingredient two: branch-wise sourcing.* A matter state $|\psi_n\rangle$ with a definite mass configuration sources a definite geometry — this is general relativity's one demand, the classical map "one source in, one geometry out" of Chapter 1. Call the geometry that the branch $|\psi_n\rangle$ sources $|g_n\rangle \in \mathcal{H}_{\text{geo}}$; operationally, $|g_n\rangle$ is the coherent state of the gravitational field whose expectation values reproduce the metric sourced by the mass configuration in $|\psi_n\rangle$ (Chapter 11 builds these $|g_n\rangle$ explicitly). For a *single* branch there is no puzzle: the pair $|\psi_n\rangle \otimes |g_n\rangle$ is what semiclassical gravity describes correctly, matter and its own geometry riding together.

Now combine the two. Prepare matter in the superposition $\sum_n c_n |\psi_n\rangle$ and let each branch source its geometry. Linearity forbids us from replacing the superposition by any single average geometry — that was the Page–Geilker failure of Chapter 1. Linearity *requires* that the joint state be the superposition of the branch-wise pairs:

$$|\Psi\rangle \;=\; \sum_n c_n \, |\psi_n\rangle \otimes |g_n\rangle \;\in\; \mathcal{H}_{\text{mat}} \otimes \mathcal{H}_{\text{geo}}. \tag{8.4}$$

This is the *duality map* — the assignment $\sum_n c_n |\psi_n\rangle \mapsto \sum_n c_n |\psi_n\rangle \otimes |g_n\rangle$ promised in Chapter 1. One line of meaning: a mass genuinely in two places is genuinely accompanied by two geometries, in superposition, correlated branch by branch. This single equation is what the theory "starts from": it is neither semiclassical averaging (which would collapse the sum to one mean geometry) nor metric quantization (which would introduce graviton dynamics of its own); it is exactly the entangled state linearity forces once you grant that each branch sources its own geometry. The state (8.4) is generically *entangled* between matter and geometry — it is not a product $|\psi\rangle \otimes |g\rangle$ whenever more than one geometry appears — and here, as Chapter 2 warned, the whole of the physics happens inside that word.

**Q 8.2** ***"Isn't (8.4) just assuming the answer? You put the geometry into a superposition by hand and call it a derivation."***

The two ingredients are the assumptions; (8.4) is forced from them, and each ingredient is something you already accept for independent reasons. If you keep linearity (every interference experiment ever done) and you keep branch-wise sourcing (a definite mass makes a definite field, which is Newtonian gravity, not a new postulate), then the superposition principle applied to the pair "matter + its sourced geometry" gives (8.4) with no further choice — the same way linearity applied to "system + its scattered photon" gave the pre-measurement state (2.10) of Chapter 2. What you *cannot* do is keep both ingredients and *avoid* (8.4): the only escapes are to break linearity (collapse models — §8.4) or to deny that superposed masses source their fields branch-wise (the mean-field route that Page–Geilker killed). So the content of the "derivation" is precisely that it closes those exits. What is *not* derived, and stays open, is whether the branch geometries $|g_n\rangle$ are the full nonperturbative states or the coherent-state approximations of Chapter 11; we use the latter, and flag it there.

### 8.2.1 Tracing out geometry recovers the decoherence functional

Equation (8.4) is a bipartite pure state, so we may do to it exactly what Chapter 2 did to every bipartite pure state: ask what the matter observer, who reaches only $\mathcal{M}_{\text{mat}}$, can see. Her state is the reduced density matrix, the partial trace over geometry (2.6):

$$\rho_{\text{mat}} \;=\; \mathrm{Tr}_{\text{geo}}\,|\Psi\rangle\langle\Psi| \;=\; \sum_{m,n} c_m c_n^*\;|\psi_m\rangle\langle\psi_n|\;\langle g_n|g_m\rangle\,, \tag{8.5}$$

where the geometry overlaps $\langle g_n|g_m\rangle$ appear because tracing keeps only the diagonal of the geometry factor — the cross terms carry $\langle g_n|g_m\rangle$ exactly as, in Worked Example 2.2, they carried $\langle 1|0\rangle$. Read (8.5) in the matter basis: the diagonal populations $|c_n|^2$ are untouched (each $\langle g_n|g_n\rangle = 1$), while every *off-diagonal* element $|\psi_m\rangle\langle\psi_n|$ — every carrier of quantum coherence between distinct branches — is multiplied by the overlap of the geometries those branches source:

$$\left(\rho_{\text{mat}}\right)_{mn} \;=\; c_m c_n^*\;\langle g_n|g_m\rangle\,, \qquad \mathcal{V}_{mn} \;=\; |\langle g_n|g_m\rangle|. \tag{8.6}$$

This is the decoherence functional — and it is, letter for letter, the visibility law (2.12) of Chapter 2, with one substitution: the environment states $|E_{L,R}\rangle$ that were gas molecules or photons are now $|g_n\rangle$, coherent states of the gravitational field the mass itself sources. One line of meaning: the coherence between two mass positions survives exactly to the extent that the geometries they source are indistinguishable, $\langle g_n|g_m\rangle \to 1$, and dies as those geometries become orthogonal, $\langle g_n|g_m\rangle \to 0$. Gravitational decoherence is ordinary decoherence with spacetime as the which-path environment.

The logical chain is now complete and worth stating bare, because Chapters 11–12 are nothing but its last step executed:

1. Principle 8.1: one state $\omega$, matter and geometry as a subalgebra and its commutant.
2. Linearity + branch-wise sourcing $\Rightarrow$ the duality map (8.4): a matter superposition is a matter–geometry entangled state.
3. Partial trace over geometry (Chapter 2) $\Rightarrow$ the reduced matter state (8.5), whose off-diagonals are the geometry overlaps $\langle g_n|g_m\rangle$.
4. Therefore the decoherence of matter coherence is controlled by one number, $\langle g_n|g_m\rangle$ — the object Chapters 11 and 12 compute, yielding the rate $\Gamma_{\text{dec}} = C\,E_{\text{G}}/\hbar$ and the benchmark $\tau_{\text{dec}} = 1.6$ ns.

**Why this is not just environmental decoherence.** The arithmetic is identical to Chapter 2; the physics differs in one decisive way. Ordinary environments are *contingent* — better vacuum, colder walls, and $\langle E_n|E_m\rangle \to 1$, restoring coherence. The geometry in (8.6) is not contingent: by the correspondence, $|g_n\rangle$ is the commutant partner *forced* to exist by branch-wise sourcing, with no gauge freedom left to screen it (the Wheeler–DeWitt constraint of Chapter 11 is what removes that freedom). The which-path record cannot be refrigerated away because it is not a record the environment happens to keep — it is the geometry the mass must source. That is the difference between removable noise and the uncoolable floor of Chapter 15.

### 8.2.2 Where the two restrictions live in one algebra

The pure-state picture (8.4) is the working form, but it presupposes a tensor factorization $\mathcal{H}_{\text{mat}} \otimes \mathcal{H}_{\text{geo}}$ that Chapter 6 taught us to distrust for local field theory (local algebras are type $\mathrm{III}_1$, Principle 6.7, and admit no such clean factorization). The algebraic statement of Principle 8.1 survives that warning intact, because it never used a tensor product: it used a subalgebra $\mathcal{M}_{\text{mat}}$ and its commutant $\mathcal{M}_{\text{geo}} = (\mathcal{M}_{\text{mat}})'$, which exist for any $\mathcal{M}$. The duality map (8.4) is then the type I *approximation* to the exact statement — valid, as Chapter 11 will justify, because the crossed-product construction of Chapter 7 (Eq. (7.20)) converts the type III geometry algebra into a type II algebra with a trace, restoring exactly enough of the tensor-product bookkeeping to make $\langle g_n|g_m\rangle$ well defined. The reader should hold (8.4) as the computational front end and Principle 8.1 as the structure it approximates; the two agree wherever the semiclassical conditions of Chapter 11 hold, and it is only there that the theory makes claims.

## 8.3 Duality or correspondence? An honest naming

We have twice now called this a "correspondence" and once flirted with "duality" (the map (8.4) does, after all, look like a dictionary between a quantum description and a geometric one). The choice of word is not decoration; it commits us to a level of claim, and the honest level is worth pinning down before Part III leans on it.

A *strict duality*, in the sense the string-theory revolution fixed the term — AdS/CFT is the exemplar — asserts two *independently complete* descriptions related by an exact dictionary: either side computes any observable, at any coupling, and the two answers always agree. QGC does not meet that standard, and it is important to say so plainly. The geometry side of our correspondence is *not* independently complete: $\mathcal{M}_{\text{geo}}$ is built as the commutant of matter, its states $|g_n\rangle$ are sourced by matter configurations, and (8.3) holds only approximately and only in the semiclassical regime. There is a manifest hierarchy — the quantum description presupposes no geometry (a density matrix and an algebra are defined on any $\mathcal{H}$), while the geometric description presupposes quantum matter to source it. That asymmetry is real, and "duality" in the AdS/CFT sense would paper over it.

What *is* established is the content of Principle 8.1 itself: that there is one state on one algebra, restricting to a matter subalgebra and its commutant, with the geometry restriction approximately classical. This is a genuine and checkable structure — it is what (8.4) and (8.5) compute with — and it is honestly a *correspondence*: a precise relationship between a more fundamental description (quantum) and a derived one (geometric), in the same spirit as, but weaker than, AdS/CFT. Two familiar precedents license the word without overclaiming. Wave–particle "duality" is universally accepted yet is exactly this: one theory (quantum mechanics) producing two complementary phenomenologies, not two complete theories with an exact dictionary. And even the gold standard calls itself, in full, the anti-de Sitter/conformal-field-theory *correspondence*. If the strongest duality in physics is a correspondence, the bar for the bare word "duality" is one we decline to claim.

> **Status of the naming.** That QGC has the structure of Principle 8.1 — one $\omega$, $\mathcal{M}_{\text{mat}}$ and its commutant, geometry approximately abelian — is [DERIVED] in the semiclassical regime (it is the content of Chapters 11–10). That this rises to a *strict* duality — a categorical equivalence with an exact two-way dictionary valid at all scales, of the kind AdS/CFT enjoys — is [CONJECTURED] and, frankly, aspirational: it would require constructing the geometry-to-matter functor nonperturbatively, which the framework does

> not do. We use "correspondence" throughout precisely to keep these apart. The rhetorical cost — "duality" sounds more profound — is paid gladly; profundity is earned by proof, not claimed by vocabulary.

## 8.4 What the correspondence does not claim

A frame is defined as much by what it excludes as by what it asserts. Four non-claims must be stated explicitly, because each is a place where a reader could reasonably over-read Principle 8.1, and because the theory's honesty is its chief asset.

**It does not quantize the metric.** Nowhere in (8.4) or (8.5) did we promote $g_{\mu\nu}$ to a fundamental operator field, sum graviton loops, or attempt a UV completion. The states $|g_n\rangle$ are coherent states sourced by matter, used at tree level; the correspondence is a statement about the *semiclassical* regime (weak curvature, slow evolution, $|h_{\mu\nu}| \ll 1$), exactly the domain where Chapter 1's power-counting wall stands far off. What happens at the Planck scale — whether geometry is discrete, stringy, or something else — QGC takes no position on. Its wager, stated in Chapter 1, is that the *measurable* quantum-gravitational physics lies below that wall, in the decoherence of (8.5), and needs no UV theory to compute.

**It does not solve the measurement problem.** The reduced state (8.5) explains why coherence between mass-position branches decays; it does not explain why one branch is *realized*, nor whence the probabilities. When we read the decohered populations $|c_n|^2$ as probabilities we invoke the Born rule, which enters here exactly as it did in Chapter 2: [ASSUMED], an input and not an output, as in every laboratory application of quantum mechanics. QGC is a theory of dynamics at the quantum–gravity seam, and it never pretends to be a solution of the measurement problem — the two are adjacent questions, and the correspondence answers only the first.

**It does not replace general relativity or quantum mechanics.** Both parent theories are *recovered*, not overturned, as the two restrictions of Principle 8.1. General relativity is $\omega|_{\mathcal{M}_{\mathrm{geo}}}$ in the regime where (8.3) makes geometry classical — and Chapter 10 will show the Einstein equations themselves emerge from the same state via entanglement equilibrium. Quantum mechanics is $\omega|_{\mathcal{M}_{\mathrm{mat}}}$, exact and linear, with not one term added to its dynamics. The correspondence lives at the seam where both shadows matter at once; away from the seam it hands each parent theory back unmodified, which is why both continue to pass every test in their own courtrooms.

**It is not a completed theory.** Principle 8.1 is a frame, and the central identification that gives it teeth — geometry's modular clock is physical time, $K = 2\pi H_{\mathrm{phys}}$ — is [CONJECTURED] at the operator level (Chapter 7), reduced to one open construction (Chapter 13). The honest grade of the theory is B+: the energy scale $E_{\mathrm{G}} = GM^2/d$ is [PROVEN], the rate is motivated and bounded but not pinned, and there is zero experimental confirmation. We state the frame cleanly here so that the reader can see exactly which load-bearing beam is still under test.

### 8.4.1 Where QGC sits among its neighbours

It helps to locate the correspondence against the three best-developed responses to the impasse. In each case QGC shares the diagnosis and differs in the cure; the comparisons are meant to be fair, not competitive.

*String theory* answers the impasse by completing gravity in the ultraviolet: it quantizes a fundamental object (the string) whose spectrum contains the graviton, tames the power-counting divergences of Chapter 1 through extended structure, and pays for this with extra dimensions and a vast landscape of vacua. QGC shares string theory's respect for the divergence problem but declines its project entirely: it makes *no* UV claim, adds no dimensions, and quantizes nothing new. Where string theory asks "what is the correct short-distance theory of gravity," QGC asks "what is the measurable long-distance consequence of matter and geometry being one state," and bets the answer needs no short-distance theory at all. The two are not rivals so much as addressed to different questions.

*Loop quantum gravity* answers by quantizing geometry directly, but nonperturbatively and background-independently: areas and volumes acquire discrete spectra, and spacetime is built from spin networks. QGC shares LQG's refusal to treat the metric as a small fluctuation on a fixed background — the algebraic statement §8.2.2 is background-light in the same spirit — but differs in that QGC does not claim geometry is fundamentally discrete, or fundamentally quantum at all: geometry is the (approximately abelian) commutant of matter, classical in the regime where the correspondence operates. LQG builds geometry from quantum atoms; QGC reads geometry off the correlations of one state and asks only that it be classical where measured.

*Dynamical collapse models* (GRW, CSL, Diósi–Penrose) answer by modifying quantum mechanics: they add a stochastic nonlinear term to the Schrödinger equation that spontaneously localizes macroscopic superpositions, and Diósi–Penrose in particular ties the collapse rate to the gravitational self-energy $E_{\mathrm{G}}$ — the same scale QGC arrives at. This is the closest neighbour, and the sharpest distinction. QGC keeps quantum mechanics *exactly linear*: the decay in (8.5) is unitary decoherence into geometry, not a real collapse of the global state. The two therefore agree on the *timescale* $\hbar/E_{\mathrm{G}}$ but disagree on the *mechanism* and, crucially, on measurable second-order signatures — whether the global state stays pure (QGC: yes, coherence is relocated not destroyed) or is genuinely reduced (collapse: no) — and Chapter 12 turns that disagreement into an experimental discriminator (the ratio $F(2t)/F(t)$ of Chapter 2's Gaussian onset). QGC is, in this company, the option that changes the least: it modifies neither quantum mechanics nor the classical field equations, and locates all the new physics in their correlation.

**Q 8.3** ***"If QGC changes neither theory and adds no new field, what is actually new here? It sounds like relabeling."***

The new content is a *prediction*, and it is quantitative and falsifiable. Relabeling would produce no numbers; Principle 8.1 produces (8.5), and (8.5) produces a decoherence rate that no combination of "QM as usual" and "GR as usual" predicts, because that rate lives in the correlation both restrictions discard. Semiclassical gravity (average the source) predicts *no* such decoherence; perturbative quantized gravity predicts it at order $G^2$, suppressed by $\sim 10^{34}$ (Chapter 12); QGC predicts it at order $G^1$, at the 1.6 ns benchmark. Three different, mutually exclusive numbers for one laboratory observable — that is not relabeling, it is a fork the experiment of Chapter 15 will resolve. What is genuinely economical about QGC is that it buys this prediction *without* new fields or modified dynamics: the frame is cheap, and the prediction is its own.

## 8.5 The shape of Part III

Principle 8.1 is the *frame* of the theory; it is deliberately contentless about *which* state $\omega$ and *which* dynamics the world runs. The remaining chapters of Part III put content in the frame, in an order that mirrors the logical chain of §8.2.1.

Chapter 9 states the three primitive axioms — Generalized Entropy, Entanglement Equilibrium, Modular Time — that select $\omega$ and its clock, each with the countermodel that shows it is independent. These are the content the frame was waiting for. Chapter 10 shows the frame plus the axioms reproduces the geometry restriction correctly: the Einstein equations *derived* from entanglement equilibrium, so that $\omega|_{\mathcal{M}_{\text{geo}}}$ really is general relativity and not merely something classical-looking. Chapters 11–12 then execute the last step of the chain: the Wheeler–DeWitt constraint builds the branch geometries $|g_n\rangle$ and their overlaps rigorously (Chapter 11), and the rate itself — with its honest $G^1$-versus-$G^2$ ledger — follows (Chapter 12). Chapter 13 isolates the single open construction on which the operator-level identification $K = 2\pi H_{\text{phys}}$ still rests. By the end of Part III the promissory note of Chapter 1 is discharged to exactly the degree mathematics currently allows, with the residue named and handed to the experiments of Part IV.

> **Read Part III as one sentence unfolded.** Everything ahead is the chain of §8.2.1 filled in: *one state on one algebra* (this chapter) $\to$ *selected by three axioms* (Ch. 9) $\to$ *whose geometry restriction is Einstein's* (Ch. 10) $\to$ *whose matter restriction decoheres at the overlap* $\langle g_n|g_m\rangle$ (Ch. 11) $\to$ *giving the rate* $C\,E_{\text{G}}/\hbar$ (Ch. 12), *modulo one open step* (Ch. 13). If you keep the chain in view, no chapter is a detour.

## Chapter FAQ

**Q: Is "geometry is the commutant of matter" a definition or a physical claim?** A physical claim, and the central one. One is always free to form the commutant $(\mathcal{M}_{\text{mat}})'$ of any algebra — that is kinematics. The claim is that this commutant *carries spacetime geometry*: that its modular structure encodes the metric matter lives in (via Bisognano–Wichmann and the crossed product of Chapter 7), and does so exactly enough that $\omega|_{(\mathcal{M}_{\text{mat}})'}$ is general relativity in the semiclassical regime. That claim is [CONJECTURED] at the operator level and earned in expectation values; it is the content of Axiom III, and Chapter 13 names what remains open.

**Q: The duality map (8.4) used a tensor product, but Chapter 6 said local algebras don't factorize. Which is it?** Both, at different standings. The exact statement is algebraic (Principle 8.1: a subalgebra and its commutant, no tensor product needed). The tensor form (8.4) is the type I approximation to it, made legitimate by the crossed product of Chapter 7, which turns the type III geometry algebra into a type II algebra with a trace and restores enough factorization to define $\langle g_n|g_m\rangle$. We compute in the approximation and claim only where the semiclassical conditions of Chapter 11 hold.

**Q: Why call QM "more fundamental" than GR if neither restriction is the whole state?** Because the asymmetry is structural, not a matter of taste. The matter algebra and its state are defined on any Hilbert space with no reference to spacetime; the geometry restriction, by contrast, is built as the commutant of matter and is only approximately classical, only in a regime. Neither restriction is the whole $\omega$ — that much is symmetric — but the geometric side presupposes the quantum side in a way the reverse does not, which is exactly why we say "correspondence," not "duality" (§8.3).

**Q: If the global state (8.4) stays pure, in what sense has anything decohered?** In exactly the sense of Chapter 2: the *matter* state (8.5) is mixed, its coherence relocated into matter–geometry correlations that no matter-only measurement can recover. The global purity is why QGC is not a collapse model; the local mixedness is why the fringes die. Both are true at once, and (8.5) states them with no contradiction — it is the same "perfect global knowledge, zero local knowledge" of the Bell pair, with geometry playing the role of the traced-out partner.

## Exercises

**Exercise 8.1.** Take the two-branch matter superposition $|\psi\rangle = (|\psi_1\rangle + |\psi_2\rangle)/\sqrt{2}$ with orthogonal matter states, sourcing geometries $|g_1\rangle, |g_2\rangle$ with real overlap $\langle g_1|g_2\rangle = s \in [0,1]$. (a) Write the joint state (8.4) and confirm it is entangled iff $s < 1$. (b) Compute $\rho_{\mathrm{mat}}$ from (8.5), give its eigenvalues, and state the fringe visibility $\mathcal{V}$ an interferometer on matter alone would measure. (c) Using the Schmidt result of Chapter 2, write the entanglement entropy $S(\rho_{\mathrm{mat}})$ as a function of $s$ and check the two limits $s = 1$ (branches source the same geometry) and $s = 0$ (orthogonal geometries). Which limit is the seam of Chapter 1?

**Exercise 8.2.** Let $\mathcal{M} = M_2(\mathbb{C}) \otimes M_2(\mathbb{C})$ act on $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$, with $\mathcal{M}_{\mathrm{mat}} = M_2(\mathbb{C}) \otimes \mathbb{1}$ the matter subalgebra. (a) Identify the commutant $\mathcal{M}_{\mathrm{geo}} = (\mathcal{M}_{\mathrm{mat}})'$ using Definition 6.2. (b) Show that $\mathcal{M}_{\mathrm{geo}}$ is a factor (trivial center) and hence *not* abelian — so in this finite toy the geometry restriction is *not* classical. (c) Explain in one or two sentences what physical ingredient, absent from this toy, makes the real geometry algebra approximately abelian in the state $\omega$ (cf. (8.3) and the callout of §8.2.1). This exercise shows that "geometry is the commutant" is necessary but not sufficient for classicality: the *state* must do the rest.

## Solutions

*Solution (Exercise 8.1).* (a) $|\Psi\rangle = (|\psi_1\rangle \otimes |g_1\rangle + |\psi_2\rangle \otimes |g_2\rangle)/\sqrt{2}$. The Schmidt rank exceeds one — so the state is entangled — exactly when $|g_1\rangle, |g_2\rangle$ are linearly independent, i.e. $s < 1$; at $s = 1$ the geometry factors out and $|\Psi\rangle = [(|\psi_1\rangle + |\psi_2\rangle)/\sqrt{2}] \otimes |g_1\rangle$ is a product. (b) From (8.5), in the orthonormal matter basis $\{|\psi_1\rangle, |\psi_2\rangle\}$,

$$\rho_{\mathrm{mat}} = \tfrac{1}{2}\begin{pmatrix} 1 & s \\ s & 1 \end{pmatrix},$$

whose eigenvalues are $(1 \pm s)/2$ (Bloch vector of length $s$; Worked Example 2.1). The visibility is $\mathcal{V} = |\langle g_1|g_2\rangle| = s$ — the geometry overlap read directly as fringe contrast, which is the whole content of (8.6). (c) $S = -\frac{1+s}{2}\ln\frac{1+s}{2} - \frac{1-s}{2}\ln\frac{1-s}{2}$. At $s = 1$: $S = 0$, full fringes, geometry carries no which-path information — no seam. At $s = 0$: $S = \ln 2$, zero fringes, the geometries are orthogonal and have recorded complete which-path information — this is the seam, where the mass's two positions have become geometrically distinguishable and coherence is gone. Chapters 11–12 compute how $s = |\langle g_1|g_2\rangle|$ falls from 1 toward 0 in time, and the rate of that fall is $\tau_{\text{dec}}$.

*Solution (Exercise 8.2).* (a) By Definition 6.2, an operator commutes with every $A \otimes \mathbb{1}$ iff it acts trivially on the first factor: hence $\mathcal{M}_{\text{geo}} = (\mathcal{M}_{\text{mat}})' = \mathbb{1} \otimes M_2(\mathbb{C})$, the matrix algebra on the second qubit. (b) Its center is $\mathcal{M}_{\text{geo}} \cap (\mathcal{M}_{\text{geo}})' = \mathcal{M}_{\text{geo}} \cap (\mathbb{1} \otimes \mathbb{1}\text{-scalars}) = \mathbb{C}\mathbb{1}$, so $\mathcal{M}_{\text{geo}}$ is a factor; and $M_2(\mathbb{C})$ is non-abelian ($[\sigma_x, \sigma_z] \neq 0$), so the geometry restriction here is a full quantum system, not classical. (c) What is absent is *decoherence driven by a state*. In the real theory the geometric observables are macroscopically many and continuously monitored by the correlations of $\omega$ (equivalently, the branch geometries $|g_n\rangle$ are near-orthogonal coherent states of a field with astronomically many modes), which suppresses $\omega([B_1, B_2])$ for accessible $B_1, B_2$ even though the algebra itself is non-abelian. Classicality is a property of the *state on* the algebra, not of the algebra alone — the same lesson as "entropy is relational" in Chapter 2.

**What to carry forward**

1. **The claim.** One von Neumann algebra $\mathcal{M}$ carries one state $\omega$; quantum mechanics is $\omega|_{\mathcal{M}_{\text{mat}}}$ on a matter subalgebra, general relativity is $\omega|_{\mathcal{M}_{\text{geo}}}$ on its commutant $\mathcal{M}_{\text{geo}} = (\mathcal{M}_{\text{mat}})'$ (Eq. (8.1), Principle 8.1), the geometry restriction being approximately abelian and hence classical. This is a dictionary, not a metaphor — the cylinder-and-shadows of Chapter 1 is retired.
2. **The map.** Linearity plus branch-wise sourcing forces the duality map $\sum_n c_n |\psi_n\rangle \mapsto \sum_n c_n |\psi_n\rangle \otimes |g_n\rangle$ (Eq. (8.4)); tracing out geometry gives a reduced matter state (Eq. (8.5)) whose off-diagonals are the geometry overlaps $\langle g_n|g_m\rangle$ — the decoherence functional Chapters 11 and 12 compute, and the same visibility law as Chapter 2 with spacetime as the which-path environment.
3. **The honesty.** QGC does not quantize the metric, does not solve the measurement problem (Born rule [ASSUMED]), and replaces neither parent theory — both are recovered as the two restrictions. It is a *correspondence*, not a strict AdS/CFT-style duality: the geometry side is not independently complete. The frame is stated here; the three axioms (Chapter 9) fill it with content, and one identification, $K = 2\pi H_{\text{phys}}$, stays [CONJECTURED].

# 9 The Three Axioms

**Synopsis.** The whole theory rests on three sentences, and they are three clauses of a single sentence. Chapter 8 wrote down the correspondence between quantum states and geometries; here we ask what minimal set of assumptions forces it, and we find that the answer is one primitive — *the physical content of a region is the observer-dressed modular structure of its causal diamond* — stated three times, once for the algebra, once for the state, once for the clock. We state each clause in the exact language Parts II built (generalized entropy, the crossed product, modular flow at $\beta = 2\pi$), unpack every term, and mark its honest status on its face. Two of the three are theorems of rigorous mathematics; the third, the identification of modular time with physical time, carries the entire conjectural weight of the framework and is the one an experiment can kill. We prove the three independent by countermodel, and work the sharpest countermodel in full: electromagnetism, which shares the algebra and the state clauses but fails the time clause — and the reason it fails is exactly the reason gravity decoheres. We close with the ledger of what the theory assumes versus what it earns.

**You will be able to.** state each of the three v3.0 axioms precisely and say which piece of machinery from Parts II each one invokes; explain why modular time is the falsifiable axiom and what observation would refute it; construct a countermodel for each axiom; and reproduce the QED contrast that singles out the Hamiltonian constraint as what makes gravity special.

**Prerequisites.** Chapter 5 (generalized entropy $S_{\text{gen}} = A/4\ell_{\text{P}}^2 + S_{\text{ext}}$), Chapter 6 (von Neumann algebra types, the crossed product), Chapter 7 (modular flow, KMS at $\beta = 2\pi$, $K = 2\pi H$), Chapter 8 (the correspondence frame). Chapter 2 for relative entropy.

## 9.1 One primitive, three clauses

By the end of Chapter 8 we had a picture but not yet a foundation. The picture was that quantum mechanics and general relativity are two restrictions of one algebra carrying one state; the missing piece is the short list of assumptions from which that picture, and the decoherence rate on the cover of this book, actually follow. This chapter supplies the list. It has three items.

The temptation — which the framework resisted only after a hard lesson we tell in §9.7 — is to make the three items independent postulates of independent content. They are not. They are the same statement read three ways. Here is the statement.

**Principle 9.1** (Modular Diamond Principle)**.** *The physical content of a region of the world is the observer-dressed modular structure of its causal diamond. Concretely: for every small causal diamond $\mathcal{D}$ (a diamond small compared to the local curvature radius, $\ell_{\text{P}} \ll \ell \ll L_{\text{curv}}$),* *(algebra)* *the gravitationally constrained algebra of $\mathcal{D}$ is the crossed product of its field algebra by its observer's modular flow;*

*(state)* *the physical state of $\mathcal{D}$ is at entanglement equilibrium;*

*(time)* *the modular flow of $\mathcal{D}$ at $\beta_{\mathrm{mod}} = 2\pi$ is $\mathcal{D}$'s physical time.*

Everything in Part II fed exactly these three clauses. The crossed product of Chapter 6 is the algebra clause; the generalized entropy of Chapter 5 is its entropy. The entanglement first law and the Raychaudhuri equation of Chapter 3 are the machinery of the state clause. The modular flow at $\beta = 2\pi$ of Chapter 7, and the identity $K = 2\pi H$, are the time clause. We now write each clause as a numbered axiom, in the operational-before-formal order this book keeps: what is measured, then what the symbols mean, then the honest scope.

**Q 9.1 *"Three axioms for a whole theory of quantum gravity — isn't that suspiciously few? Every rival programme has more moving parts."***

Fewer premises is the goal, not a warning sign; it was Einstein's whole method in 1905, two postulates and no more. But be precise about the claim. The three axioms do not build spacetime from nothing at the Planck scale — the theory takes no position on that (Chapter 1, the wager). What they do is much narrower and much more testable: they specify, at the semiclassical seam where a heavy mass superposes, exactly how the one algebra restricts to "quantum mechanics" and to "geometry," and at what rate the two shadows lose their mutual coherence. Two of the three, moreover, are not really new postulates at all — they are theorems of mathematics (the crossed product) and of an existing programme (Jacobson's), repackaged as clauses so that the one genuinely new commitment stands out in isolation. That isolation is the point of the whole exercise, and §9.4.4 is where you see why.

## 9.2 Axiom I: Generalized Entropy — the algebra clause

### 9.2.1 Operational statement

Fix a region of the world bounded by a surface $\mathcal{X}$ that is either a causal horizon (the boundary beyond which an observer cannot see) or a quantum extremal surface (the surface that extremizes the entropy we are about to write). The measurable content of that region is a number, its *generalized entropy*, and the axiom says what the number is and how it behaves.

**Axiom 9.2** (Generalized Entropy)**.** *For a region bounded by a quantum extremal surface or causal horizon $\mathcal{X}$, with the gravitational constraints imposed, the physical observables form a Type II (crossed-product) algebra whose von Neumann entropy is the generalized entropy*

$$S_{\mathrm{gen}}(\mathcal{X}) \;=\; \frac{A(\mathcal{X})}{4\ell_{\mathrm{P}}^2} \;+\; S_{\mathrm{ext}}(\mathcal{X}) \;+\; \mathrm{const}, \tag{9.1}$$

*defined up to an additive constant, together with the three sub-clauses*

*(I.a)* Global unitarity*: for the total state on a complete Cauchy slice, $S_{\mathrm{gen}}$ is invariant under evolution — a statement about the full slice, not $d/dt$ of a subregion.*

*(I.b)* Generalized second law*: for open subregions, $\Delta S_{\mathrm{gen}} \geq 0$ along future-directed horizon cuts.*

*(I.c)* Surface selection*: $\mathcal{X}$ is the surface that extremizes the generalized entropy,* $\partial_\lambda S_{\text{gen}} = 0$ *— taken here as a postulate.*

The formula (9.1) is the object we built in Chapter 5, now promoted from observation to axiom. Read it term by term.

### 9.2.2 Unpacking the terms

*The area term* $A/4\ell_P^2$. This is Bekenstein–Hawking entropy: a horizon of area $A$ carries $A/4\ell_P^2$ nats, one quarter of its area in Planck units. Chapter 5 met it as the ceiling on how much information can hide behind a surface. Here it is the geometric half of the entropy ledger. The axiom does not derive this term — it *posits* it. That is a deliberate honesty, flagged below: the area law is the irreducible geometric input of the whole framework.

*The matter term* $S_{\text{ext}}$. This is the ordinary von Neumann entanglement entropy of Chapter 2, computed for the quantum fields outside (or across) $\mathcal{X}$ — the entropy of the reduced state an observer bounded by $\mathcal{X}$ actually holds. It is the informational half of the ledger.

*"Type II (crossed-product) algebra."* This is the load-bearing phrase, and it is why Axiom 9.2 is called the *algebra* clause and not merely the entropy clause. Chapter 6 showed that the algebra of quantum fields in a region is Type $\text{III}_1$: it has no trace, no density matrix, and its entanglement entropy is literally infinite (every scale contributes, as Chapter 2's reader question warned). You cannot write $-\text{Tr}\,\rho \ln \rho$ on it because there is no $\rho$ and no Tr. The fix, built in Chapter 6, was the crossed product: adjoin the observer's own clock — an operator conjugate to the modular flow — and the algebra changes type, from $\text{III}_1$ to II, acquiring exactly the trace that lets a finite entropy exist. The remarkable theorem (Chandrasekaran, Longo, Penington, Witten; Chandrasekaran, Penington, Witten) is that the finite Type II entropy so produced *is* the generalized entropy (9.1). So the axiom is not asserting two separate facts ("the algebra is Type II" and "its entropy is $A/4\ell_P^2 + S_{\text{ext}}$"); it is asserting one, because in the crossed product the second follows from the first. [PROVEN] at CFT-proxy standing — the crossed-product theorem is rigorous; its realization for the physical causal diamond is the open construction R1 of Chapter 13.

### 9.2.3 The sub-clauses

The three sub-clauses say how $S_{\text{gen}}$ behaves, and each one was met in Part II.

*(I.a) Global unitarity.* On a complete slice of the whole world, the state is pure and its entropy does not change — closed-system evolution is unitary (Chapter 2: $S(U\rho U^\dagger) = S(\rho)$). Note the careful wording: this is a statement about the entire slice, *not* about the time derivative of a subregion's entropy. That distinction is the whole content of §9.7's first lesson, so hold it.

*(I.b) Generalized second law.* For an open region — one that can exchange information with its surroundings — the generalized entropy cannot decrease along a future horizon: $\Delta S_{\text{gen}} \geq 0$. This is the second law of thermodynamics with the horizon area counted as entropy. It was proved by Wall, given the quantum null energy condition; we quote it, we do not reprove it.

*(I.c) Surface selection. Which* surface $\mathcal{X}$? The one that extremizes $S_{\text{gen}}$. This is the quantum extremal surface prescription (Engelhardt–Wall) that fixed the black-hole information paradox by producing the Page curve; Chapter 5 met it as the rule that determines where the entropy ledger balances. The axiom takes it as a postulate rather than deriving it, because its known justification (the gravitational path integral) is not prior to the dynamics we are trying to found.

### 9.2.4 Honest scope

Three caveats sit on the face of Axiom 9.2, and honesty requires stating them in the main text, not a footnote (Chapter 1's discipline).

> *First: the split is scheme-dependent.* The division of $S_{\text{gen}}$ into an "area piece" $A/4\ell_P^2$ and a "matter piece" $S_{\text{ext}}$ is a renormalization convention, not a physical fact. Only the *sum*, its *differences* between configurations, and its *monotonicity* are physical. No "two separate ledgers" reading is asserted — this is exactly the lesson of Chapter 2, where a region's entanglement entropy diverged and only relative entropy survived. The generalized entropy inherits the same fine print.
> *Second: the area law is input, not output.* That a horizon carries entropy $A/4\ell_P^2$ is the one geometric ingredient the theory assumes rather than earns. Everything downstream — Einstein's equations in Chapter 10, the rate in Chapter 12 — is derived *given* this term. [ASSUMED] for the area law specifically; the rest of Axiom 9.2 (the Type II structure, the sub-clauses) is DERIVED or PROVEN as marked.
> *Third: the dead conservation clause.* An earlier version of this axiom carried a fourth sub-clause, "$dS_{\text{gen}}/dt = 0$ for closed systems." It was withdrawn. Why is the subject of §9.7; the short version is that it was vacuous where true and false where interesting, and no result ever used it.

## 9.3 Axiom II: Entanglement Equilibrium — the state clause

### 9.3.1 Operational statement

Axiom 9.2 said what entropy a region has. Axiom II says what state it is in: the equilibrium state, the one that makes the entropy stationary. The operational content is a comparison. Take any small causal diamond; perturb it slightly at fixed volume; the generalized entropy does not change to first order.

**Axiom 9.3** (Entanglement Equilibrium)**.** *Every small causal diamond $\mathcal{D}$ ($\ell_P \ll \ell \ll L_{\text{curv}}$) is at entanglement equilibrium: the generalized entropy of Axiom 9.2 is stationary at fixed volume,*

$$\delta S_{\text{gen}}(\mathcal{D}) \;=\; 0. \tag{9.2}$$

One line of meaning: the diamond's state is not just any state but the one that extremizes the entropy it can hold — the maximum-entropy state consistent with its energy, exactly the equilibrium condition that Chapter 2's Klein inequality attached to a Gibbs state. Equation (9.2) is a variational principle, and like every variational principle it has two kinds of consequence, one for the geometry and one for the state.

### 9.3.2 What it yields

*Varying the geometry gives Einstein's equations.* This is the headline, and it is the whole business of Chapter 10, so we only state it here. Stationarity (9.2), combined with the first law of entanglement ($\delta S_{\text{ext}} = \delta\langle K\rangle$, from Chapter 7) and the Raychaudhuri equation (from Chapter 3), forces the semiclassical Einstein equations

$$G_{\mu\nu} = \frac{8\pi G}{c^4} \langle \hat{T}_{\mu\nu} \rangle \tag{9.3}$$

to hold at every point (Jacobson's theorem). The geometric and matter prefactors cancel exactly; no Einstein–Hilbert action is inserted by hand. Einstein's equations are not postulated in this framework — they are the equilibrium condition of Axiom 9.3 read as a statement about curvature.

*Varying the state gives KMS at $\beta = 2\pi$.* The same stationarity, read as a statement about the quantum state rather than the geometry, fixes the diamond's state to satisfy the KMS condition at modular inverse-temperature $\beta_{\mathrm{mod}} = 2\pi$ — the thermal (Gibbs) statics of Chapter 7. This is the input the time clause will need in §9.4: the equilibrium state is a thermal state at a definite modular temperature, and $2\pi$ is not adjustable.

### 9.3.3 Honest scope

"Einstein derived" means *given the area law.* The cancellation of prefactors in Jacobson's argument uses the coefficient $1/4\ell_{\mathrm{P}}^2$ from Axiom 9.2 as input. So (9.3) is a theorem of Axioms 9.2 and 9.3 *jointly*, not of the equilibrium principle alone. Chapter 10 is scrupulous about this; it is the exact point on which an earlier "derive everything from thermodynamics" proposal was found to overclaim, and the corrected statement is the one here. Jacobson's derivation also carries its own conditions (small diamonds, first-order variations, conformal-matter subtleties, the cosmological constant entering as an undetermined integration constant), all inherited here. [DERIVED] on the gravity-as-thermodynamics programme, given the area law.

Note also what Axiom 9.3 does *not* do: it does not evolve states in time. The dynamics of quantum matter is ordinary unitary quantum field theory on the background; what equilibrium fixes is which backgrounds and states are self-consistent, i.e. the coupled field equations. Time evolution is the next axiom's job, and keeping the two jobs separate is what lets the conjectural weight concentrate where it does.

**Q 9.2** ***"If Axiom II derives Einstein's equations, and Axiom I already contains Bekenstein–Hawking entropy which Wald's theorem gets from the Einstein–Hilbert action, aren't the two axioms circular — each smuggling in the other?"***

A fair worry, and the answer is a clean no, established by counting content. Wald's theorem gives $S = A/4G$ from the Einstein–Hilbert action, but *only for stationary black holes with bifurcate Killing horizons* — a measure-zero sliver of the surfaces Axiom 9.2 covers. It says nothing about the additive structure $S_{\mathrm{gen}} = A/4\ell_{\mathrm{P}}^2 + S_{\mathrm{ext}}$ (Wald gives the area piece alone), nothing about the generalized second law (a statement about the arrow of time, which no time-symmetric action can produce), nothing about the quantum extremal surface rule (a different variational problem on a different space), and nothing about the universal scope across cosmological, Rindler, and dynamical horizons. The sliver of Axiom 9.2's content that overlaps with what Axiom 9.3 plus Wald can reach is a special case; the bulk of it — the type-II algebra, the QES rule, the universal scope — is genuinely independent. The overlap is a consistency check, not a circularity. And the formal proof that no axiom follows from the other two is §9.5, by explicit countermodel.

## 9.4 Axiom III: Modular Time — the time clause

### 9.4.1 Operational statement

The first two axioms fixed the algebra and the equilibrium state; neither says how anything changes in time. The third axiom supplies time itself, and it is the axiom on which the whole theory stands or falls.

The operational setup is the one from Chapters 1 and 2: a mass $M$ in a superposition of two locations a distance $d$ apart; the measured quantity is the interference visibility of the two branches as a function of elapsed time. Axiom III predicts the rate at which that visibility decays. But it says something deeper first — it says what "elapsed time" *is* for the diamond's observer.

**Axiom 9.4** (Modular Time — the Modular–Physical Correspondence)**.** *For the diamond's observer, the modular flow of the constrained (Type II, observer-dressed) algebra at the KMS temperature* $\beta_{\text{mod}} = 2\pi$ is *physical time evolution. At the operator level,*

$$K \;=\; 2\pi\, H_{\text{phys}} \;+\; O(G^2). \tag{9.4}$$

Here $K$ is the modular Hamiltonian of Chapter 7 — the generator of the flow the algebra defines all by itself, from the state, with no clock put in by hand — and $H_{\text{phys}}$ is the physical Hamiltonian that generates laboratory time. The axiom asserts they are the same operator up to the factor $2\pi$ and a correction of second order in Newton's constant. Equation (9.4) is *the central equation of the book.*

### 9.4.2 Why the factor is $2\pi$ and not adjustable

The $2\pi$ is not a fit. Chapter 7 proved that a thermal state's modular flow satisfies the KMS condition with modular period $2\pi$ (Bisognano–Wichmann for the Rindler wedge gave $K = 2\pi H_{\text{boost}}$ exactly), and Chapter 4 met the same $2\pi$ as the Unruh temperature's denominator, $T_{\text{U}} = \hbar a/2\pi c k_{\text{B}}$. Axiom 9.3 then fixed the diamond's state to be KMS at precisely $\beta_{\text{mod}} = 2\pi$. So the $2\pi$ in (9.4) is the same $2\pi$ that has been following us since the Unruh calculation — the geometric factor of a thermal circle — and the axiom's only new content is the word *is*: that the modular flow the algebra generates is not merely *like* time but *is* the observer's physical time.

### 9.4.3 What it predicts

The consequence is gravitational decoherence at first order in $G$. When the mass superposes, the two branches source distinguishable geometries; the constraint of Chapter 11 entangles matter with field; and the identification (9.4) converts that entanglement into an irreversible decay of coherence at the rate

$$\Gamma_{\text{dec}} \;=\; C\,\frac{E_{\text{G}}}{\hbar}, \qquad E_{\text{G}} \;=\; \frac{GM^2}{d}, \qquad C \in \left[\tfrac{2}{\pi},\, 1\right]. \tag{9.5}$$

One line of meaning: the visibility of a superposed grain decays at a rate set by the gravitational self-energy $E_G$ that distinguishes its two branches, divided by $\hbar$, with a coefficient $C$ of order one. This is the $\tau_{\text{dec}} = \hbar d/GM^2$ of Chapter 1, now traced to its source. The coefficient window — natural value $C = 1$ (Markovian dephasing, the Diósi–Penrose rate), floor $C = 2/\pi$ (the Margolus–Levitin quantum speed limit) — is pinned by this axiom alone, and Chapter 12 does the pinning. The full derivation of the rate is the business of Chapters 11–12; here we are only stating what Axiom 9.4 commits to.

### 9.4.4 The falsifiable axiom — honest status, no silent upgrade

Now the crucial candor, because this is where a theory earns or loses trust.

> The operator equation (9.4) is [CONJECTURED]. It is [PROVEN] in expectation values and in solvable models — exactly the Gibbs-state and Rindler-wedge computations of Chapter 7, where $K = 2\pi H$ holds on the nose — but as a *full operator identity* for the interacting theory it is not proved. The gap has been reduced to one sharply-posed open construction, R1, which is the entire subject of Chapter 13: the residual $O(G^2)$ term and the interacting split constant are R1's two open values, both bounded in controlled models but not computed exactly.

We will not upgrade this label anywhere in the book. When the central equation appears in a derivation, it appears as a conjecture, and the results that lean on it inherit the conjectural status. That discipline is not modesty for its own sake — it is what makes the theory scientific, because it isolates one commitment that an experiment can refute.

And an experiment can. The framework *commits* to (9.5): a first-order-in-$G$ decoherence timescale, with the $M^2/d$ scaling and the $C$-window, for a preparable superposition. The alternative, from perturbative quantum field theory alone, is a second-order-in-$G$ rate, slower by the enormous factor $G$ carries — some $10^{34}$ at the benchmark (Chapter 12 computes the gap). If a levitated-nanosphere interferometer holds coherence out to the $G^2$ timescale rather than the $G^1$ one, Axiom 9.4 is false, and with it the distinctive content of the theory. There is no product-state escape hatch — the commitment is unconditional (Chapter 15). Few approaches to quantum gravity hand the experimentalist so clean a verdict.

**Q 9.3 *"You say $K = 2\pi H_{\text{phys}}$ is "proven in expectation values." If it holds whenever you take an average, in what sense is it not just true?"***

Because an operator equation is far stronger than the equation of its averages. Two operators can have identical expectation values in every state of a particular family and still differ — they can disagree on variances, on commutators, on the off-diagonal matrix elements that carry *decoherence* (recall Chapter 2: the whole phenomenon lives in off-diagonal elements). What is established is $\langle K\rangle = 2\pi\langle H_{\text{phys}}\rangle$ in coherent states, and the equation $K = 2\pi H$ in every model solvable enough to check both sides fully. What is open is whether the full operators coincide in the interacting theory, including the pieces that never show up in a first-moment average. The honest reading: the physics that *averages* see is settled; the physics that *fluctuations* see — which is exactly the decoherence rate — rests on the conjecture. R1 (Chapter 13) is the work of closing that last gap, and the experiment is what decides it if the mathematics does not first.

## 9.5 The three axioms are logically independent

Three clauses of one principle, yes — but genuinely three, not one wearing three hats. To show that, we exhibit for each axiom a world satisfying the other two and failing it. A model where a statement fails while its supposed premises hold is a proof that the statement does not follow from those premises; this is the standard countermodel argument, and we run it three times.

**Theorem 9.5** (Independence of the three axioms)**.** *No one of Axioms 9.2, 9.3, 9.4 follows from the other two.*

*Derivation.* We give a countermodel per axiom.

*Axiom 9.2 independent of the rest — a world with no area term.* Take modular time on the algebra of an ordinary, *non-gravitational* quantum field theory: the flat Rindler wedge of Chapter 7, where Bisognano–Wichmann gives $K = 2\pi H_{\text{boost}}$ exactly (the time clause holds), and the equilibrium state is the Minkowski vacuum (the state clause holds). But the algebra is Type III$_1$: there is no $A/4G$ term, no crossed product, no generalized entropy. Axiom 9.2 fails while the others hold — because without gravity there is no area to count.

*Axiom 9.3 independent — a non-Einstein background.* Take a spacetime whose region algebras do carry the Type II generalized entropy (Axiom 9.2 holds), and impose modular time (the time clause holds), but let the diamonds be sourced by out-of-equilibrium entanglement, so $\delta S_{\text{gen}} \neq 0$. The entropy functional exists; the field equations are simply not Einstein's. Axiom 9.3 fails while the others hold — the equilibrium condition is a genuine extra demand, not a consequence of having an entropy.

*Axiom 9.4 independent — electromagnetism.* This is the sharpest and most instructive countermodel, because electromagnetism satisfies the analogues of Axioms 9.2 and 9.3 and yet does *not* decohere superposed charges. It gets its own section (§9.6), because understanding *why* it fails is understanding what singles out gravity.

## 9.6 The QED countermodel, worked in full

Here is a puzzle that should worry you, and that worried the framework hard enough to become one of its strongest consistency checks. Everything the coherent-state picture of gravitational decoherence uses seems to apply, word for word, to electromagnetism.

### 9.6.1 The parallel that almost works

Put a charge $q$ (an electron, say) in a superposition of two locations a distance $d$ apart, $|\psi\rangle = (|L\rangle + |R\rangle)/\sqrt{2}$. Gauss's law, $\nabla\cdot\mathbf{E} = \rho/\varepsilon_0$, forbids the charge from existing in a bare superposition: the electric field must match the charge, so the physical state is dressed,

$$|\Psi^{\text{EM}}_{\text{phys}}\rangle = \frac{1}{\sqrt{2}}\Big(|L\rangle\,|E_L\rangle + |R\rangle\,|E_R\rangle\Big), \tag{9.6}$$

exactly the form of the gravitational state $(|L\rangle\,|\Phi_L\rangle + |R\rangle\,|\Phi_R\rangle)/\sqrt{2}$. There is even an energy scale: the Coulomb self-energy of the field difference,

$$E_C = \frac{q^2}{4\pi\varepsilon_0\, d}, \tag{9.7}$$

the perfect analogue of $E_G = GM^2/d$. So one is tempted to predict decoherence at $\Gamma \sim E_C/\hbar$. For an electron split by a micron that is a timescale of $\sim 5 \times 10^{-13}$ s — and it is flatly wrong. Electron interferometers hold coherence over nanoseconds; superconducting qubits over microseconds; trapped ions longer still. Charged particles in superposition do not decohere from their own Coulomb fields. If the gravitational mechanism predicted otherwise for electromagnetism, it would be dead on arrival, killed by seventy years of interferometry.

So it must be that the parallel breaks somewhere. Tracking exactly where is the payoff.

### 9.6.2 Where the two theories agree

The coherent-state argument that leads to $K = 2\pi H_{\text{phys}}$ runs through a chain of steps, and one can run the identical chain for photons. It works — step for step — almost all the way:

| **Step** | **Gravity** | **QED** |
|---|---|---|
| Modular flow is the boost (Bisognano–Wichmann) | $K_0 = 2\pi H_{\text{boost}}$ | identical |
| The sourced field is a coherent state | graviton, $O(G^{1/2})$ | photon, $O(\alpha^{1/2})$ |
| Modular Hamiltonian shifts covariantly under it | yes | yes |
| The shift is local, no bilocal terms | yes | yes (abelian: even cleaner) |

Every one of these is a generic property of a free relativistic field and its coherent states. Photons have them as surely as gravitons. If the mechanism depended only on these, electromagnetism *would* decohere, and we would have a contradiction with experiment. It does not. The chain has one more step, and that step is where the theories part.

### 9.6.3 The step that fails: constraint structure

The last step identifies modular flow with physical time. In gravity, this rests on the Wheeler–DeWitt constraint

$$\hat{H}_{\text{total}} |\Psi_{\text{phys}}\rangle = 0 \tag{9.8}$$

— a *Hamiltonian* constraint: the total energy of the closed system, matter plus geometry, vanishes on physical states. Its consequences are exactly what the time clause needs. Because $H_{\text{total}} = 0$, there is no external time; time must emerge relationally from correlations within the state (Page–Wootters, Chapter 11). And because the matter Hamiltonian equals minus the gravitational one, the modular Hamiltonian $K$ of the region is forced to track the physical energy: $K = 2\pi H_{\text{phys}}$. The constraint that eliminates external time is the very thing that makes modular time *be* physical time.

In electromagnetism the constraint is Gauss's law,

$$\nabla \cdot \hat{\mathbf{E}} |\Psi_{\text{phys}}\rangle = \frac{\hat{\rho}}{\varepsilon_0} |\Psi_{\text{phys}}\rangle , \tag{9.9}$$

and this is a *spatial* constraint: one equation per point of space, at each instant, constraining the field configuration, not the total energy. The QED Hamiltonian is nonzero and unconstrained; time is a background parameter given by Minkowski spacetime, not something that has to emerge. Gauss's law generates internal $U(1)$ gauge transformations at fixed time — it commutes with the Hamiltonian but is not proportional to the modular generator. So imposing it is *not* imposing modular invariance, and $[K^{\mathrm{EM}}, H_{\mathrm{phys}}] \neq 0$: modular flow and time evolution are genuinely different flows. The identification that gravity gets for free is simply unavailable.

> The one-line difference. *Gravity's constraint is the total Hamiltonian ($H_{\mathrm{total}} = 0$); electromagnetism's is a spatial divergence ($\nabla \cdot \mathbf{E} = \rho/\varepsilon_0$).* A Hamiltonian constraint ties matter and field together at the level of *dynamics* and makes time emerge, which is what forces $K = 2\pi H_{\mathrm{phys}}$. A spatial constraint ties them together only at the level of *kinematics* — the field configuration must match the charge at each instant — and leaves the dynamics an ordinary unitary evolution that preserves coherence. This single structural fact is what selects Axiom 9.4, and it is why the theory decoheres masses but not charges.

### 9.6.4 Kinematic versus dynamical entanglement

The physical picture behind the algebra: the entanglement in (9.6) is real — charge and field genuinely are entangled — but it is *kinematic*. It is present at every instant as a consequence of the constraint, it does not grow, and tracing over the (non-dynamical, Coulomb-gauge longitudinal) field leaves the charge's coherence intact; the Coulomb self-energy $E_C$ shows up only as a unitary phase $e^{-iE_C t/\hbar}$ on the off-diagonal element, which shifts the interference fringes but does not shrink them. The genuine decoherence of a charge — from real transverse-photon emission — is a *dynamical*, perturbative process, and it scales as $\alpha^2$: two coupling vertices, the electromagnetic counterpart of $G^2$. There is no $\alpha^1$ channel, exactly as experiment demands.

In gravity the Hamiltonian constraint *converts* kinematic entanglement into dynamical decoherence, because the constraint itself governs time evolution. Chapter 11 is where that conversion is computed; here the point is only that the constraint structure is what does it, and that electromagnetism's constraint cannot.

**For the curious — Why only gravity has a Hamiltonian constraint — a theorem, not an accident**

It is natural to suspect this is special pleading: perhaps some other force could be found or engineered with a Hamiltonian constraint. It cannot, and the reason is a theorem. Hojman, Kuchar and Teitelboim (1976) showed that any theory whose constraints reproduce the hypersurface-deformation algebra of spacetime — the algebra that includes a Hamiltonian constraint with structure *functions* rather than structure *constants* — is necessarily a metric theory of gravity. Yang–Mills theories, electromagnetism included, have structure constants and only spatial constraints. So no non-gravitational gauge theory can carry a Hamiltonian constraint without secretly being gravity. This lifts the QED countermodel from one worked example to a general statement: the $G^1$ mechanism of Axiom 9.4 cannot operate in any non-gravitational force. It also makes a prediction — in Kaluza–Klein theories, where a gauge force literally *is* extra-dimensional geometry, the higher-dimensional Hamiltonian constraint applies and the mechanism should return for that gauge sector. The worked contrast returns in Chapter 15, where the charge-versus-mass distinction becomes an experimental control.

## 9.7 How the axioms were hardened: the v2.0 teardown

The three axioms above are version 3.0. There was a version 2.0, and telling honestly what happened to it is not an embarrassment to bury but a case study in how an axiom set is supposed to be stress-tested. Each of the three earlier axioms was found broken under adversarial fire, and repaired; the repairs turned out to be three faces of the one principle (9.1). Candor about the breakage is a feature of the method (Chapter 1's discipline again; Einstein correcting his own record in a footnote).

### 9.7.1 The withdrawn conservation clause

Version 2.0's Axiom I carried a fourth sub-clause: $dS_{\text{gen}}/dt = 0$ for closed systems. It reads plausibly — entropy is conserved — but it is vacuous where true and false where interesting.

**Worked Example 9.1 — The trivial-or-false dilemma**

Two cases exhaust the meaning of "closed system." *Case one*: $\mathcal{X}$ bounds the whole world. Then the global state is pure, $S_{\text{ext}} = 0$ at all times, and $dS_{\text{gen}}/dt = 0$ merely restates unitarity — true, but adding nothing Axiom (I.a) did not already say. *Case two*: $\mathcal{X}$ is a dynamical interior surface, say the horizon of an evaporating black hole. Then the Page curve — the single most important fact about black-hole information — is precisely the statement that $A/4\ell_{\text{P}}^2$ and $S_{\text{ext}}$ do *not* cancel slice by slice: $dS_{\text{gen}}(\text{horizon})/dt \neq 0$ generically. The proposed clause is false here. The constancy that genuinely holds is a property of a different object (the island/QES prescription for the radiation region), obeying a different theorem.

So the clause was vacuous in case one, false in case two, and used by no derivation in either. It was withdrawn. Axiom 9.2 keeps only the two survivors — global unitarity on a full slice (I.a) and the generalized second law for open regions (I.b) — and loses zero predictions in the process. That last fact is the tell: a clause that can be deleted with no downstream damage was never load-bearing.

### 9.7.2 The demoted entropic action

Version 2.0's Axiom II was not the equilibrium principle but an "entropic action," a free-energy functional $\mathcal{S}[\rho, g] = \int \sqrt{-g}\,[\langle \hat{H} \rangle_\rho + c^4 R/16\pi G - \beta^{-1} s(\rho)]$ that was to be varied to yield the dynamics. It was demoted for a precise reason: it dressed *statics* as *dynamics*. Varying it with respect to $\rho$ gives a frozen Gibbs state, not an equation of motion — and worse, the theory's own flagship configuration, a *pure* superposition (entropy $s(\rho) = 0$, a rank-one state), is not even an allowed input, because $\ln \rho$ diverges on it. The action was never actually varied in any headline derivation: the decoherence timescale, the MOND acceleration $a_0 = cH_0/2\pi$, the dark-energy density — all were obtained without it. So it was parasitic: present, impressive-looking, and load-bearing for nothing. Version 3.0 replaces it with the entanglement-equilibrium principle the derivations *actually use* (Axiom 9.3), and keeps the action only as an equivalent coarse-grained repackaging of the equilibrium statics, in an appendix, varied for nothing.

### 9.7.3 The retired Planck interpolation

Version 2.0's Axiom III was the most ambitious and the most broken: a Planck-scale interpolation $\mathcal{O} = \rho_q\, f(r/\ell_P) + \rho_g\, [1 - f(r/\ell_P)]$ meant to blend the quantum and geometric descriptions smoothly as one approaches the Planck length. It is formally incoherent — trivial where it has no content, and inconsistent where it does.

**Worked Example 9.2 — Trivial-or-Bianchi-violating**

Wherever the semiclassical Einstein equation holds (which is wherever Axioms 9.2+9.3 hold), the geometric density equals the quantum one, $\rho_g = \rho_q$, so the interpolation is between a quantity and itself: $f$ drops out and no prediction can depend on it. Trivial. The only way to give it content is to assert the Einstein equation *fails* at $r \sim \ell_P$. Then $\mathcal{O}$ is a modified field equation whose effective source is not conserved,

$$\nabla^\mu T^{\text{eff}}_{\mu\nu} = \Big(\langle T_{\mu\nu}\rangle - \frac{c^4}{8\pi G} G_{\mu\nu}\Big)\nabla^\mu f(r/\ell_P) \;\neq\; 0, \tag{9.10}$$

because the geometric side is covariantly conserved by the contracted Bianchi identity ($\nabla^\mu G_{\mu\nu} = 0$, Chapter 3) and the matter side is conserved on its own, so their weighted blend is conserved only where $\nabla^\mu f = 0$. Both factors on the right are nonzero exactly at $r \sim \ell_P$, where the two descriptions differ and $f$ varies — a diffeomorphism inconsistency precisely at the one scale where the axiom claims to say something. Recognise the shape of this failure: it is the Bianchi clash of Chapter 1, the same identity that killed naive semiclassical gravity, returning to kill a naive UV patch.

The interpolation is retired. Its slot in version 3.0 is taken by the Modular Time postulate (Axiom 9.4) — which in version 2.0 had been an un-axiomatized "Saturation Principle," a candidate fourth axiom that quietly carried the theory's real falsifiable content. Promoting it to Axiom III put the theory's one genuine commitment where it belongs, in plain sight. What survives of the old interpolation is modest, honest phenomenology: a minimum length $\sim \ell_P$ and a modified dispersion relation, derived from the generalized uncertainty principle with explicit order-one uncertainty, not from the discredited blend. Chapter 15 handles that phenomenology; it is [MOTIVATED], not derived, and labelled so.

**Q 9.4 *"If all three original axioms broke, why should I trust the current three any more than the last three?"***

Because the failure modes were different in kind, and the survivors are of a different character. The v2.0 axioms broke because they *overclaimed*: a conservation law that wasn't, a dynamical action that was only statics, a UV completion that was inconsistent. The v3.0 replacements deliberately claim *less*. Two of them — the algebra clause and the state clause — are no longer bold postulates at all but theorems of established mathematics (the crossed product) and an established programme (Jacobson's), imported with their caveats attached. The entire remaining boldness is concentrated in one clause, the time clause, and that clause is *explicitly* flagged conjectural and tied to a specific experiment. You should trust the current three exactly as far as: two are as solid as their cited theorems, and the third is a conjecture with an execution date. That is a more defensible epistemic position than any version that spread its risk silently across all three — which is the whole reason the teardown was worth doing.

## 9.8 The assumed ledger

Honesty about a theory means a clean accounting of what it takes in versus what it gives out. Here is the ledger for the three axioms.

| Statement | Status | Where |
|---|---|---|
| Born rule $p_n = \|c_n\|^2$ | [ASSUMED] | Ch 2 |
| Standard QM linearity (no collapse term) | [ASSUMED] | Ch 2 |
| Area law $S = A/4\ell_P^2$ (geometric input) | [ASSUMED] | Ax 9.2 |
| Algebra is Type II; $S_{\rm gen}$ its entropy | [PROVEN] † | Ax 9.2, Ch 6 |
| Generalized second law | [PROVEN] | Ax 9.2 (Wall) |
| QES surface selection | [ASSUMED] | Ax 9.2 (I.c) |
| Einstein equations $G_{\mu\nu} = 8\pi G\langle\hat{T}_{\mu\nu}\rangle/c^4$ | [DERIVED] | Ax 9.2+9.3, Ch 10 |
| Bekenstein–Hawking, Unruh, holographic bound | [DERIVED] | Ax 9.2 |
| Central equation $K = 2\pi H_{\rm phys} + O(G^2)$ | [CONJECTURED] | Ax 9.4, Ch 13 |
| Gravitational decoherence $\Gamma_{\rm dec} = CE_G/\hbar$ | [CONJECTURED] | Ax 9.4, Ch 12 |
| Minimum length, GUP dispersion | [MOTIVATED] | §9.7, Ch 15 |

† PROVEN as a theorem; realized for the physical diamond only at CFT-proxy standing (the open construction R1, Chapter 13).

Read the ledger top to bottom and the shape of the theory is visible at a glance. It *assumes* the same things ordinary quantum mechanics and general relativity assume — the Born rule, linearity, the area law — and adds nothing speculative on the assumption side. It *derives*, as genuine theorems, Einstein's equations and the standard results of horizon thermodynamics. And it *conjectures* exactly one thing, the identification of modular with physical time, from which the one novel prediction — gravitational decoherence at order $G$ — follows. The conjectural burden of the entire foundation sits in one clause of one principle, and that clause is the one an experiment measures.

That is the honest trade the Modular Diamond Principle makes. It gives up the grander rhetoric an earlier telling reached for — a single conserved matter–geometry ledger, a slogan that "QM *equals* GR" — in exchange for a foundation two of whose three clauses are theorems and whose one open question has an address in the laboratory. Chapter 10 now cashes out the state clause (Einstein's equations in full), and Chapters 11–12 cash out the time clause (the rate, step by step, with the constraint doing the work). The conjecture at the centre is carried, unupgraded, all the way to Chapter 15, where the experiment that will settle it is laid out.

## Chapter FAQ

**Q: Are the "three axioms" really three, if they are clauses of one principle?** Both statements are true and neither is a dodge. Logically they are three: no one follows from the other two, proved by the countermodels of §9.5. Structurally they are one: the same underlying object — the observer-dressed modular structure of a causal diamond — specified for its algebra, its state, and its time. The unity is what makes the theory economical; the independence is what makes each clause carry real content. A crystal has one structure and three lattice directions; nobody objects that the directions are "not really three" because they belong to one crystal.

**Q: Why is the area law assumed rather than derived? Doesn't that leave a hole at the foundation?** It leaves a clearly-marked hole, which is better than a hidden one. Deriving the area law would mean deriving Bekenstein–Hawking entropy from something more primitive — a full theory of quantum gravity's microstates, which nobody has in a form this framework could import. An earlier in-house proposal tried to make the area law a theorem via a Clausius relation, and the June 2026 review found it overclaimed and deferred it. So the framework does the honest thing: it takes the one geometric fact it cannot yet earn as an explicit input, labels it [ASSUMED], and derives everything else given it. The hole is the same one every approach to quantum gravity has; the difference is that this one is not papered over.

**Q: The QED contrast turns on the Wheeler–DeWitt constraint $H_{\text{total}} = 0$. Is that constraint itself uncontroversial?** The constraint is standard canonical gravity — it is the quantum form of the statement that general relativity has no preferred time, which is the content of diffeomorphism invariance (Chapter 3). What is a research question is not the constraint but its *consequence*: whether it forces the full operator identity $K = 2\pi H_{\text{phys}}$ or only the expectation-value version. That is the R1 gap of Chapter 13, and it is exactly the piece flagged [CONJECTURED]. The constraint is solid; the strength of what it implies is the open question.

**Q: If the central equation is only a conjecture, why build a whole book on it?** Because a sharp, falsifiable conjecture with a derivation programme behind it is worth more than a vague framework immune to test. The conjecture is proven in every solvable model, reduced to one well-posed open construction, and pinned to a laboratory number with a five-sigma protocol (Chapter 15). If it is wrong, an experiment now within engineering reach will say so within the decade. That is how physics is supposed to work: state the boldest claim precisely, isolate it, and let nature rule. The book is built on the conjecture the way 1905 was built on the constancy of $c$ — as a postulate whose consequences are worth chasing to the point where they can be checked.

## Exercises

**Exercise 9.1.** An electron ($q = e = 1.602 \times 10^{-19}$ C) is split by $d = 1\,\mu$m. (a) Compute the Coulomb self-energy $E_C = q^2/(4\pi\varepsilon_0 d)$ using $\varepsilon_0 = 8.854 \times 10^{-12}$ F/m. (b) If (wrongly) one applied the gravitational-style rate $\Gamma = E_C/\hbar$, what decoherence timescale would follow? (c) Electron interferometers maintain coherence over $\sim 10^{-9}$ s. By what factor does the naive prediction fail, and which step of §9.6 tells you it *should* fail?

**Exercise 9.2.** Axiom 9.4 sets the modular temperature at $\beta_{\text{mod}} = 2\pi$. Trace this $2\pi$ back through the book: (a) where did Chapter 4 produce a $2\pi$ in the same physical role, and in what formula? (b) where did Chapter 7 produce it as the modular period of a thermal state? (c) In one sentence, state why the coefficient in $K = 2\pi H_{\text{phys}}$ is *not* a fitted parameter.

**Exercise 9.3.** For each of the three countermodels in the proof of Theorem 9.5, name the two axioms it satisfies and the one it violates, and give in one phrase the physical feature responsible for the violation. (This is a check that you can run the countermodel argument yourself, not recite it.)

## Solutions

*Solution (Exercise 9.1).* (a) $E_C = \dfrac{(1.602 \times 10^{-19})^2}{4\pi(8.854 \times 10^{-12})(10^{-6})} = \dfrac{2.566 \times 10^{-38}}{1.113 \times 10^{-16}} = 2.3 \times 10^{-22}$ J. (b) $\tau = \hbar/E_C = (1.0546 \times 10^{-34})/(2.3 \times 10^{-22}) = 4.6 \times 10^{-13}$ s $\approx 0.5$ ps. (c) The naive prediction of half a picosecond is wrong by a factor of order $10^{-9}/5 \times 10^{-13} \approx 2 \times 10^3$: interference survives thousands of times longer than $\Gamma = E_C/\hbar$ would allow. It fails at the constraint step (§9.6): Gauss's law is a *spatial* constraint, so the charge–field entanglement is kinematic, $E_C$ enters only as a unitary phase, and there is no $\alpha^1$ decoherence channel. Real QED decoherence (transverse-photon emission) is $\alpha^2$ and far slower still. The energy scale $E_C$ exists; the constraint that would turn it into a rate does not.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch9 QED contrast values)`

*Solution (Exercise 9.2).* (a) In the Unruh temperature $T_\mathrm{U} = \hbar a/2\pi c k_\mathrm{B}$ (Chapter 4): the $2\pi$ is the circumference of the Euclidean thermal circle an accelerated observer sees. (b) In Chapter 7, Bisognano–Wichmann gave the Rindler-wedge modular flow as the boost with $K = 2\pi H_\mathrm{boost}$, i.e. a thermal (KMS) state of modular period $2\pi$. (c) It is not fitted because it is the fixed geometric period of the modular/thermal circle — the same $2\pi$ in both prior appearances — and Axiom 9.3 pins the diamond's equilibrium state to exactly that period; there is no dial to turn.

*Solution (Exercise 9.3).* *Non-gravitational QFT (flat Rindler wedge):* satisfies Axioms 9.3 (Minkowski vacuum is equilibrium) and 9.4 (Bisognano–Wichmann gives $K = 2\pi H_\mathrm{boost}$ exactly); violates Axiom 9.2 — no gravity, no area term, algebra stays Type $\mathrm{III}_1$. Feature: absence of the gravitational area/entropy. *Non-Einstein background:* satisfies Axioms 9.2 (Type II entropy exists) and 9.4 (modular time imposed); violates Axiom 9.3 — $\delta S_\mathrm{gen} \neq 0$, diamonds out of equilibrium. Feature: out-of-equilibrium entanglement sourcing the geometry. *Electromagnetism:* satisfies the analogues of Axioms 9.2 and 9.3; violates Axiom 9.4 — $[K^\mathrm{EM}, H_\mathrm{phys}] \neq 0$, first-order rate is zero. Feature: only a spatial (Gauss) constraint, no Hamiltonian constraint, so modular flow is not physical time.

**What to carry forward**

1. The theory rests on *one* primitive — the observer-dressed modular structure of a causal diamond (the Modular Diamond Principle) — read as three independent-but-inseparable axioms: *generalized entropy* (Ax 9.2, the algebra: $S_\mathrm{gen} = A/4\ell_\mathrm{P}^2 + S_\mathrm{ext}$ on a Type II algebra), *entanglement equilibrium* (Ax 9.3, the state: $\delta S_\mathrm{gen} = 0$, which with the area law yields Einstein's equations), and *modular time* (Ax 9.4, the clock: $K = 2\pi H_\mathrm{phys} + O(G^2)$).
2. Modular time is the falsifiable axiom. It is [CONJECTURED] as a full operator equation ([PROVEN] only in expectation values and solvable models; reduced to the open construction R1), and it commits the theory to first-order-in-$G$ decoherence $\Gamma_\mathrm{dec} = C E_\mathrm{G}/\hbar$, $C \in [2/\pi, 1]$. A $G^2$ timescale would falsify it. The label is never silently

upgraded.

3. The QED countermodel is load-bearing: electromagnetism shares the algebra and state clauses and an energy scale $E_C = q^2/4\pi\varepsilon_0 d$, yet does not decohere charges, because Gauss's law is a *spatial* constraint while gravity's Wheeler–DeWitt constraint is a *Hamiltonian* one — and only a Hamiltonian constraint makes modular flow *be* physical time. That single fact singles out gravity (Hojman–Kuchar–Teitelboim: no non-gravitational gauge theory can have it). The v2.0 teardown — a vacuous conservation clause, a statics-mislabeled action, an inconsistent Planck interpolation, all repaired — is the case study in earning that clarity.

# 10 From Axioms to Einstein's Equations

**Synopsis.** This is the theory's first great payoff. Two of the three axioms of Chapter 9 — generalized entropy (Axiom I) and entanglement equilibrium (Axiom II) — turn out to *contain* all of general relativity. We give the derivation twice. First, Jacobson's 1995 route: demand that a local Rindler horizon through every point obey the thermodynamic Clausius relation $\delta Q = T\,\delta S$, with $T$ the Unruh temperature and $S$ the horizon area law, feed in the Raychaudhuri equation for how the area responds to curvature, and Einstein's equations fall out — as an *equation of state*, not a fundamental field law. Second, the entanglement-equilibrium route (Jacobson 2015, the one Axiom II actually uses): demand that the generalized entropy of every small causal diamond be stationary, $\delta S_{\text{gen}} = 0$, and the same equations appear, read through the axioms. We are scrupulous about the one thing that is borrowed rather than earned — the area law $A/4\ell_{\text{P}}^2$ — and honest that "derived" here means *given that*. The chapter closes by cashing the result: because Einstein's equations follow, so does everything built on them, and the correspondence does not compete with general relativity but contains it.

**You will be able to.** set up a local Rindler horizon at any point and in any null direction; run the Clausius relation and the Raychaudhuri equation to obtain $R_{\mu\nu}k^\mu k^\nu \propto T_{\mu\nu}k^\mu k^\nu$; fix the trace term by the Bianchi identity to recover $G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G\,T_{\mu\nu}/c^4$ with $\Lambda$ an integration constant; and reproduce the same equations from $\delta S_{\text{gen}} = 0$ over a causal diamond, stating cleanly what is proved and what is assumed.

**Prerequisites.** Chapter 3 (Raychaudhuri equation (3.39), Rindler horizons, the Bianchi identity (3.16), metric and Christoffels); Chapter 4 (Unruh temperature (4.27)); Chapter 5 (the area law (5.12), generalized entropy (5.17), the Clausius/first-law reasoning); Chapter 7 (modular flow, KMS at $\beta = 2\pi$, (7.13)); Chapter 9 (Axioms I, II, III).

## 10.1 The claim, and why it should unsettle you

Chapter 1 presented Einstein's equations

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}\,T_{\mu\nu} \tag{10.1}$$

as one of the two flawless rulebooks — a fundamental law of nature, on the same footing as Maxwell's equations, to be revered and, one day, quantized. This chapter unsettles that reading. We will show that (10.1) is not fundamental at all, but a *thermodynamic equation of state*: a relation that emerges when you ask that a certain entropy be extremal, in exactly the way the ideal gas law $PV = Nk_{\text{B}}T$ emerges from the statistical mechanics of molecules. Nobody thinks $PV = Nk_{\text{B}}T$ is a fundamental law of the universe; it is what the molecular bookkeeping looks like from far away. The claim of this chapter — Jacobson's claim, adopted as a theorem of the axioms — is that Einstein's equations have the same humble status.

The operational content we must define before writing symbols is this: at *every* point $p$ of spacetime, and in *every* null direction $k^\mu$, we place a small horizon and impose one thermodynamic law across it. There is nothing special about $p$ or $k^\mu$; the construction is run everywhere and in all directions at once, and the requirement that it be mutually consistent is what pins the geometry. Let us build the horizon first, then the law, then watch the geometry fall out.

**Q 10.1** ***"If Einstein's equations are "merely" thermodynamics, does that make general relativity less true?"***

Not in the slightest — it makes it *more* understood. Thermodynamics is not false because statistical mechanics underlies it; the second law is as ironclad as any statement in physics. What changes is the interpretation. A fundamental field equation demands to be quantized directly — and that is the program that failed by power counting in Chapter 1. An equation of state does not: you do not quantize $PV = Nk_\mathrm{B}T$, you quantize the molecules. The lesson Jacobson draws, and the one this book is built on, is that looking for "the quantum theory of the metric" may be like looking for the quantum theory of temperature — a category error. The metric is a collective, thermodynamic variable, and the quantum degrees of freedom live underneath it.

## 10.2 Route one: the Einstein equation of state (Jacobson 1995)

### 10.2.1 A local Rindler horizon at every point

Fix a spacetime point $p$. Chapter 3 taught us that a uniformly accelerated observer carries a *Rindler horizon*: a null surface beyond which she can neither send nor receive signals, appearing at proper distance $c^2/a$ behind her. That was a global construction in flat space. Jacobson's first move is to localise it.

Choose any null direction $k^\mu$ at $p$. Because spacetime is a Lorentzian manifold, we may always find a local inertial frame at $p$ (Chapter 3: the equivalence principle, $g_{\mu\nu}(p) = \eta_{\mu\nu}$, $\Gamma(p) = 0$), and in that frame the past-directed null geodesics leaving $p$ along $-k^\mu$ generate a patch of a null surface — the boundary of the region an observer accelerating along $k^\mu$ through $p$ can causally influence. This patch is a *local Rindler horizon* $\mathcal{H}$ through $p$. It exists for every $p$ and every $k^\mu$, and near $p$ it looks exactly like the flat-space Rindler horizon of Chapter 3, because to first order in the distance from $p$ every spacetime is flat.

Let $\lambda$ be an affine parameter along the null generators of $\mathcal{H}$, running toward the future, with $\lambda = 0$ at the point $p$ (more precisely, at a spacelike cross-section through $p$ called the bifurcation surface, where the generators momentarily neither converge nor diverge). The local boost that generates $\mathcal{H}$ — the approximate Killing vector of the near-$p$ Rindler geometry — is

$$\chi^\mu \;=\; -\,\kappa\,\lambda\,k^\mu, \tag{10.2}$$

where $\kappa$ is the surface gravity (the acceleration of the Rindler observer, in the sense of Chapter 3). The factor $-\kappa\lambda$ is just the statement that the boost Killing vector vanishes at the bifurcation surface ($\lambda = 0$) and grows linearly along the generators; we met exactly this linear-in-$\lambda$ boost structure in the modular flow of Chapter 7. Equation (10.2) is the whole kinematic input; everything else is thermodynamics and the Raychaudhuri equation.

**For the curious — Why a horizon can be built at every point**

The equivalence principle guarantees a locally inertial frame at any $p$; a Lorentzian metric then guarantees a light cone at $p$; and picking one null generator of that cone plus an acceleration direction fixes a Rindler wedge whose edge is $\mathcal{H}$. No global symmetry is needed — the Killing vector (10.2) is only *approximate*, exact at $p$ and degrading away from it, which is why the argument controls only the leading behaviour near $p$ and delivers a *local* (pointwise) equation. That locality is a feature: an equation imposed at every point and in every null direction is enormously overdetermined, and it is precisely this overdetermination that will force the full tensor equation (10.1).

### 10.2.2 The Clausius relation across the horizon

Now the physics. We treat $\mathcal{H}$ as a thermodynamic system and impose the Clausius relation — the first law in the form heat divided by temperature equals entropy change,

$$\delta Q \;=\; T\,\delta S, \tag{10.3}$$

the same relation Chapter 5 used to force the black-hole area law out of the Hawking temperature. Three ingredients enter, each imported from an earlier chapter.

*The temperature* is the Unruh temperature of the accelerated observer for whom $\mathcal{H}$ is a horizon (Chapter 4, (4.27)):

$$T \;=\; \frac{\hbar\,\kappa}{2\pi\,c\,k_{\mathrm{B}}}. \tag{10.4}$$

The observer sees the vacuum as thermal at this temperature; the horizon is hot because acceleration is hot. This is where the $2\pi$ that runs through the whole book first enters the derivation.

*The entropy* is the area law of Chapter 5, (5.12): the horizon carries entropy proportional to its area,

$$S \;=\; \eta\,A, \qquad \eta \;=\; \frac{k_{\mathrm{B}}}{4\ell_{\mathrm{P}}^2} \;=\; \frac{k_{\mathrm{B}}c^3}{4G\hbar}, \tag{10.5}$$

with $\eta$ the entropy per unit area. We write $\eta$ as an unknown constant for the moment and let the derivation *demand* its value; the punchline will be that consistency forces exactly the Bekenstein–Hawking 1/4. This is the one piece we import rather than derive — flagged now, audited fully in §10.4.

*The heat* is the energy of matter that flows across $\mathcal{H}$. For a stress–energy tensor $T_{\mu\nu}$, the heat flux through the horizon is the flux of boost energy, integrated over the horizon:

$$\delta Q \;=\; \int_{\mathcal{H}} T_{\mu\nu}\,\chi^{\mu}\,d\Sigma^{\nu} \;=\; -\,\kappa \int_{\mathcal{H}} T_{\mu\nu}\,k^{\mu}k^{\nu}\,\lambda\,d\lambda\,dA, \tag{10.6}$$

where in the second equality we used the Killing vector (10.2) and wrote the horizon area element as $d\Sigma^{\nu} = k^{\nu}\,d\lambda\,dA$ (the generators point along $k^{\nu}$; $dA$ is the transverse cross-sectional area). The one line of meaning: the heat that crosses the horizon is the boost energy $T_{\mu\nu}k^{\mu}k^{\nu}$ of the matter, weighted by how far up the horizon ($\lambda$) it crosses, because the Killing energy grows with $\lambda$.

### 10.2.3 Area change from the Raychaudhuri equation

The right-hand side of the Clausius relation needs $\delta S = \eta\,\delta A$, so we need the area change $\delta A$ of the horizon. This is pure geometry, and Chapter 3 already did the hard part: the expansion $\theta$ of the null generators — the fractional rate of change of their transverse area, $\theta = d(\ln \delta A)/d\lambda$ — obeys the Raychaudhuri equation (3.39). For a null congruence, dropping the quadratic self-focusing $-\theta^2/(d-2)$ and the shear $-\sigma^2$ (both second order in the small quantities near the bifurcation surface, where $\theta$ and $\sigma$ start at zero) and the twist (zero for a hypersurface-orthogonal congruence like a horizon), it reads

$$\frac{d\theta}{d\lambda} \;=\; -\,R_{\mu\nu}\,k^\mu k^\nu \qquad \text{(near the bifurcation surface).} \tag{10.7}$$

Integrate once, using $\theta(0) = 0$ at the bifurcation surface:

$$\theta(\lambda) \;=\; -\,R_{\mu\nu}\,k^\mu k^\nu\,\lambda. \tag{10.8}$$

The expansion is negative when $R_{\mu\nu}k^\mu k^\nu > 0$: matter focuses the horizon generators, exactly the "gravity attracts" content of Chapter 3. Now the area change. Since $\theta = d(\ln \delta A)/d\lambda$, the fractional area change is $d(\delta A) = \theta\,\delta A\,d\lambda$, and integrating the leading contribution (with $\delta A \approx dA$ constant to this order) from the bifurcation surface outward gives

$$\delta A \;=\; \int \theta\,d\lambda\,dA \;=\; -\int R_{\mu\nu}k^\mu k^\nu\,\lambda\,d\lambda\,dA. \tag{10.9}$$

One line of meaning: the horizon's area shrinks in exact proportion to the Ricci curvature along $k^\mu$ — curvature is the geometric shadow of the heat that (10.6) is pouring across.

**Worked Example 10.1 — Matching the two sides**

Put the Clausius relation (10.3) together. The left side is the heat (10.6):

$$\delta Q \;=\; -\,\kappa \int T_{\mu\nu}k^\mu k^\nu\,\lambda\,d\lambda\,dA.$$

The right side is $T\,\delta S = T\,\eta\,\delta A$, with $T$ from (10.4) and $\delta A$ from (10.9):

$$T\,\eta\,\delta A \;=\; \frac{\hbar\kappa}{2\pi c\,k_\mathrm{B}}\,\eta\,\left(-\int R_{\mu\nu}k^\mu k^\nu\,\lambda\,d\lambda\,dA\right).$$

Setting $\delta Q = T\,\delta S$, the factor $\kappa$ cancels on both sides (a happy accident that removes the surface gravity, which was never a property of the geometry but of our arbitrary choice of accelerated observer), and so do the identical integrals $\int(\cdots)\lambda\,d\lambda\,dA$. What remains is a pointwise relation between the integrands, holding for the given $k^\mu$:

$$T_{\mu\nu}\,k^\mu k^\nu \;=\; \frac{\hbar\,\eta}{2\pi c\,k_\mathrm{B}}\,R_{\mu\nu}\,k^\mu k^\nu. \tag{10.10}$$

The two sides carry the same null vector $k^\mu$ contracted twice. And this holds for *every* null $k^\mu$ at *every* $p$ — the overdetermination promised in §10.2.1. That is enough to fix the full tensors, up to a term that vanishes on null vectors, as we show next.

### 10.2.4 From null contractions to the full tensor equation

Equation (10.10) says that the two symmetric tensors $T_{\mu\nu}$ and $\dfrac{\hbar\eta}{2\pi c\, k_{\rm B}} R_{\mu\nu}$ have equal contractions $t_{\mu\nu}k^\mu k^\nu$ for *all* null $k^\mu$. Two symmetric tensors that agree on all null vectors can differ only by a multiple of the metric: if $t_{\mu\nu}k^\mu k^\nu = 0$ for all null $k$, then, because a null vector satisfies $g_{\mu\nu}k^\mu k^\nu = 0$, the tensor $t_{\mu\nu}$ must be proportional to $g_{\mu\nu}$, say $t_{\mu\nu} = f\, g_{\mu\nu}$ for some scalar field $f$ (this is a standard fact of linear algebra — the null cone is large enough to detect any tensor *except* a pure trace piece, which annihilates null vectors; you prove it in Exercise 10.2). Applying it to the difference of the two sides of (10.10),

$$T_{\mu\nu} \;=\; \frac{\hbar\,\eta}{2\pi c\, k_{\rm B}}\, R_{\mu\nu} \;+\; f\, g_{\mu\nu}, \tag{10.11}$$

with $f(x)$ an as-yet-undetermined scalar. We are one constraint away from Einstein's equations. The unknown $f$ will be fixed — not fitted — by a consistency requirement, in the spirit of how Chapter 3 pinned the factor $\varphi(v)$ by symmetry: let the identities the objects already obey do the work.

### 10.2.5 Fixing $f$ by the Bianchi identity

Two conservation laws are available, and together they determine $f$ completely.

*First*, the stress–energy tensor is locally conserved — this is the statement that energy and momentum are not created or destroyed, a property of matter independent of gravity:

$$\nabla^\mu T_{\mu\nu} \;=\; 0. \tag{10.12}$$

*Second*, the Einstein tensor $G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2}R\, g_{\mu\nu}$ obeys, as a pure geometric *identity* true for every metric whatsoever, the contracted Bianchi identity of Chapter 3 (Theorem 3.3, (3.16)):

$$\nabla^\mu G_{\mu\nu} \;=\; 0 \qquad\Longleftrightarrow\qquad \nabla^\mu R_{\mu\nu} \;=\; \tfrac{1}{2}\,\nabla_\nu R. \tag{10.13}$$

Now take the divergence of (10.11). Write $\alpha \equiv \hbar\eta/(2\pi c\, k_{\rm B})$ for the constant coefficient. Using (10.12) on the left and $\nabla^\mu g_{\mu\nu} = \nabla_\nu$ (the metric is covariantly constant, Chapter 3) on the right:

$$0 \;=\; \alpha\,\nabla^\mu R_{\mu\nu} \;+\; \nabla_\nu f \;=\; \alpha\,\tfrac{1}{2}\,\nabla_\nu R \;+\; \nabla_\nu f, \tag{10.14}$$

where the second equality used the Bianchi identity (10.13). This is a gradient equation: $\nabla_\nu\big(\tfrac{\alpha}{2}R + f\big) = 0$, so the quantity in parentheses is a spacetime *constant*. Name it:

$$f \;=\; -\frac{\alpha}{2}\,R \;+\; \alpha\,\Lambda, \tag{10.15}$$

where $\Lambda$ is a constant of integration — forced to be constant, but not otherwise determined by the derivation. Substitute (10.15) back into (10.11):

$$T_{\mu\nu} \;=\; \alpha\Big(R_{\mu\nu} - \tfrac{1}{2}R\, g_{\mu\nu}\Big) \;+\; \alpha\,\Lambda\, g_{\mu\nu} \;=\; \alpha\big(G_{\mu\nu} + \Lambda\, g_{\mu\nu}\big). \tag{10.16}$$

The Einstein tensor has assembled itself: the trace term supplied by the Bianchi identity is exactly the $-\frac{1}{2}R\,g_{\mu\nu}$ needed to complete $R_{\mu\nu}$ into $G_{\mu\nu}$. Rearranged,

$$G_{\mu\nu} + \Lambda\, g_{\mu\nu} \;=\; \frac{1}{\alpha}\, T_{\mu\nu} \;=\; \frac{2\pi c\, k_{\mathrm{B}}}{\hbar\,\eta}\, T_{\mu\nu}. \tag{10.17}$$

### 10.2.6 The coefficient, and the punchline

Everything is derived except the number $\eta$, the entropy per unit area, which we left open in (10.5). Compare (10.17) with the target (10.1): they agree if and only if

$$\frac{2\pi c\, k_{\mathrm{B}}}{\hbar\,\eta} \;=\; \frac{8\pi G}{c^4} \qquad \Longleftrightarrow \qquad \eta \;=\; \frac{c^3 k_{\mathrm{B}}}{4G\hbar} \;=\; \frac{k_{\mathrm{B}}}{4\ell_{\mathrm{P}}^2}. \tag{10.18}$$

The consistency of horizon thermodynamics with the Raychaudhuri geometry *demands* the Bekenstein–Hawking entropy density, $1/4$ and all. Feed that $\eta$ back in and the derivation closes on the equation we set out to reach:

$$\boxed{G_{\mu\nu} + \Lambda\, g_{\mu\nu} \;=\; \frac{8\pi G}{c^4}\, T_{\mu\nu}} \tag{10.19}$$

Einstein's field equations, with the cosmological constant $\Lambda$ appearing as a constant of integration, obtained from three ingredients we built in earlier chapters — the Unruh temperature, the area law, the Raychaudhuri equation — plus one identity, the Bianchi identity, and one demand, the Clausius relation at every point.

**Theorem 10.1** (Einstein equations from horizon thermodynamics)**.** *Suppose that across the local Rindler horizon through every spacetime point, and for every null direction, the Clausius relation $\delta Q = T\,\delta S$ holds with $T$ the Unruh temperature* (10.4) *and $S$ the area law* (10.5)*. Then the geometry obeys the Einstein field equations* (10.19)*, with the cosmological constant $\Lambda$ an undetermined constant of integration, and the entropy density is forced to the Bekenstein–Hawking value $\eta = k_{\mathrm{B}}/(4\ell_{\mathrm{P}}^2)$.* [DERIVED] *— given the area law* (10.5) *as input (itself* [ASSUMED]*; see §10.4). The chain Clausius → Raychaudhuri → Bianchi is exact; the borrowed piece is the $1/4$.*

**Q 10.2** ***"You imposed the Clausius relation*** **at every point** ***in*** **every** ***null direction. Isn't that an enormous number of equations to demand at once — how can they be mutually consistent?"***

They can be consistent for exactly one reason, and it is the beautiful part. A priori, imposing $T_{\mu\nu}k^\mu k^\nu \propto R_{\mu\nu}k^\mu k^\nu$ for all $k^\mu$ at all points is wildly overdetermined — far more conditions than a symmetric tensor field has components. It could easily have had *no* solution, in which case horizon thermodynamics would have been incompatible with Lorentzian geometry and the whole idea would have died. Instead it has exactly the solutions permitted by the Bianchi identity: the geometry is free to be *any* solution of (10.19), and no more. That the count works — that the overdetermined thermodynamic demand is satisfiable at all, and precisely by Einstein's equations — is the content of the theorem, and it is why Jacobson could call (10.19) an "equation of state." The demand does not fit the geometry; it *selects* the geometries nature already allows.

## 10.3 Route two: entanglement equilibrium (Jacobson 2015)

Route one is complete and beautiful, but it leans on the Clausius relation, which strictly holds only in thermal equilibrium, and on treating the horizon as a hot membrane. The 2015 rederivation — the one Axiom II of Chapter 9 is built from — replaces "thermal equilibrium" with the sharper and more fundamental "entanglement equilibrium," and replaces the hot membrane with a small causal diamond. The physics is the same; the reading through the axioms is cleaner, and the tools are the rigorous ones of Chapters 2 and 7.

### 10.3.1 The setup: a small causal diamond

Take a spacetime point $p$ and a small ball $B_\ell$ of proper radius $\ell$ around it, with $\ell_\mathrm{P} \ll \ell \ll L$ where $L$ is any curvature scale. The *causal diamond* $\mathcal{D}$ is everything that can both be reached from and reach $B_\ell$ — the intersection of the future of the bottom tip and the past of the top tip. Its boundary is a null surface, its "horizon," with transverse area $A$. Unlike the Rindler horizon of route one, the diamond is defined purely by causal structure: no accelerated observer, no choice of frame. This is why it is the natural arena for a *local* statement of the axioms.

Two axioms now speak. Axiom I (generalized entropy, Chapter 9) assigns the diamond the generalized entropy of Chapter 5, (5.17):

$$S_\mathrm{gen} \;=\; \frac{A}{4\ell_\mathrm{P}^2} \;+\; S_\mathrm{ext}, \tag{10.20}$$

the horizon area in Planck units plus the matter entanglement entropy of the fields in the diamond. Axiom II (entanglement equilibrium) demands that, for the vacuum, this generalized entropy is *stationary* under any small first-order perturbation of the state and geometry at fixed diamond volume:

$$\delta S_\mathrm{gen} \;=\; 0. \tag{10.21}$$

The physical reading: the vacuum is the state of maximal generalized entropy for the diamond — "gravitational equilibrium" at every point — and a geometry is admissible only if perturbing it costs no generalized entropy to first order. We now compute the two pieces of $\delta S_\mathrm{gen}$ and set their sum to zero.

### 10.3.2 The area piece: geometry via Raychaudhuri

How does the diamond's boundary area respond to curvature? The same Raychaudhuri equation as route one, integrated over the diamond's null boundary rather than a Rindler horizon, gives the deficit of the area relative to its flat-space value. Carrying out the integral for a small diamond in $d = 4$ (the transverse geometry is a 2-sphere; this is the causal-diamond version of (10.9), done in full in the technical note cited in §10.4):

$$\delta A \;=\; -\,\frac{\pi\,\ell^5}{3}\,R_{\mu\nu}\,k^\mu k^\nu \;+\; O(\ell^6). \tag{10.22}$$

Curvature makes the diamond smaller than it would be in flat space, in exact proportion to $R_{\mu\nu}k^\mu k^\nu$ — the identical geometric content as route one, now attached to a diamond. Hence the area piece of the generalized-entropy variation is

$$\frac{\delta A}{4\ell_{\rm P}^2} \;=\; -\,\frac{\pi\,\ell^5}{12\,\ell_{\rm P}^2}\,R_{\mu\nu}\,k^\mu k^\nu. \tag{10.23}$$

### 10.3.3 The matter piece: the first law of entanglement

The matter piece $\delta S_{\rm ext}$ is where modular theory (Chapter 7) does its work. For the vacuum restricted to the diamond, the Bisognano–Wichmann theorem ((7.13), the load-bearing $2\pi$) gives the modular Hamiltonian $K$ generating the diamond's modular flow. For a small ball it takes the explicit local form

$$K \;=\; 2\pi \int_{B_\ell} \frac{\ell^2 - r^2}{2\ell}\,T_{00}\,dV, \tag{10.24}$$

the $2\pi$ inherited from (7.13) and the weight $(\ell^2 - r^2)/2\ell$ the diamond analogue of the linear boost weight $\lambda$ of route one. The *first law of entanglement* — the statement, rigorous for a first-order perturbation of the vacuum, that the change in entanglement entropy equals the change in modular energy — then reads

$$\delta S_{\rm ext} \;=\; \delta\langle K\rangle. \tag{10.25}$$

This is the entanglement-theoretic replacement for the Clausius relation of route one: instead of "heat over temperature equals entropy change," we have "modular-energy change equals entanglement-entropy change," and it is a *theorem* (Chapter 7) rather than a thermodynamic assumption. Inserting (10.24) and integrating over the small ball against a slowly varying $\delta\langle T_{\mu\nu}\rangle$ gives, in $d = 4$,

$$\delta S_{\rm ext} \;=\; \frac{2\pi^2\,\ell^5}{15}\,\delta\langle T_{\mu\nu}\rangle\,k^\mu k^\nu \;+\; O(\ell^6). \tag{10.26}$$

### 10.3.4 Equilibrium forces the Einstein equations

Now impose Axiom II, $\delta S_{\rm gen} = 0$, by adding (10.23) and (10.26):

$$\underbrace{-\,\frac{\pi\,\ell^5}{12\,\ell_{\rm P}^2}\,R_{\mu\nu}k^\mu k^\nu}_{\text{area piece}} \;+\; \underbrace{\frac{2\pi^2\,\ell^5}{15}\,\langle T_{\mu\nu}\rangle k^\mu k^\nu}_{\text{matter piece}} \;=\; 0. \tag{10.27}$$

The common factor $\ell^5$ cancels — the balance is scale-independent, as it must be for a *local* law — and solving for the curvature contraction,

$$R_{\mu\nu}k^\mu k^\nu \;=\; \frac{2\pi^2/15}{\pi/(12\ell_{\rm P}^2)}\,\langle T_{\mu\nu}\rangle k^\mu k^\nu \;=\; \frac{8\pi\,\ell_{\rm P}^2}{5}\,\langle T_{\mu\nu}\rangle k^\mu k^\nu. \tag{10.28}$$

As in route one, this holds for every null $k^\mu$, so by the same null-cone argument of §10.2.4 the full tensors agree up to a metric term, $R_{\mu\nu} = (\text{const})\,\langle T_{\mu\nu}\rangle + f g_{\mu\nu}$, and the same Bianchi-plus-conservation argument of §10.2.5 completes $R_{\mu\nu}$ into $G_{\mu\nu}$ and produces the integration constant $\Lambda$. The numerical prefactor $24\pi\ell_{\rm P}^2/5$ collapses — once $\ell_{\rm P}^2 = G\hbar/c^3$ is restored and the diamond normalisation is tracked honestly — to the same $8\pi G/c^4$ as before, for the same reason: the $1/4$ in the area law was chosen (by Bekenstein and Hawking, and here by Axiom I) to make exactly this happen. The result is identical to (10.19):

$$G_{\mu\nu} + \Lambda\, g_{\mu\nu} \;=\; \frac{8\pi G}{c^4}\,\langle T_{\mu\nu}\rangle. \tag{10.29}$$

**The two routes are one physics.** Route one imposes $\delta Q = T\,\delta S$ across a Rindler horizon; route two imposes $\delta S_{\text{gen}} = 0$ over a causal diamond. They share the same geometric engine (Raychaudhuri: curvature shrinks horizon area), the same load-bearing $2\pi$ (Unruh temperature in route one, Bisognano–Wichmann modular Hamiltonian in route two — and these are the *same* $2\pi$, as Chapter 7 showed), the same closing identity (Bianchi), and the same borrowed input (the area-law 1/4). Route two is the version stated in the language of the axioms: Axiom I supplies $S_{\text{gen}}$, Axiom II supplies $\delta S_{\text{gen}} = 0$, and Einstein's equations are their joint theorem — the theory's Theorem II.1, Einstein equations from Axioms I + II.

**Q 10.3** ***"Route one used a* thermal *temperature and route two used* entanglement*. Are those really the same thing, or is the agreement a coincidence?"***

They are genuinely the same thing, and Chapter 7 is where that identity was forged. The Unruh temperature of route one is the statement that the accelerated vacuum is a KMS (thermal) state at inverse temperature $\beta = 2\pi$ with respect to boost time. The modular Hamiltonian of route two is, by Bisognano–Wichmann (7.13), exactly $2\pi$ times that same boost generator. "Thermal at $\beta = 2\pi$" and "modular flow of the vacuum" are two names for one mathematical fact — the vacuum's entanglement structure across a horizon *is* a thermal structure, with the temperature fixed by the geometry. So the two routes are not analogous; they are the same computation, done once in the thermodynamic dialect and once in the entanglement dialect. That this book can pass between the dialects freely is the whole reason it can also pass between "general relativity" and "quantum mechanics."

## 10.4 Honest scope: what is derived, what is assumed

Both routes end at (10.19), and it would be easy to overclaim. The house discipline of this book forbids that, so we state plainly what has and has not been shown. The audited status (full theory review, June 2026) is that Einstein's equations are [DERIVED] from the axioms *given the area law as input*, not [PROVEN] from nothing.

1. **The area law is imported, not earned.** The coefficient 1/4 in $S = A/4\ell_{\text{P}}^2$ — equivalently the entropy density $\eta = k_{\text{B}}/(4\ell_{\text{P}}^2)$ forced by (10.18) — is not derived within this framework. It encodes the microscopic density of quantum-gravitational degrees of freedom per Planck area, a number on the same footing as the fine-structure constant: known empirically (and computed, for special black holes, by string microstate counting), but not delivered by thermodynamics alone. Every "derivation" of the area law inside the theory (including Chapter 5's first-law argument) presupposes the Hawking temperature or the equivalent, and so is not a derivation from nothing. [ASSUMED] — this is the one irreducible geometric input, and the theory is honest that it is thermodynamic, not microscopic, in the same sense that thermodynamics assumes the ideal-gas law and leaves its explanation to statistical mechanics.

2. **Matter must be well-behaved.** The heat flux (10.6) and the first law (10.25) assume ordinary, locally conserved stress–energy (10.12) satisfying the usual energy conditions (so that curvature focuses rather than defocuses). Exotic matter that violates them is outside the argument's scope.

3. **Equilibrium / stationarity is assumed.** Route one needs the Clausius *equality* (thermal equilibrium); out of equilibrium it degrades to an inequality $\delta Q \leq T\,\delta S$ and yields dissipative corrections rather than the clean Einstein equations. Route two needs the vacuum to be the maximal-entropy state to first order. Both fail, as principles, precisely in strongly non-equilibrium regimes — including, notably, a genuine superposition of distinct geometries, which is exactly the regime where this book's *other* predictions live (Chapters 11–12). The Einstein equations derived here are the equations of the *decohered* (equilibrium) branch; the inter-branch physics is a separate story.

4. **Equations, not the full action; and $\Lambda$ left open.** The derivation delivers the field equations (10.19), not uniquely the Einstein–Hilbert action — boundary terms, and higher-curvature terms that vanish on shell, are not fixed by the classical equations of motion (they matter only for the quantum path integral). And $\Lambda$ appears as a constant of integration: the derivation says it is constant, not what its value is. The cosmological-constant *problem* — why $\Lambda$ is what it is — is not solved here; it is the subject of Part IV (Chapter 14).

**Scope.** The result of this section is [DERIVED] in the precise sense of the status ledger (Appendix E): Einstein's equations follow from Axioms I and II *given the area law* $A/4\ell_{\mathrm{P}}^2$ as the one geometric input. Nothing above is claimed to be stronger than that.

**Q 10.4** ***"If the area law is assumed, and Einstein's equations follow from it, isn't this circular — the axioms already "know" the answer?"***

It is a fair charge, and the honest reply is: the circularity is real but has been *reduced*, not hidden. The naive worry is that the axioms secretly contain the Einstein–Hilbert action and "deriving" Einstein's equations just varies it — a tautology. That is *not* what happened here. Axiom I is a statement about *how many* degrees of freedom a surface carries (an entropy count); Einstein's equations are a statement about the *dynamics* of the geometry. These are logically independent — knowing the number of molecules per unit area does not tell you the gas law — and the bridge between them (Raychaudhuri plus Bianchi plus the first law) is a genuine, non-trivial computation, not a substitution. What remains irreducibly assumed is the *one* number 1/4, and the theory says so out loud: one geometric input (the area law) instead of two (area law *and* action). That is progress, honestly labelled, not a conjuring trick.

## 10.5 What this buys: general relativity, contained

Step back and take in what (10.19) means for the shape of the whole theory.

Once Einstein's equations follow from the axioms, *everything general relativity builds on them follows too.* The Schwarzschild solution and its black holes, the Friedmann–Robertson–Walker cosmologies and the expanding universe, gravitational waves, the bending of starlight, the perihelion precession of Mercury, the binary-pulsar spin-down, GPS clock corrections, the LIGO waveforms — every one of these is a *consequence* of (10.19), and (10.19) is now a theorem of the correspondence. The Quantum–Geometric Correspondence therefore does not *compete* with general relativity; it *contains* it. Everything the older rulebook got right — and Chapter 1 emphasised that it gets essentially everything right — the correspondence inherits automatically, because it reproduces the equation from which all of it flows.

> **The correspondence subsumes, it does not overturn.** A theory that replaced general relativity would have to re-explain every success in the list above. This one does not: it *derives* the equation those successes rest on, so they carry over untouched. What the correspondence *adds* is physics in the one regime general relativity forfeits — superpositions of geometry (Chapter 1) — where (10.19), an equilibrium relation, falls silent and the axioms' *non*-equilibrium content (Chapters 11–12) takes over. Contains the old, predicts the new.

And there is a second dividend, deferred to Part IV but worth naming now. The *same* two axioms that gave Einstein's equations, applied to the *cosmological* horizon rather than a local diamond, will fix the dark-energy density and the MOND acceleration scale $a_0 = cH_0/2\pi$ with no new parameters (Chapter 14). The cosmological constant $\Lambda$ that appeared here as a free integration constant is, in that chapter, tied to the horizon by the same entanglement-equilibrium condition, turning an open constant into a prediction. The machinery of this chapter is not a one-off trick for reproducing GR; it is the engine of the theory's quantitative predictions, run on different horizons.

**Q 10.5** ***"If the axioms contain all of general relativity, and GR passes every classical test, what is left for the theory to predict that GR does not?"***

Precisely the seam of Chapter 1. The derivation of this chapter is an *equilibrium* statement: it produces the geometry of a single, definite, decohered matter configuration. When matter is instead in a genuine superposition of distinct configurations — the microgram grain split by a millimetre — there is no single classical geometry, the Clausius equality and the entanglement-equilibrium condition no longer apply, and (10.19) makes no statement. That is the gap the rest of Part III fills: the axioms' behaviour *away* from equilibrium gives the branch-dependent geometries, the modular clock $K = 2\pi H_{\text{phys}}$, and the decoherence rate $\Gamma = C\,E_{\text{G}}/\hbar$ — the 1.6-nanosecond prediction that GR alone cannot make and that experiment can test. Same axioms; equilibrium gives you Einstein, non-equilibrium gives you the new physics.

## Chapter FAQ

**Q: Why does the surface gravity $\kappa$ cancel out? Shouldn't the answer depend on how hard the observer accelerates?** It cancels because it appears on both sides of the Clausius relation: in the temperature $T \propto \kappa$ (Unruh) *and*, through the Killing vector $\chi^\mu = -\kappa\lambda k^\mu$, in the heat flux $\delta Q \propto \kappa$. The acceleration is a property of the arbitrary observer we chose to define the horizon, not of the geometry, so it *must* drop out of any statement about the geometry — and it does. Its cancellation is a consistency check that the construction is observer-independent, exactly as the final equation (10.19) is.

**Q: Where did the cosmological constant come from, physically?** From the freedom in the trace term $f\, g_{\mu\nu}$ of (10.11), fixed by the Bianchi identity only up to an additive constant (10.15). Thermodynamically, $\Lambda$ is like a chemical-potential/pressure offset that the equation of state does not determine — an integration constant of the field equations. A thermodynamic derivation *naturally* leaves it free; that it is *small* rather than Planckian is a separate puzzle (the cosmological-constant problem), addressed in Part IV where the same axioms tie it to the cosmological horizon.

**Q: Both routes used the Raychaudhuri equation. Is that the real source of the "gravity" in the answer?** Yes, in a precise sense: Raychaudhuri is what carries *curvature* into the entropy balance. It is a pure identity of Lorentzian geometry (no field equation assumed) that says a congruence of light rays is focused by $R_{\mu\nu}k^\mu k^\nu$. The thermodynamics (Clausius / entanglement equilibrium) then *equates* that geometric focusing with the flux of matter energy, and the equality is Einstein's equation. So the division of labour is: geometry (Raychaudhuri) says how area responds to curvature; thermodynamics (the axioms) says area-change is heat/entropy; matter conservation + Bianchi glue them into a tensor equation. Remove any one and the derivation fails.

**Q: Does this mean gravity is not a force and there is no graviton?** It means the *metric* is a collective, thermodynamic variable, like temperature or pressure — and you no more quantize it directly than you quantize temperature. A "graviton" can still exist as the quantized ripple of the metric in the effective, low-energy regime (Chapter 1 built it), just as phonons are the quantized ripples of a crystal without the crystal's density being fundamental. What the equation-of-state view denies is that quantizing that ripple is the *route to* quantum gravity; the fundamental degrees of freedom are the ones the area law is counting, and they live underneath the geometry.

## Exercises

**Exercise 10.1.** Carry out the cancellation of §10.2 explicitly. Starting from the heat flux (10.6) and $T\,\delta S = T\eta\,\delta A$ with $T$ from (10.4) and $\delta A$ from (10.9), show that the factors of $\kappa$ and the integrals $\int(\cdots)\lambda\, d\lambda\, dA$ cancel, leaving the pointwise relation (10.10). Identify at which step the *same* affine weight $\lambda$ on both sides is what makes the integrals match.

**Exercise 10.2.** Prove the linear-algebra fact used in §10.2.4: if a symmetric tensor $t_{\mu\nu}$ satisfies $t_{\mu\nu}k^\mu k^\nu = 0$ for *all* null $k^\mu$ (in signature $(-,+,+,+)$), then $t_{\mu\nu} = f\, g_{\mu\nu}$ for some scalar $f$. (Hint: work at a point in an orthonormal frame; take $k = (1, \hat{n})$ for unit 3-vectors $\hat{n}$, and show that vanishing for all $\hat{n}$ forces the spatial block to be a multiple of $\delta_{ij}$ equal to $t_{00}$, and the mixed components to vanish.)

**Exercise 10.3.** Verify the coefficient match (10.18): show that requiring the derived equation (10.17) to coincide with (10.1) forces $\eta = c^3 k_{\mathrm{B}}/(4G\hbar) = k_{\mathrm{B}}/(4\ell_{\mathrm{P}}^2)$, using $\ell_{\mathrm{P}}^2 = G\hbar/c^3$. Comment on why the appearance of exactly the Bekenstein–Hawking $1/4$ is a consistency *output* of the thermodynamic argument rather than an independent assumption, given the discussion in §10.4.

## Solutions

*Solution (Exercise 10.1).* The Clausius relation is $\delta Q = T\,\eta\,\delta A$. Substitute:

$$\underbrace{-\kappa\int T_{\mu\nu}k^\mu k^\nu\,\lambda\,d\lambda\,dA}_{\delta Q} = \underbrace{\frac{\hbar\kappa}{2\pi c\,k_{\rm B}}}_{T}\,\eta\,\underbrace{\left(-\int R_{\mu\nu}k^\mu k^\nu\,\lambda\,d\lambda\,dA\right)}_{\delta A}.$$

Both sides carry an overall factor $-\kappa$; divide it out. Both sides carry the *same* measure $\int(\cdots)\lambda\,d\lambda\,dA$ — crucially, the same affine weight $\lambda$, which on the left came from the Killing vector $\chi^\mu = -\kappa\lambda k^\mu$ (heat grows with boost energy $\propto \lambda$) and on the right came from integrating the Raychaudhuri expansion $\theta = -R_{\mu\nu}k^\mu k^\nu\,\lambda$. Because the two $\lambda$-weightings are identical, the integrals cancel as *integrals*, term by term over the horizon, leaving the integrand identity $T_{\mu\nu}k^\mu k^\nu = (\hbar\eta/2\pi c k_{\rm B})\,R_{\mu\nu}k^\mu k^\nu$, which is (10.10). Had the weights differed, only an integrated (non-local) relation would survive; it is the matching of the affine weight that makes the result *pointwise* — and hence a local field equation.

*Solution (Exercise 10.2).* Work at one point in an orthonormal frame, $g = \mathrm{diag}(-1,1,1,1)$. A null vector is $k^\mu = (1,\hat n)$ with $|\hat n| = 1$. Then

$$0 = t_{\mu\nu}k^\mu k^\nu = t_{00} - 2\,t_{0i}n^i + t_{ij}n^i n^j \qquad \text{for all unit } \hat n.$$

This is a function on the sphere of directions that must vanish identically. Average over $\hat n$: the linear term $-2t_{0i}n^i$ averages to zero, and $\langle n^i n^j\rangle = \frac{1}{3}\delta^{ij}$, giving $t_{00} + \frac{1}{3}\,t_{ii} = 0$. Next, replace $\hat n \to -\hat n$ and subtract: this kills the even terms and forces the odd term to vanish for all $\hat n$, so $t_{0i} = 0$. With the linear term gone, $t_{00} + t_{ij}n^i n^j = 0$ for all $\hat n$ means the quadratic form $t_{ij}n^i n^j$ is constant on the sphere, which for a symmetric matrix forces $t_{ij} = c\,\delta_{ij}$ (all eigenvalues equal); then $t_{00} = -c$. Collecting: $t_{\mu\nu} = c\,\mathrm{diag}(-1,1,1,1) = c\,g_{\mu\nu}$, i.e. $t_{\mu\nu} = f g_{\mu\nu}$ with $f = c$. So a symmetric tensor is invisible to all null vectors exactly when it is a multiple of the metric — which is why (10.10) determines $T_{\mu\nu}$ and $\alpha R_{\mu\nu}$ only up to the term $f g_{\mu\nu}$ that §10.2.5 then pins.

*Solution (Exercise 10.3).* The derived equation is $G_{\mu\nu} + \Lambda g_{\mu\nu} = (2\pi c k_{\rm B}/\hbar\eta)\,T_{\mu\nu}$ and the target is $G_{\mu\nu} + \Lambda g_{\mu\nu} = (8\pi G/c^4)T_{\mu\nu}$. Equate the coefficients:

$$\frac{2\pi c\,k_{\rm B}}{\hbar\,\eta} = \frac{8\pi G}{c^4} \implies \eta = \frac{2\pi c\,k_{\rm B}\,c^4}{8\pi G\,\hbar} = \frac{c^3 k_{\rm B}}{4G\hbar}.$$

Using $\ell_{\rm P}^2 = G\hbar/c^3$, i.e. $c^3/(G\hbar) = 1/\ell_{\rm P}^2$, this is $\eta = k_{\rm B}/(4\ell_{\rm P}^2)$ — the Bekenstein–Hawking entropy density, coefficient $1/4$ included. Why this is an *output*, not a smuggled input: the thermodynamic argument (Clausius + Unruh + Raychaudhuri + Bianchi) is structurally complete before $\eta$ is fixed — it delivers Einstein's equations for *any* constant $\eta$, just with the wrong overall strength of gravity. Demanding that the strength match the observed $8\pi G/c^4$ then *forces* $\eta = 1/4\ell_{\rm P}^2$. As §10.4 stresses, this does not derive the $1/4$ from microscopics; it shows that the $1/4$ is the *unique* value making horizon thermodynamics consistent with the measured Newton constant. The $1/4$ is thus the single irreducible number the theory imports — the analogue of an equation of state's one empirical constant.

**What to carry forward**

1. **Einstein's equations are a theorem, not an axiom.** Imposing the Clausius relation $\delta Q = T\,\delta S$ across a local Rindler horizon at every point (route one), or entanglement equilibrium $\delta S_{\text{gen}} = 0$ over every causal diamond (route two, Axioms I+II), yields $G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G\,T_{\mu\nu}/c^4$. The engine is the same in both: Raychaudhuri carries curvature into the area, the load-bearing $2\pi$ sets the temperature, and the Bianchi identity fixes the trace term and coughs up $\Lambda$ as an integration constant. [DERIVED].
2. **One number is borrowed.** The area law's coefficient $1/4$ ($\eta = k_{\text{B}}/4\ell_{\text{P}}^2$) is [ASSUMED] — the single irreducible geometric input, forced by consistency with the measured $8\pi G/c^4$ but not derived from microscopics. The derivation is thermodynamic (like the ideal gas law), honest about its one empirical constant, and valid only in the equilibrium/decohered regime; the cosmological-constant *value* stays open until Chapter 14.
3. **The correspondence contains general relativity.** Because Einstein's equations follow, so does everything GR builds on them — Schwarzschild, FRW, gravitational waves, every classical test. QGC does not compete with GR; it subsumes it, and the *same* axioms will, on the cosmological horizon, fix the dark-energy and MOND scales (Chapter 14). The new physics lives where this chapter's equilibrium derivation falls silent: superpositions of geometry (Chapters 11–12).

# 11 The Constraint and the Influence Functional

**Synopsis.** This is the chapter the whole book has been walking toward. We compute, with no skipped steps, the object that decides everything: the overlap of the two gravitational fields a superposed mass sources. First we learn what a *constraint* is, building from the electromagnetic Gauss law up to the gravitational one — the Wheeler–DeWitt equation $\hat{H}_{\text{tot}} |\Psi\rangle = 0$, which says the total Hamiltonian annihilates physical states. We see why this single fact forces the gravitational field to be entangled with the matter it dresses, and why — unlike gas molecules or photons — that record can be neither screened nor refrigerated away. Then we build the Feynman–Vernon influence functional from scratch on a toy oscillator bath, and turn the machine on the real problem. The two mass branches source two coherent states of the linearized field (Chapter 4); their overlap is the visibility law of Chapter 2; the mode integral evaluates cleanly; and out of it drops the energy scale $E_{\text{G}} = GM^2/d$ — the gravitational self-energy of the split, the same $6.7 \times 10^{-26}$ J that set the cover's nanosecond. What is [PROVEN] here (the scale) and what is not (that it is a decoherence *rate*) we keep rigorously apart.

**You will be able to.** state the Wheeler–DeWitt constraint and explain operationally why it forces matter–field entanglement; derive the Feynman–Vernon influence functional for a bath of oscillators; compute the exact coherent-state decoherence functional $F(t)$ mode by mode and extract $E_{\text{G}} = GM^2/d$; say precisely why wider superpositions decohere *slower*.

**Prerequisites.** Chapter 2 (visibility law $\mathcal{V} = |\langle E_L|E_R\rangle|$, the Gaussian onset), Chapter 4 (coherent states, the multimode overlap, the graviton coherent state $|g_A\rangle$), Chapter 3 (Newtonian potential, the $3+1$ split). Nothing else new.

## 11.1 What a constraint is, in one force we already trust

Before we touch gravity, let us settle a question that sounds abstract and turns out to be the crux of the whole matter: *which field configurations does a theory actually permit?* The naive answer — "any of them, the field is a free variable" — is wrong for every gauge theory, and understanding why it is wrong in electromagnetism, where we have complete confidence, is the fastest route to understanding what gravity does.

### 11.1.1 The Gauss law forbids most fields

Take electromagnetism. One is tempted to say: the electric field $\vec{E}(\vec{x})$ can be anything, and Maxwell's equations tell it how to change in time. This is false. At every instant, before any talk of time evolution, the field must satisfy the *Gauss constraint*

$$\nabla \cdot \vec{E}(\vec{x}) \;=\; \frac{\rho(\vec{x})}{\epsilon_0}, \tag{11.1}$$

where $\rho$ is the charge density. Read (11.1) operationally: it is not a law of motion — there is no time derivative in it — but a law of *allowed states*. A field configuration that violates it does not describe a late-arriving or mistaken electromagnetism; it describes *no* electromagnetism at all. It is not in the theory's vocabulary.

Why is a static equation a constraint rather than a dynamical law? Because it is the piece of Maxwell's equations that contains no time derivative of the thing it constrains. In the Hamiltonian formulation, $\nabla \cdot \vec{E} - \rho/\epsilon_0$ is a quantity whose Poisson bracket with the Hamiltonian generates the gauge transformations $\vec{A} \to \vec{A} + \nabla\lambda$ — the redundancies of the description. That it generates a symmetry rather than a motion is exactly what makes it a constraint. We will meet the same structure in gravity, generating a deeper symmetry.

**Q 11.1 *"If most field configurations are forbidden, how can I ever write one down? I set up charges and solve for $\vec{E}$ all the time."***

You never write $\vec{E}$ down freely — you write the charges down freely and let (11.1) *solve* for the longitudinal (divergence-carrying) part of $\vec{E}$. That is what "solving for the Coulomb field" is: enforcing the constraint. The only genuinely free electromagnetic data are the transverse (radiative) degrees of freedom, the photons; the longitudinal part is a slave to the charges. Hold this decomposition — free radiation versus slaved near-field — because it is the fulcrum of the whole chapter. In gravity the mass will slave its Newtonian field in exactly this way, and the slaving is what records which-path.

The physical consequence for a quantum charge is immediate and, once seen, hard to un-see. Put an electron in a superposition of two positions, $(|L\rangle + |R\rangle)/\sqrt{2}$. Each branch carries a definite charge density, so (11.1) forces each branch to carry its own Coulomb field:

$$|\Psi\rangle \;=\; \frac{1}{\sqrt{2}}\big(|L\rangle\,|E_L\rangle + |R\rangle\,|E_R\rangle\big), \qquad \vec{E}_A = \text{Coulomb field of a charge at } A. \tag{11.2}$$

The electron *cannot* sit in superposition while its field stays in one place; that would violate the constraint in each branch. So constraint-forced entanglement between a charged particle and its field is not exotic — it is the mundane content of the Gauss law. We will ask, at the end of the chapter, why this does *not* decohere the electron (Chapter 12 answers in full), and the answer will pin down exactly what is special about gravity.

### 11.1.2 From a spatial law to a law about time

Here is the fork that separates gravity from every other force. The Gauss constraint (11.1) is a *spatial* law: it constrains the field at each point of space and generates internal (gauge) transformations that have nothing to do with time. The energy of the electromagnetic field — its Hamiltonian — is a perfectly ordinary, non-zero, unconstrained quantity, and time is an external parameter ticking along in the background, just as in non-relativistic quantum mechanics.

Gravity is different because in gravity there is no external background to tick against. Spacetime *is* the gravitational field. There is no stage behind the stage. When we build the constraint for gravity, it will not be a statement about the field at each point of space — it will be a statement about *time itself*, and it will say something startling: that the total energy of a closed gravitating system is zero. To see why, we need the $3+1$ split of general relativity, which we now sketch just far enough to write the constraint down honestly.

## 11.2 The Wheeler–DeWitt constraint

### 11.2.1 Splitting spacetime into space and time

General relativity, as Chapter 3 built it, is a theory of a four-dimensional geometry $g_{\mu\nu}$ with no preferred slicing into "space" and "time." To do Hamiltonian mechanics — to ask "what is the state now, and how does it evolve?" — we must nonetheless choose a slicing: foliate spacetime into a stack of three-dimensional spatial slices labelled by a time coordinate. This is the Arnowitt–Deser–Misner ($3+1$, or ADM) formulation, and we need only its shape, not its every index.

On each slice lives a three-dimensional geometry, encoded in the spatial metric $h_{ij}(\vec{x})$ — the genuine dynamical field, the analogue of $\vec{A}$ in electromagnetism. Its conjugate momentum $\pi^{ij}$ (built from how fast the slice is curving, the extrinsic curvature) is the analogue of $\vec{E}$. Two functions tell us how successive slices are stacked: the *lapse* $N(\vec{x})$ (how much proper time separates a slice from the next, at each point) and the *shift* $N^i(\vec{x})$ (how the spatial coordinates slide from slice to slice). The lapse and shift are not dynamical — they encode our free choice of coordinates, our freedom in *how* to slice.

When one writes the Einstein–Hilbert action in these variables and passes to the Hamiltonian, a remarkable thing happens: the total Hamiltonian is a sum of the lapse and shift multiplying two quantities,

$$H_{\text{tot}} \;=\; \int d^3x \left[ N(\vec{x})\,\mathcal{H}_\perp(\vec{x}) + N^i(\vec{x})\,\mathcal{H}_i(\vec{x}) \right], \tag{11.3}$$

and there is *nothing else*. The Hamiltonian is built entirely out of $\mathcal{H}_\perp$ (the *Hamiltonian constraint*, riding the lapse) and $\mathcal{H}_i$ (the *momentum constraints*, riding the shift). This is the signature of a theory with no external time: the whole energy is a sum of constraints.

**For the curious — Why the Hamiltonian is pure constraint**

For an ordinary system — a pendulum, a field on a fixed spacetime — the Hamiltonian is a definite function of the state, and it generates evolution in an external time. For general relativity there is no external time to evolve in; the only "motion" is the relabelling of coordinates, which is a redundancy, not a physical change. A theory all of whose evolution is redundancy has a Hamiltonian that is a sum of constraints, each generating one of the redundancies: the momentum constraints $\mathcal{H}_i$ generate spatial coordinate changes (diffeomorphisms of the slice), and the Hamiltonian constraint $\mathcal{H}_\perp$ generates the pushing of the slice forward in time — which, since the slicing was arbitrary, is *also* a redundancy. This is Dirac's analysis of "generally covariant" systems, and it is forced by diffeomorphism invariance alone. The lapse and shift $N, N^i$ are Lagrange multipliers: varying the action with respect to them yields the constraints $\mathcal{H}_\perp = 0$, $\mathcal{H}_i = 0$.

### 11.2.2 The equation that annihilates the state

Quantizing, the metric and its momentum become operators, and a physical quantum state $|\Psi\rangle$ is a wavefunctional of three-geometries. The classical requirement $\mathcal{H}_\perp = 0$ becomes an operator statement: the physical state must be *annihilated* by the Hamiltonian-constraint operator. Including matter, whose stress-energy also enters $\mathcal{H}_\perp$, and integrating against the lapse, this is the *Wheeler–DeWitt equation*:

$$\hat{H}_{\text{tot}}\,|\Psi_{\text{phys}}\rangle \;=\; 0. \tag{11.4}$$

In words, and this is the sentence to carry: *the total Hamiltonian of a closed gravitating universe — gravity plus matter — gives zero when applied to any physical state.* The total energy is not merely conserved; it vanishes identically as an operator statement on the physical Hilbert space.

The reason is the one from §11.1.2, now made sharp. $\hat{H}_{\text{tot}}$ is the generator of "evolution in the time coordinate," but the time coordinate was our free choice — pushing the slice forward is a diffeomorphism, a redundancy. A generator of a pure redundancy must annihilate physical states, exactly as $\nabla \cdot \vec{E} - \rho/\epsilon_0$ (the generator of electromagnetic gauge) annihilates physical states in QED. The difference, all the difference, is *which* symmetry: gauge in one case, *time* itself in the other.

### 11.2.3 The problem of time, stated plainly

Equation (11.4) carries a shock. If $\hat{H}_{\text{tot}}\,|\Psi\rangle = 0$, then the Schrödinger equation $i\hbar\,\partial_t\,|\Psi\rangle = \hat{H}_{\text{tot}}\,|\Psi\rangle$ reads $i\hbar\,\partial_t\,|\Psi\rangle = 0$: the physical state *does not evolve.* There is no external time in which anything happens. This is the *problem of time*, and it is not a paradox to be dissolved but a feature to be understood: in a closed universe there is no clock outside the universe, so "time" cannot be an external parameter. It must be *relational* — reconstructed from correlations *within* the state, one subsystem serving as a clock for another. Chapter 7 built the mathematical machine for exactly this (the crossed product that hands an observer a clock); here we need only the physical upshot.

**Q 11.2 *"If the physical state never changes, how can anything — let alone decoherence — happen at all?"***

Nothing happens *to the whole state*; things happen *relative to a clock built from part of it.* The situation is the one from Chapter 2's Bell pair: the joint state is a fixed, definite vector, yet an observer with access to only one half sees rich statistics. Here the fixed vector is the whole universe; the "observer" is the laboratory subsystem; and what looks like time evolution of the lab is the correlation of the lab with a degree of freedom playing clock. The decoherence we are about to compute is a statement about the reduced state of the matter, read at successive clock readings — and *that* changes, even though the global $|\Psi_{\text{phys}}\rangle$ does not. The one physical input this whole chapter rests on is (11.4); everything downstream is a consequence of taking it seriously.

## 11.3 The constraint forces the field to keep a record

We now cash the constraint out for the one configuration this book cares about: a mass in superposition. The argument is short, it uses only what we have built, and it is the physical heart of the chapter — so we take it slowly.

### 11.3.1 Each branch drags its own field

In the weak-field, slow-matter (Newtonian) regime — which, as Chapter 1's power counting showed, is where all laboratory physics lives — the Hamiltonian constraint reduces to a statement we can read at a glance. Writing the Newtonian potential $\Phi$ for the metric perturbation, the $\mathcal{H}_\perp = 0$ condition becomes the quantized Poisson equation:

$$\nabla^2 \hat{\Phi}(\vec{x}) \;=\; 4\pi G\, \hat{\rho}(\vec{x}), \tag{11.5}$$

the exact gravitational analogue of the Gauss law (11.1), with $4\pi G$ where electromagnetism had $1/\epsilon_0$. Note the single power of $G$: the constraint is *linear* in Newton's constant. This will matter enormously in Chapter 12, where the whole $G^1$-versus-$G^2$ question turns on whether the physics inherits its $G$ from the constraint (once) or from a coupling squared (twice).

Now put the mass in superposition, $(|L\rangle + |R\rangle)/\sqrt{2}$. Branch $|L\rangle$ has mass density $\rho_L(\vec{x}) = M\delta^3(\vec{x} - \vec{x}_L)$; branch $|R\rangle$ has $\rho_R(\vec{x}) = M\delta^3(\vec{x} - \vec{x}_R)$. The constraint (11.5) must hold *in each branch*, so it solves, branch by branch, for the field:

$$\Phi_L(\vec{x}) = -\frac{GM}{|\vec{x} - \vec{x}_L|}, \qquad \Phi_R(\vec{x}) = -\frac{GM}{|\vec{x} - \vec{x}_R|}. \tag{11.6}$$

There is no single $\Phi$ that satisfies (11.5) for both densities at once — one field cannot be sourced by a mass in two places. The would-be state $(|L\rangle + |R\rangle)/\sqrt{2} \otimes |\Phi_{\text{one}}\rangle$ is therefore *not physical*: it violates the constraint in at least one branch. The only constraint-satisfying state is the entangled one,

$$|\Psi_{\text{phys}}\rangle \;=\; \frac{1}{\sqrt{2}}\Big(|L\rangle\,|g_L\rangle + |R\rangle\,|g_R\rangle\Big), \tag{11.7}$$

where $|g_A\rangle$ is the gravitational field sourced by the mass at $A$. The constraint has *forced* the matter and its field to be entangled — not as a dynamical process that unfolds over time, but as a condition on the allowed states, imposed at every instant. [PROVEN] in linearized gravity; this is the content of (11.5), no more.

### 11.3.2 The record is mandatory, ungaugeable, uncoolable

Set (11.7) beside the decoherence template of Chapter 2. There, an environment particle scattered off the mass and recorded which branch it was in, giving $(|L\rangle\,|E_L\rangle + |R\rangle\,|E_R\rangle)/\sqrt{2}$, and the surviving coherence was the overlap $\mathcal{V} = |\langle E_L|E_R\rangle|$ (equation eq:visibility-law). Equation (11.7) is *the same structure* — but the recording environment is now the gravitational field the mass sources, and the record is there because a *law* put it there, not because a stray molecule happened by.

This is the decisive difference, and it deserves its own statement.

**Ordinary decoherence is contingent; gravitational decoherence is mandatory.** A gas molecule's which-path record can be removed — pump out the gas, cool the walls, shield the photons — because the coupling that wrote it is an accident of the environment, tunable to zero in principle. The gravitational record cannot be removed, because the coupling that writes it is the constraint (11.5), a law of the theory. There is no vacuum good enough, no refrigerator cold enough, no shield dense enough: as long as the mass has mass, its field satisfies the constraint, and the constraint entangles. There is no gauge freedom left in which to hide the record — the Newtonian field is the constraint-solved, longitudinal part (§11.1.1), fully determined by the source, with no adjustable transverse piece to absorb the difference.

This is precisely the answer promised in Chapter 2's reader question — "is gravitational which-path information contingent too, and can it be undone?" The answer is no, and now we see why: the record lives in the constraint-slaved field, and the constraint is not optional. That "no" is what will make the decoherence floor of Chapter 15 *uncoolable*, and the theory falsifiable. Everything now reduces to one number — the overlap $\langle g_L|g_R\rangle$ and its time dependence — and computing it, honestly and in full, is the rest of the chapter.

**Q 11.3** ***"You showed in §11.1.1 that electromagnetism forces exactly the same entanglement* (11.2)*. Why doesn't my electron decohere from its own Coulomb field, then?"***

This is the sharpest possible objection, and the honest answer is that the entanglement alone does *not* settle it — both fields keep a record. What differs is whether the record turns into an irreversible *loss* of coherence or into a mere reversible *phase*. That distinction is the entire business of Chapter 12, and it hinges on the constraint's *type*: the electromagnetic Gauss law is a spatial constraint (it generates gauge, time stays external, and the Coulomb entanglement produces a recoverable phase rotation); the gravitational constraint is a *Hamiltonian* constraint (it generates time, $\hat{H}_{\text{tot}} = 0$, and there is no external clock to make the phase reversible). We flag the distinction here and earn it there. For this chapter we compute the object common to both cases — the field overlap — and identify the energy scale it carries.

## 11.4 The influence functional, built from scratch

To turn $\langle g_L|g_R\rangle$ into a time-dependent, calculable quantity, we need the tool physicists use whenever a system of interest is coupled to a large "bath" we do not track: the Feynman–Vernon *influence functional*. Rather than quote it, we build it on the simplest bath that contains the full mechanism — a two-state system coupled to one harmonic oscillator — and then read off the general structure. When we meet the gravitational field, it will be a bath of such oscillators (one per mode), and we will already own the machine.

### 11.4.1 The setup: a qubit watched by an oscillator

Let the system be a qubit with basis $|L\rangle, |R\rangle$ (our two mass branches), and let the bath be a single harmonic oscillator of frequency $\omega$ (one field mode). The coupling records which branch: the oscillator is pushed one way in branch $|L\rangle$ and the other way in branch $|R\rangle$. The cleanest model of "the bath is displaced according to the branch" is

$$H \;=\; \hbar\omega\, a^\dagger a \;+\; \hbar\big(f_s\, a^\dagger + f_s^*\, a\big), \qquad s \in \{L, R\}, \tag{11.8}$$

where the branch label $s$ selects a complex drive $f_s$ — the amplitude with which branch $s$ pushes this mode. (For gravity, $f_s$ will be built from the Fourier component of the Newtonian potential of a mass at position $s$; here it is just a number.) The oscillator starts in its ground state $|0\rangle$. Everything we want follows from one fact we proved in Chapter 4: a harmonic oscillator driven by a c-number stays a coherent state.

### 11.4.2 Solving the driven oscillator

*Derivation.* In branch $s$, the Hamiltonian (11.8) is a displaced oscillator. Its Heisenberg equation for the annihilation operator is

$$\dot a \;=\; \frac{i}{\hbar}[H, a] \;=\; -i\omega a - i f_s,$$

a linear first-order equation with the solution

$$a(t) \;=\; a(0)\, e^{-i\omega t} \;-\; \frac{f_s}{\omega}\big(1 - e^{-i\omega t}\big).$$

Correspondingly the state, starting from $|0\rangle$, is at time $t$ a coherent state $|\chi_s(t)\rangle$ (up to an overall phase) with amplitude

$$\chi_s(t) \;=\; -\frac{f_s}{\omega}\big(1 - e^{-i\omega t}\big). \tag{11.9}$$

The amplitude starts at zero (the mode has not yet learned anything at $t = 0$), swings out toward its equilibrium displacement $-f_s/\omega$, and — because the drive was switched on abruptly — oscillates around it forever after. That $\chi_s(0) = 0$ is the mathematical face of a physical fact we will lean on: the which-path record is not present at $t = 0$; it accumulates.

### 11.4.3 Integrating out the bath

The joint state at time $t$, for the qubit prepared in $(|L\rangle + |R\rangle)/\sqrt{2}$ and the oscillator in $|0\rangle$, is

$$|\Psi(t)\rangle = \frac{1}{\sqrt{2}}\Big(|L\rangle\,|\chi_L(t)\rangle + |R\rangle\,|\chi_R(t)\rangle\Big). \tag{11.10}$$

We do not track the oscillator, so we trace it out (Chapter 2's partial trace). The off-diagonal element of the qubit's reduced density matrix — the one interference reads — is multiplied by the bath overlap. This factor, viewed as a functional of the two branch histories, *is* the Feynman–Vernon influence functional:

$$\mathcal{F}(t) \;=\; \langle \chi_L(t)\,|\,\chi_R(t)\rangle\,. \tag{11.11}$$

It is the qubit's memory of the record the oscillator kept, and its modulus is the surviving visibility — the same statement as $\mathcal{V} = |\,\langle E_L|E_R\rangle\,|$, now for the oscillator bath.

### 11.4.4 Evaluating the toy functional

**Worked Example 11.1 — The one-mode influence functional**

Use the coherent-state overlap from Chapter 4, $|\langle\alpha|\beta\rangle|^2 = \exp(-|\alpha-\beta|^2)$. The modulus of (11.11) is therefore

$$|\mathcal{F}(t)|^2 = \exp(-|\chi_L(t) - \chi_R(t)|^2). \tag{11.12}$$

The difference of the two amplitudes, from (11.9), is

$$\chi_L(t) - \chi_R(t) = -\frac{f_L - f_R}{\omega}(1 - e^{-i\omega t}), \tag{11.13}$$

so, using $|1 - e^{-i\omega t}|^2 = 2(1 - \cos\omega t)$,

$$-\ln|\mathcal{F}(t)|^2 = \frac{|f_L - f_R|^2}{\omega^2}\, 2(1 - \cos\omega t). \tag{11.14}$$

Three features are worth naming, because all three survive to the real problem. *One:* the exponent depends only on the *difference* $f_L - f_R$ of the drives — a mode pushed identically by both branches keeps no record and contributes nothing. *Two:* at short times, expanding $1 - \cos\omega t \approx \frac{1}{2}\omega^2 t^2$, the exponent is $|f_L - f_R|^2 t^2$ — *quadratic* in $t$, the Gaussian onset of Chapter 2 (eq:gaussian-onset), forced by unitarity to start flat. *Three:* the exponent does not grow without bound — it oscillates, capped at $4|f_L - f_R|^2/\omega^2$. A single mode cannot decohere anything permanently; it loses the record and gets it back every period. Only a *continuum* of modes, whose periods never realign, can produce irreversible decay — which is exactly what the gravitational field provides, and exactly the subtlety Chapter 12 must handle with care.

### 11.4.5 The general structure

For a bath of many oscillators, indexed by mode $\vec{k}$ with frequency $\omega_k$ and branch-difference drive $\delta f(\vec{k}) \equiv f_L(\vec{k}) - f_R(\vec{k})$, the overlap factorizes over modes (independent oscillators; Chapter 4's multimode overlap eq:multimode-overlap). The exponents add, and the influence functional's modulus is

$$-\ln|\mathcal{F}(t)|^2 \;=\; \int \frac{d^3k}{(2\pi)^3}\, \frac{|\delta f(\vec{k})|^2}{\omega_k^2}\, 2(1 - \cos\omega_k t). \tag{11.15}$$

This is the master formula. It says: the loss of coherence is a sum over bath modes of (branch-difference coupling)$^2\times$(a universal time factor). Compare it, mode for mode, with the fully solvable dephasing model of Chapter 2 (equation eq:gaussian-onset) — it is the same object, with the abstract coupling $g_k$ now identified as the physical drive $\delta f(\vec{k})/\omega_k$. We have built the machine. We now feed it gravity.

## 11.5 The main event: the gravitational decoherence functional

Everything is now in place. The two mass branches source two coherent states of the linearized gravitational field — Chapter 4 proved this and handed us the amplitude (equation eq:graviton-coherent). The decoherence functional is their overlap. We compute it mode by mode, exactly, and read off what it means.

### 11.5.1 The two field states and their difference

From Chapter 4, the field sourced by the mass at position $\vec{x}_A$ is the graviton coherent state $|g_A\rangle = D(\alpha_A)\,|0\rangle$ with mode amplitude

$$\alpha_A(\vec{k}) \;=\; -\frac{4\pi GM}{k^2}\,\frac{e^{-i\vec{k}\cdot\vec{x}_A}}{\sqrt{2\hbar\omega_k}}, \tag{11.16}$$

the Fourier transform of the Newtonian potential (11.6): the $1/k^2$ is the Fourier face of the Poisson constraint $\nabla^2\Phi = 4\pi G\rho$, and the phase $e^{-i\vec{k}\cdot\vec{x}_A}$ places the mass at $\vec{x}_A$. The decoherence functional is the overlap of the two, evolved in the relational time $t$; by the driven-oscillator solution of §11.4.2 the branch difference at time $t$ carries the same $(1 - e^{-i\omega_k t})$ factor as the toy model, so the mode-difference amplitude is

$$\delta\alpha(\vec{k}) \;\equiv\; \alpha_L(\vec{k}) - \alpha_R(\vec{k}) \;=\; -\frac{4\pi GM}{k^2\sqrt{2\hbar\omega_k}}\,(e^{-i\vec{k}\cdot\vec{x}_L} - e^{-i\vec{k}\cdot\vec{x}_R}). \tag{11.17}$$

This is the gravitational $\delta f(\vec{k})$: the difference in how the two branches push mode $\vec{k}$. Everything hinges on its magnitude squared.

### 11.5.2 The mode-difference squared

*Derivation.* Take the modulus squared of (11.17). The two source phases combine through

$$\left|e^{-i\vec{k}\cdot\vec{x}_L} - e^{-i\vec{k}\cdot\vec{x}_R}\right|^2 = 2 - 2\cos(\vec{k}\cdot(\vec{x}_L - \vec{x}_R)) = 2(1 - \cos(\vec{k}\cdot\vec{d})),$$

with $\vec{d} = \vec{x}_L - \vec{x}_R$ the separation vector, $|\vec{d}| = d$. Hence

$$|\delta\alpha(\vec{k})|^2 = \frac{(4\pi GM)^2}{k^4\,(2\hbar\omega_k)}\,2(1 - \cos(\vec{k}\cdot\vec{d})). \tag{11.18}$$

Read the ingredients: the $(GM)^2/k^4$ is the squared strength of the Newtonian source in Fourier space; the $1/(2\hbar\omega_k)$ is the zero-point normalization of each mode; and the $2(1 - \cos\vec{k}\cdot\vec{d})$ is the interference between the two source positions — it vanishes for modes so long-wavelength that $\vec{k}\cdot\vec{d} \to 0$ (they cannot resolve the split) and reaches its maximum for modes with wavelength $\sim d$ (which resolve it best). The split is invisible to modes that cannot fit inside it: this is the physics that will make wider splits decohere *slower*.

### 11.5.3 Assembling the decoherence exponent

Insert (11.18) into the master formula (11.15), carrying the extra $2(1 - \cos\omega_k t)$ time factor from the mode evolution (equation §11.4.2). Writing the decoherence exponent as $\Gamma(t) \equiv -\ln|F(t)|^2$, where $F(t) = \langle g_L(t)|g_R(t)\rangle$ is the gravitational decoherence functional,

$$\Gamma(t) \;=\; \int \frac{d^3k}{(2\pi)^3}\,\frac{(4\pi GM)^2}{k^4\,(2\hbar\omega_k)}\,2(1 - \cos(\vec{k}\cdot\vec{d}))\,2(1 - \cos\omega_k t). \tag{11.19}$$

This is the exact gravitational decoherence functional in the linearized free-field theory — the object this chapter set out to compute. Every symbol was earned: $(4\pi GM)^2/k^4$ from the constraint-solved source, $1/(2\hbar\omega_k)$ from the mode normalization, one cosine factor from the separation, one from the time evolution. It is the exact analogue of the toy model's (11.14), now integrated over the continuum of graviton modes. The visibility the laboratory measures is $\mathcal{V}(t) = |F(t)| = \exp(-\frac{1}{2}\Gamma(t))$.

### 11.5.4 Doing the angular integral

*Derivation.* Go to spherical coordinates in $\vec{k}$ with the polar axis along $\vec{d}$, so $\vec{k}\cdot\vec{d} = kd\cos\theta \equiv kd\,\mu$. The only angular dependence is in $\cos(\vec{k}\cdot\vec{d})$; averaging it over the sphere,

$$\frac{1}{2}\int_{-1}^{1} d\mu\,\left(1 - \cos(kd\,\mu)\right) = 1 - \frac{\sin(kd)}{kd},$$

where the elementary integral $\int_{-1}^{1}\cos(kd\mu)\,d\mu = 2\sin(kd)/(kd)$ was used. The angular factor $1 - \sin(kd)/(kd)$ is the isotropic average of the two-source interference: it goes as $(kd)^2/6$ for $kd \ll 1$ (long-wavelength modes barely resolve the split) and settles to 1 for $kd \gg 1$ (short-wavelength modes see two distinct sources). After the angular integral, using $d^3k = 4\pi k^2\,dk$ with the average folded in,

$$\Gamma(t) = \frac{(4\pi GM)^2}{2\pi^2\hbar}\int_0^\infty \frac{dk}{k^2}\,\frac{1}{\omega_k}\left(1 - \frac{\sin kd}{kd}\right)2(1 - \cos\omega_k t). \tag{11.20}$$

The angular structure has done its job; what remains is one radial integral over $k$, with the mode dispersion $\omega_k$ still to be specified.

### 11.5.5 The energy scale that emerges

We do not need the full time-resolved radial integral to extract the physics that matters most — and this is fortunate, because (as Chapter 12 will show in painful honesty) the time dependence of (11.20) hides a genuine subtlety. What we *can* read off cleanly, rigorously, and without controversy is the *energy scale* the functional carries.

The cleanest route is to recognize that the decoherence exponent is measuring the gravitational energy stored in the *difference* of the two field configurations. In position space, the branch-difference density is $\Delta\rho = \rho_L - \rho_R$, and the gravitational self-energy of this difference is the quantity

$$E_{\mathrm{G}} \;=\; \frac{G}{2}\int d^3x\,d^3y\,\frac{\Delta\rho(\vec{x})\,\Delta\rho(\vec{y})}{|\vec{x}-\vec{y}|}. \tag{11.21}$$

This is the same energy that appears in (11.19) when one strips the time factor: the $(GM)^2/k^4$ source strength, integrated against the Newtonian kernel $4\pi/k^2$ (whose position-space face is $1/|\vec{x}-\vec{y}|$), is exactly the Fourier representation of (11.21). Let us evaluate it for the mass split.

**Worked Example 11.2 — The self-energy of a superposed mass**

With $\Delta\rho(\vec{x}) = M[\delta^3(\vec{x}-\vec{x}_L) - \delta^3(\vec{x}-\vec{x}_R)]$, expand the double integral (11.21) into four terms:

$$\int d^3x\, d^3y\, \frac{\Delta\rho(\vec{x})\Delta\rho(\vec{y})}{|\vec{x}-\vec{y}|} = M^2\Big[\underbrace{\tfrac{1}{|\vec{x}_L-\vec{x}_L|} + \tfrac{1}{|\vec{x}_R-\vec{x}_R|}}_{\text{self}} - \underbrace{\tfrac{2}{|\vec{x}_L-\vec{x}_R|}}_{\text{cross}}\Big].$$

The two self terms $1/|\vec{x}_A - \vec{x}_A|$ diverge — the Coulomb self-energy of a point source. Once the mass is given its true finite radius $R$ they become a fixed constant $\propto M^2/R$, identical in the two branches and independent of $d$. (We treat this honestly in Chapters 12 and 15: the finite radius sets a floor.) Now keep the signs honest, because the mode integral (11.19) has already fixed them: its integrand is manifestly positive, and translating it to position space with the Newtonian kernel produces exactly the bracket above — *(regularized self)* minus *(cross)*, that is $2M^2[\text{const}(R) - 1/d]$, positive whenever $d > R$. The self piece is a $d$-independent offset. The only part that *changes* when the separation changes — and therefore the only part an interference experiment can probe — is the cross term, of magnitude

$$E_G = \frac{G}{2}\cdot\frac{2M^2}{d} = \frac{GM^2}{d}. \tag{11.22}$$

The factor of 2 from the two identical cross terms cancels the $\frac{1}{2}$ in (11.21). This is the gravitational interaction energy between the two branch positions — the amount by which the difference-field's energy content varies with $d$ — and it carries exactly *one* power of $G$, inherited from the single $G$ in the constraint (11.5).

$$\boxed{E_G = \frac{GM^2}{d}} \qquad \text{[PROVEN] (classical gravity, rigorous).} \tag{11.23}$$

This is the load-bearing result of the chapter, and its status is worth stating without hedging. The energy scale $E_G = GM^2/d$ — the gravitational self-energy of the split, the energy by which the two branches' field configurations differ — is [PROVEN]. It follows from classical gravity by the computation you just watched: the constraint solves for the field, the field of a superposition differs between branches, and the energy of that difference is $GM^2/d$. No conjecture, no modular theory, no operator identity enters (11.23) — only the Poisson equation and an integral you could hand to an undergraduate.

### Worked Example 11.3 — The benchmark energy

Evaluate $E_G$ at the cover's benchmark, $M = 10^{-9}$ kg (one microgram), $d = 10^{-3}$ m (one millimetre):

$$E_G = \frac{GM^2}{d} = \frac{(6.67430\times10^{-11})(10^{-9})^2}{10^{-3}} = \frac{6.67430\times10^{-29}}{10^{-3}} = 6.7\times10^{-26}\ \text{J}. \tag{11.24}$$

The associated time is $\tau = \hbar/E_G = (1.054571817\times10^{-34})/(6.67\times10^{-26}) = 1.58\times10^{-9}$ s $\approx 1.6$ ns — the number on the cover. What we have earned so far: the *energy* $6.7\times10^{-26}$ J is real, derived, and uncoolable. What we have *not* yet earned: that $\hbar/E_G$ is a decoherence *time*. That is a claim about a rate, and we turn to it with deliberate caution.

▶ Run it: `TOOLS/verify-paper-numerics.py` (cross-paper benchmark X: $E_G$, $\tau_{\text{dec}}$)

### 11.5.6 What is proven, and what is not

Let us be scrupulous, because the whole credibility of the enterprise rests on not overclaiming here. Equation (11.19) is exact and [PROVEN] — it is the free-field coherent-state overlap, computed with no approximation beyond linearized gravity. From it, the energy scale $E_G = GM^2/d$ is [PROVEN]. But converting the exponent $\Gamma(t)$ into a decoherence *rate* — a claim that $\Gamma(t)$ grows *linearly* in $t$ with slope $E_G/\hbar$, so that $\mathcal{V}(t) = e^{-E_G t/2\hbar}$ — requires two further things that this chapter does *not* supply:

> *(i) The saturation problem.* The free-field integral (11.19), taken at face value, does *not* grow linearly forever — it *saturates* to a finite constant, as the toy model warned (a single mode cannot decohere permanently, and it turns out the continuum here does not either without further input). The linear-in-$t$ growth that a genuine rate requires comes only from imposing the Hamiltonian constraint on the dynamics, not from the free-field influence functional alone. This is the central subtlety of Chapter 12, and we do not paper over it: *the influence functional gives the energy scale; the constraint gives the rate.*
> *(ii) The modular identity.* Identifying the constraint-enforced growth rate with $E_G/\hbar$ rests on the operator equation $K = 2\pi H_{\text{phys}}$ relating modular time to physical time. That identity is [CONJECTURED] as a full operator equation — [PROVEN] in expectation values and in every solvable model, reduced to one well-posed open construction R1 (Chapter 13), but not established in general.

So the honest ledger at the end of this chapter reads: energy scale $E_G = GM^2/d$, [PROVEN]; the identification of $\hbar/E_G$ with a decoherence time, [CONJECTURED], with the coefficient pinned to the window $C \in [2/\pi, 1]$ and its natural value $C = 1$. Chapter 12 takes up the rate; Chapter 13 takes up the modular identity; Chapter 15 lets the experiment decide what mathematics has not.

## 11.6 Reading the result: three surprises, explained

The functional (11.19) and the scale (11.23) carry three features that seem, at first, to point the wrong way. Each dissolves under inspection, and dissolving them is the best test that we understand what we have computed.

### 11.6.1 Why the phase winds at $E_G/\hbar$

The two branches carry field configurations differing in energy by $E_G$. A relative energy $E_G$ between two branches of a superposition means their relative quantum phase winds at rate $E_G/\hbar$ — the elementary content of $e^{-iEt/\hbar}$. Because $\hbar$ is minuscule, even the tiny energy $6.7 \times 10^{-26}$ J winds a full radian of relative phase in 1.6 ns. This is why a gravitational effect far too weak to move a test mass measurably can nonetheless act fast on *coherence*: coherence is a phase, phases are cheap to wind, and $\hbar$ is small. Weak energy, fast clock.

### 11.6.2 Why wider splits decohere *slower*

Here is the feature that surprises everyone, and it is a genuine prediction, not a sign error. The energy scale is $E_G = GM^2/d$: it *decreases* as the separation $d$ grows. A wider superposition decoheres *slower* ($\tau = \hbar/E_{\mathrm{G}} = \hbar d/GM^2$ grows with $d$), the exact opposite of every mundane decoherence source, all of which get worse as the object spreads out.

The physics is in (11.21). The energy that distinguishes the branches is the gravitational *interaction* energy between the two branch positions, $GM^2/d$. Two masses far apart interact weakly — their mutual gravitational energy is small — so the two field configurations they source differ by a small energy, and the fields are hard to tell apart. Bring the branch positions closer, and the interaction energy $GM^2/d$ grows, the field configurations differ more sharply, and the record is stronger. It is counterintuitive only if one pictures "wider" as "more macroscopic, therefore more classical." The correct picture is the interaction energy: the self-energy of a *wider* split is *less*, so the branches are gravitationally more alike, and coherence survives longer.

**Q 11.4 *"This still feels backward. Surely a mass smeared over a* larger *region is "more classical" and should lose quantum behaviour faster?"***

The intuition conflates two different things. "More classical for larger $M$" is right and is in the formula: heavier masses decohere faster, as $M^2$. But "more classical for larger $d$" is wrong for gravitational decoherence, and the formula is emphatic about it. The quantity that matters is not how spread out the object is but how *different the two gravitational fields are*, and that difference — the interaction energy $GM^2/d$ — shrinks with distance. The signature is a gift to experimenters: gravitational decoherence gets *slower* with $d$, every environmental source gets *faster* with $d$, so a measured rate that *falls* as the split widens is a gravitational fingerprint that nothing mundane can fake (Chapter 15 builds the discriminator on exactly this).

### 11.6.3 Why this is genuine decoherence, not a removable phase

A skeptic's last refuge: perhaps (11.19) is just an overall phase $e^{-iE_{\mathrm{G}}t/\hbar}$ on one branch — a relabelling of the clock, removable by redefining zero energy, no loss of coherence at all. The refuge is closed by the fact that $F(t)$ has a *modulus* less than one, not merely a phase. Look back at the derivation: the modulus came from $|\langle\alpha|\beta\rangle|^2 = \exp(-|\alpha-\beta|^2)$, the genuine reduction in overlap between two *distinct* coherent field states. The branches are entangled with field configurations that are not the same state and not related by a global phase; tracing the field out leaves the qubit's off-diagonal element multiplied by a real number below one. That is loss of visibility — decoherence — not phase.

Whether the loss is *permanent* (a true rate) or *recovers* (a saturating oscillation) is the separate question flagged in §11.5.6 and settled in Chapter 12. But that some coherence is genuinely lost, and not merely rotated, is settled here: the modulus is below one, and the constraint that put it there cannot be gauged away.

## Chapter FAQ

**Q: Is the Wheeler–DeWitt equation $\hat{H}_{\mathrm{tot}}\,|\Psi\rangle = 0$ established physics, or part of the speculation?** The constraint structure is established: any diffeomorphism-invariant theory has a Hamiltonian that is a sum of constraints, and general relativity is such a theory (this is Dirac's analysis, textbook since the 1960s). What is *not* established is the full non-perturbative quantization — writing $\hat{H}_{\mathrm{tot}}$ as a well-defined operator on a well-defined Hilbert space is one of the hard open problems of quantum gravity. But we never need the full operator. We need only its *linearized*, weak-field face, the quantized Poisson equation (11.5), which is under complete control and is all that laboratory physics probes. The single physical input of this chapter is that face.

**Q: You kept saying "relational time." Did you actually use a clock in the derivation, or wave at one?** For the energy scale $E_G$ — the proven result — we used no clock at all: it is a static property of the constraint-entangled state (11.7), the energy of the field difference. The clock (relational time) enters only when we ask for the *rate*, which is precisely the part we have flagged as [CONJECTURED] and deferred to Chapters 12 and 13. So the honest reading is: the chapter's solid deliverable needs no clock; the rate needs one, and the clock is where the open questions live.

**Q: The self-energy integral (11.21) diverged and you threw the divergence away. Is that legitimate?** Yes, and for a reason cleaner than usual renormalization hand-waving. The divergent self terms are *identical* in the two branches — the same point mass, the same infinite self-energy — so they cancel exactly in any quantity that compares branches, which is all decoherence ever does. The finite, physical answer $GM^2/d$ is the branch-to-branch *difference*, and differences are where the divergence is guaranteed to drop out. Give the mass its true finite radius $R$ and the self terms become a finite constant $\propto 1/R$; it sets an extended-body floor (Chapter 12) but never contaminates the leading $GM^2/d$.

**Q: Why does the free-field integral (11.19) give the energy scale correctly but not the rate? That sounds like having it both ways.** It is not having it both ways — it is respecting what each part of the computation controls. The *amplitude* of $\Gamma(t)$ — how big the exponent can get — is fixed by the coherent-state overlap and carries the energy scale $E_G$ faithfully; that part is solid. The *time profile* of $\Gamma(t)$ — whether it grows linearly (a rate) or saturates (no rate) — depends on the dynamics, and the free-field dynamics happens to saturate. Getting the size right while the growth law needs extra input is not a contradiction; it is the statement that the influence functional fixes the scale and the constraint fixes the growth, which is exactly Chapter 12's thesis.

## Exercises

**Exercise 11.1.** Write the electromagnetic Gauss constraint (11.1) and the gravitational Poisson constraint (11.5) side by side. (a) Identify the coupling that plays the role of $4\pi G$ in electromagnetism. (b) For an electron ($e = 1.602 \times 10^{-19}$ C, $\epsilon_0 = 8.854 \times 10^{-12}$ F/m) in a superposition of separation $d = 1\,\mu$m, compute the electromagnetic self-energy of the split, $E_{EM} = e^2/(4\pi\epsilon_0 d)$, and compare it with the gravitational self-energy $GM^2/d$ of the same electron. (c) The two differ by an enormous factor; state it, and explain in one sentence why — despite $E_{EM} \gg E_G$ — the electron nonetheless does not gravitationally decohere and barely electromagnetically decoheres either (the answer is the constraint *type*, §11.1.2, not the energy size).

**Exercise 11.2.** Consider the angular factor $1 - \sin(kd)/(kd)$ from §11.5.4. (a) Expand it for $kd \ll 1$ and confirm the leading behaviour $(kd)^2/6$. (b) Show it tends to 1 as $kd \to \infty$. (c) In one sentence each, say which gravitational modes (long- or short-wavelength relative to $d$) carry the which-path record and which are blind to the split, and connect this to why the mode-difference $|\delta\alpha(\vec{k})|^2$ peaks near $k \sim 1/d$.

**Exercise 11.3.** Using $E_\mathrm{G} = GM^2/d$ and $\tau = \hbar/E_\mathrm{G}$: (a) A nanosphere of density $\rho = 2000\,\mathrm{kg/m^3}$ and radius 50 nm is split by $d = 100$ nm. Compute $M$, then $E_\mathrm{G}$ and $\tau$. (b) Keeping $M$ fixed, by what factor does $\tau$ change if the split is widened to $d = 1\,\mu$m? State whether decoherence speeds up or slows down, and why this is the opposite of environmental decoherence. (c) A referee objects that "bigger superpositions are more classical, so $\tau$ should shrink with $d$." Rebut in two sentences using (11.21).

## Solutions

*Solution (Exercise 11.1).* (a) Electromagnetism has $1/\epsilon_0$ where gravity has $4\pi G$; writing $\nabla \cdot \vec{E} = \rho/\epsilon_0$ against $\nabla^2\Phi = 4\pi G\rho$, the gravitational "charge" is the mass and the coupling is $4\pi G$ (equivalently, $G \leftrightarrow 1/4\pi\epsilon_0$ in the Coulomb-kernel bookkeeping). (b) $E_\mathrm{EM} = e^2/(4\pi\epsilon_0 d) = (1.602 \times 10^{-19})^2/[(1.113 \times 10^{-10})(10^{-6})] = 2.3 \times 10^{-22}$ J. Gravitationally, with $M = 9.109 \times 10^{-31}$ kg, $E_\mathrm{G} = GM^2/d = (6.674 \times 10^{-11})(8.30 \times 10^{-61})/(10^{-6}) = 5.5 \times 10^{-65}$ J. (c) The ratio $E_\mathrm{EM}/E_\mathrm{G} \approx 4 \times 10^{42}$: electromagnetism's self-energy dwarfs gravity's by forty-two orders. Yet the electron does not gravitationally decohere (the scale is negligible) and barely electromagnetically decoheres because the Gauss law is a *spatial* constraint — it entangles the electron with a Coulomb field but, with time external, that entanglement is a recoverable phase, not an irreversible loss (§11.1.2; the full contrast is Chapter 12). The lesson: what decides decoherence is the constraint's *type*, not the energy's *size*.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch11 gauss-analogy energies)`

*Solution (Exercise 11.2).* (a) With $\sin x = x - x^3/6 + \cdots$, $\sin(kd)/(kd) = 1 - (kd)^2/6 + \cdots$, so $1 - \sin(kd)/(kd) \to (kd)^2/6$ for $kd \ll 1$. (b) As $kd \to \infty$, $\sin(kd)/(kd) \to 0$ (bounded numerator, growing denominator), so the factor $\to 1$. (c) Long-wavelength modes ($kd \ll 1$, wavelength $\gg d$) are suppressed by $(kd)^2$ — they cannot resolve the split and carry almost no record; short-wavelength modes ($kd \gg 1$, wavelength $\ll d$) see two clearly distinct sources and carry the record fully. Combined with the $1/k^4$ source falloff in $|\delta\alpha|^2$, the product peaks at $k \sim 1/d$: the modes whose wavelength matches the separation are the ones that best distinguish the branches, which is why $d$ sets the scale of the whole effect.

*Solution (Exercise 11.3).* (a) $M = \rho \cdot \frac{4}{3}\pi R^3 = 2000 \cdot \frac{4}{3}\pi(5 \times 10^{-8})^3 = 2000 \cdot 5.24 \times 10^{-22} = 1.05 \times 10^{-18}$ kg. Then $E_\mathrm{G} = GM^2/d = (6.674 \times 10^{-11})(1.10 \times 10^{-36})/(10^{-7}) = 7.3 \times 10^{-40}$ J, and $\tau = \hbar/E_\mathrm{G} = (1.055 \times 10^{-34})/(7.3 \times 10^{-40}) = 1.4 \times 10^5$ s — about a day and a half. (Gravitational decoherence at this small mass is glacial; the effect is a strong function of $M$.) (b) Widening $d$ tenfold, from 100 nm to 1 $\mu$m, decreases $E_\mathrm{G}$ tenfold and *increases* $\tau$ tenfold, to $\sim 1.4 \times 10^6$ s. Decoherence *slows down*, because $\tau = \hbar d/GM^2 \propto d$ — the opposite of environmental decoherence, whose rate rises as the object spreads. (c) The referee pictures "wider = more macroscopic = more classical," but (11.21) shows the relevant quantity is the gravitational *interaction* energy between the branch positions, $GM^2/d$, which *falls* with $d$. Farther-apart branches source gravitational fields that differ by *less* energy, so they are gravitationally more alike and coherence survives longer.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch11 EG scaling values)`

**What to carry forward**

1. A *constraint* restricts which field configurations a theory permits. Electromagnetism's is the spatial Gauss law $\nabla \cdot \vec{E} = \rho/\epsilon_0$; gravity's is the *Hamiltonian* constraint, the Wheeler–DeWitt equation $\hat{H}_{\text{tot}} |\Psi_{\text{phys}}\rangle = 0$ — the total energy annihilates physical states, making time relational. This single fact is the chapter's only physical input.

2. The constraint *forces* the gravitational field to be entangled with the matter it dresses: a superposed mass must carry a superposition of field configurations, $|\Psi_{\text{phys}}\rangle = (|L\rangle |g_L\rangle + |R\rangle |g_R\rangle)/\sqrt{2}$. Unlike gas or photons, this which-path record is mandatory, ungaugeable, and uncoolable.

3. Feeding the two coherent field states into the influence-functional machine gives the exact decoherence functional $\Gamma(t) = -\ln |F(t)|^2$ (equation (11.19)), whose energy scale is $E_{\text{G}} = GM^2/d = 6.7 \times 10^{-26}$ J at one microgram and one millimetre — [PROVEN]. That $\hbar/E_{\text{G}}$ is a decoherence *time* is [CONJECTURED]: it needs the saturation resolution and the modular identity of Chapters 12 and 13. Wider splits decohere *slower* because $E_{\text{G}} = GM^2/d$ falls with $d$.

# 12

# The Rate

**Synopsis.** Chapter 11 handed us two things: an decoherence exponent $\Gamma(t)$ built from the gravitational field the mass sources, and the constraint $H_{\text{total}} = 0$ that keeps that field welded to the mass. Naively $\Gamma(t)$ should grow forever and give a decoherence rate; it does not. Left to itself the free-field integral *saturates* — it reaches a fixed exponent $\Gamma_{\text{sat}} \approx 0.154$ and stops — and that number is not a rate at all but a one-time overlap, an accident of parameters near $1/2\pi$. The linear-in-time rate comes instead from the constraint, through the modular identity $K = 2\pi H_{\text{phys}}$ of Chapter 9: $\Gamma_{\text{dec}} = C\, E_{\text{G}}/\hbar$, with $C$ boxed into $[2/\pi, 1]$ by two quantum speed limits, natural value $C = 1$. We then face the theory's most-attacked claim head-on — that this rate is $G^1$ while textbook graviton physics gives $G^2$ — and show the two compute different questions (manifestation versus generation), with electromagnetism as the clean control that fails exactly where gravity's constraint does the work.

**You will be able to.** state and explain the Saturation Theorem, and why $\Gamma_{\text{sat}}$ is a static exponent and not a rate; trace the rate to the constraint via $K = 2\pi H_{\text{phys}}$; derive the coefficient window $C \in [2/\pi, 1]$ from Margolus–Levitin and Mandelstam–Tamm; explain the $G^1$-versus-$G^2$ resolution and the $\sim 3 \times 10^{34}$ gap; and state, with the correct status labels, what is proven and what is conjectured.

**Prerequisites.** Chapters 2 (visibility, Gaussian onset), 7 (modular flow, KMS), 9 ($K = 2\pi H_{\text{phys}}$, the QED countermodel), 11 (the decoherence functional and its exponent $\Gamma(t)$, $E_{\text{G}} = GM^2/d$, the Wheeler–DeWitt constraint). No new mathematics is introduced.

## 12.1 Where we stand, and what is missing

By the end of Chapter 11 we had reduced the whole question of gravitational decoherence to a single object. The off-diagonal element of the grain's reduced density matrix — the coherence between the "mass at $L$" and "mass at $R$" branches — is

$$\rho_{LR}(t) \;=\; \tfrac{1}{2}\,\langle \Phi_R(t) | \Phi_L(t) \rangle\,, \qquad |\rho_{LR}(t)| \;=\; \tfrac{1}{2}\, e^{-\Gamma(t)/2}, \tag{12.1}$$

where $|\Phi_{L,R}\rangle$ are the coherent states of the gravitational field sourced by the mass in each branch (Chapter 11), and $\Gamma(t) = -\ln |\,\langle \Phi_R(t)|\Phi_L(t)\rangle\,|^2$ is the decoherence exponent computed there, equation (11.19). This is the visibility law (2.12) of Chapter 2, now with the "environment" being spacetime itself. Written spectrally,

$$\Gamma(t) \;=\; 2\sum_k |\delta\alpha_k|^2 \,(1 - \cos\omega_k t) \;=\; 2\int_0^\infty d\nu\, \varrho(\nu)\,(1 - \cos\nu t), \tag{12.2}$$

a sum over field modes of (coupling to the mass difference)$^2$ times a universal time factor $1 - \cos\omega_k t$ — exactly the mode-sum shape the solvable dephasing model of Worked Example 2.3 predicted, with $g_k$ replaced by the gravitational coupling $\delta\alpha_k$ of mode $k$. Here $\varrho(\nu) \geq 0$ is the spectral density of the distinguishing modes, $\omega_k = c|k|$, and $\varrho$ is peaked at $\nu \sim c/d$, the inverse light-crossing time of the separation.

The energy scale in the problem is settled. The gravitational self-energy that distinguishes the two branches is

$$E_{\rm G} \;=\; \frac{GM^2}{d} \;=\; 6.7\times10^{-26}\ \text{J} \qquad (\text{at } 1\,\mu\text{g},\ 1\,\text{mm}), \tag{12.3}$$

a classical-gravity computation you did in Chapter 11. Its status is [PROVEN]: nothing quantum, nothing conjectural, just the field energy of two Newtonian potentials.

What is *not* settled is the step from an energy scale to a *rate*. To go from (12.3) to a decoherence time $\tau_{\rm dec} \sim \hbar/E_{\rm G}$ we need $\Gamma(t)$ to grow *linearly* in $t$ — $\Gamma(t) \to 2\Gamma_{\rm dec}\,t$, so that $|\rho_{LR}(t)| = \frac{1}{2}\,e^{-\Gamma_{\rm dec}t}$ with a well-defined $\Gamma_{\rm dec}$. The naive expectation is that (12.2) does this automatically. It does not. This chapter is the honest accounting of where the rate really comes from — and it comes from the constraint, not from the free field.

## 12.2 The Saturation Theorem

### 12.2.1 The integral that refuses to grow

Look again at (12.2). Each mode contributes a bounded, oscillating factor:

$$0 \;\leq\; 1-\cos\omega_k t \;\leq\; 2 \qquad \text{for all } t. \tag{12.4}$$

No single mode's contribution grows without limit; it swings between 0 and 2 forever. The only way the *sum* could grow linearly is if there were always fresh modes, at ever-lower frequency, just beginning their first quarter-oscillation — a reservoir of arbitrarily slow modes feeding the integral. Whether that reservoir exists is a question about the spectral density $\varrho(\nu)$ at low frequency, and for the free gravitational field the answer is: it does not.

**Theorem 12.1** (Saturation of the free-field exponent)**.** *For the free graviton bath sourced by a mass at two positions, the exponent* (12.2) *converges to a finite constant as $t \to \infty$:*

$$\Gamma(t) \;\xrightarrow{t\to\infty}\; \Gamma_{\rm sat} \;<\; \infty. \tag{12.5}$$

*There is no sustained linear growth. The limit is a finite, one-time overlap reduction, reached once the field modes have equilibrated to the source.* [PROVEN]

*Derivation.* Two ingredients. *First*, boundedness gives an immediate ceiling: from (12.4),

$$\Gamma(t) \;\leq\; 4\int_0^\infty d\nu\,\varrho(\nu) \;=\; 4\,\|\delta\alpha\|^2,$$

so the exponent can never exceed four times the total squared norm of the field difference, *provided* that norm is finite. *Second*, the oscillatory part washes out. By the Riemann–Lebesgue lemma, for any integrable $\varrho$,

$$\int_0^\infty d\nu\,\varrho(\nu)\cos\nu t \;\xrightarrow{t\to\infty}\; 0,$$

so $\Gamma(t) \to 2\int d\nu\, \varrho(\nu) = 2\|\delta\alpha\|^2 \equiv \Gamma_{\rm sat}$ (using the time-averaged value 1 for $1-\cos$, not its peak 2). The one thing left to check is that $\|\delta\alpha\|^2 < \infty$, i.e. that $\varrho$ is integrable. This is where the physics of gravity enters, and it is the subject of the next subsection.

### 12.2.2 Why the reservoir is empty: the super-Ohmic bath

The convergence hinges entirely on how $\varrho(\nu)$ behaves as $\nu \to 0$ — the slow modes. For the free gravitational field the spectral density *vanishes* there:

$$\varrho(\nu) \;\sim\; \nu\, e^{-\nu/\nu_c} \quad (\text{low } \nu), \qquad \text{so} \qquad \varrho(0^+) = 0. \tag{12.6}$$

A bath whose spectral density vanishes at zero frequency is called *super-Ohmic*. Physically (12.6) says: long-wavelength gravitational modes barely couple to the difference between two positions separated by $d$. A mode of wavelength $\lambda \gg d$ cannot resolve which of two sites $d$ apart the mass sits at — it sees essentially one source either way — so its distinguishing coupling $|\delta\alpha_k|^2$ is suppressed as $(kd)^2 \sim (\nu d/c)^2$. There is no reservoir of arbitrarily slow modes carrying which-path information; the slow modes are blind to the separation.

With $\varrho(0^+) = 0$ the integral $\int d\nu\, \varrho(\nu)$ converges at the low end, $\|\delta\alpha\|^2 < \infty$, and Theorem 12.1 holds. The exponent climbs, levels off, and sits.

> **The three regimes of $\Gamma(t)$, once and for all.** The free-field exponent (12.2) passes through three phases, set by the light-crossing time $d/c$:
>
> - *Zeno* ($t \ll d/c$): no mode has completed even a quarter turn; $1 - \cos\nu t \approx (\nu t)^2/2$, so $\Gamma \approx M_2 t^2$ grows *quadratically*. Zero initial slope.
> - *Crossover* ($t \sim d/c$): the dominant modes ($\nu \sim c/d$) are turning over.
> - *Saturated* ($t \gg d/c$): all modes have equilibrated; $\Gamma \to \Gamma_{\rm sat}$, oscillating and decaying toward the plateau as a power law, *not* exponentially.
>
> Nowhere is there a stretch of sustained linear growth. The free field gives an energy scale (through $\Gamma_{\rm sat}$) but no rate.

### 12.2.3 The number, and what it is not

Evaluating $\Gamma_{\rm sat} = \|\delta\alpha\|^2$ for the free graviton bath with a Planck-scale ultraviolet cutoff gives a logarithmically dominated integral. The result, to leading-log accuracy, is

$$\Gamma_{\rm sat} \;=\; \xi\, \ln\frac{d}{\ell_{\rm P}}, \qquad \xi \;\equiv\; \frac{GM^2}{\hbar c} \;=\; \left(\frac{M}{m_{\rm P}}\right)^2, \tag{12.7}$$

where $\xi$ is the dimensionless gravitational coupling — the ratio of the gravitational self-energy to the Compton-scale quantum, equivalently $(M/m_{\rm P})^2$ with $m_{\rm P}$ the Planck mass of Worked Example 1.2.

**Worked Example 12.1 — Evaluating $\Gamma_{\text{sat}}$ at the benchmark**

For $M = 1\,\mu\text{g} = 10^{-9}$ kg and $d = 1$ mm, with $\ell_\text{P} = 1.616 \times 10^{-35}$ m:

$$\xi = \frac{GM^2}{\hbar c} = \frac{(6.674 \times 10^{-11})(10^{-18})}{(1.0546 \times 10^{-34})(2.998 \times 10^{8})} = 2.11 \times 10^{-3}, \tag{12.8}$$

$$\ln\frac{d}{\ell_\text{P}} = \ln\frac{10^{-3}}{1.616 \times 10^{-35}} = \ln(6.19 \times 10^{31}) = 73.2,$$

$$\Gamma_\text{sat} = (2.11 \times 10^{-3})(73.2) = 0.154.$$

The coherence reduction it implies is a *one-time* $|\langle\Phi_L|\Phi_R\rangle|^2 = e^{-\Gamma_\text{sat}} = e^{-0.154} = 0.86$: a 14% loss of contrast, delivered once the field has settled, and then no more. Not decoherence in the sense of an ongoing decay — a static overlap.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch12 Gamma_sat = xi ln(d/l_P))`

Three things about (12.7) must be nailed down before we go on, because misreading this number is the single most common error made about the theory.

*It is not a rate.* $\Gamma_\text{sat}$ is dimensionless. It is the *total* overlap exponent after the field equilibrates, a coherence loss, not a coherence loss *per unit time*. It has no "per second" in it. You cannot write $|\rho_{LR}(t)| \sim e^{-\Gamma_\text{sat} t}$ — that is dimensionally nonsense.

*It depends on $M^2$ and on a cutoff.* Through $\xi = (M/m_\text{P})^2$ the saturation exponent scales as $M^2$, and through $\ln(d/\ell_\text{P})$ it depends logarithmically on the Planck cutoff. It is not a universal constant. Change the mass by a factor of 3 and it moves by a factor of 9; over plausible $(M, d)$ it ranges from $\sim 0.0015$ to $\sim 15$.

*Its closeness to $1/2\pi$ is a coincidence.* At the benchmark, $\Gamma_\text{sat} = 0.154$ sits within 3% of $1/(2\pi) = 0.159$. This is happenstance: the smooth function $\xi \ln(d/\ell_\text{P})$ happens to pass near 0.159 somewhere around $(1\,\mu\text{g}, \sim 1\,\text{cm})$, and a different mode-counting convention gives 0.098 instead (a $2/\pi$ factor the leading-log estimate drops). Nothing forces $\Gamma_\text{sat}$ to equal any pure number.

**Q 12.1** ***"You just spent a page computing $\Gamma_\text{sat} \approx 0.154$. Isn't that the decoherence rate? It looks like the answer."***

It is emphatically not, and this is the trap. Three separate reasons, any one fatal. *(i) Dimensions.* A rate carries units of inverse time; $\Gamma_\text{sat}$ is a pure number. It cannot be the coefficient of $e^{-\Gamma t}$. *(ii) It saturates.* By Theorem 12.1 $\Gamma(t)$ climbs to 0.154 and *stops*. If this governed decoherence, a $1\,\mu$g grain split by a millimetre would lose 14% of its coherence, once, and then keep the remaining 86% forever — which is exactly the objection a careful critic raises, and exactly what the free field alone predicts. *(iii) It is not even the right category.* It is a *static overlap exponent* — how distinguishable the two field configurations are once equilibrated — whereas a rate is an *ongoing* loss. Conflating the two is a category error. The free field, honestly computed, gives no rapid decoherence at all. Something else must supply the linear-in-$t$ growth, and finding it is the rest of this chapter. Notice we have just conceded the critic's point in full before answering it: that is the honest order of business.

## 12.3 The modular route to the rate

### 12.3.1 What the constraint changes

The free-field calculation of §12.2 treated the gravitational field as an independent bath that "switches on" at $t = 0$: at $t = 0$ the field is in its vacuum, and it then relaxes toward the source. That picture is unphysical for gravity, and the reason is the Wheeler–DeWitt constraint of Chapter 11,

$$H_{\text{total}} \;=\; H_{\text{matter}} + H_{\text{grav}} \;=\; 0. \tag{12.9}$$

The constraint is not a dynamical process; it is a condition on the space of allowed states, enforced at every instant. There is no moment when the field is decoupled from the matter and then couples. The field is *always* welded to the mass configuration — in the Newtonian limit, $\nabla^2\Phi = 4\pi G\rho$ holds as an operator equation at all times. The physical state is never the product $(|L\rangle + |R\rangle) \otimes |0\rangle$; it is the constraint-entangled

$$|\Psi_{\text{phys}}\rangle \;=\; \tfrac{1}{\sqrt{2}}\big(|L\rangle\,|\Phi_L\rangle + |R\rangle\,|\Phi_R\rangle\big), \tag{12.10}$$

with the field already carrying which-path information at $t = 0$. The free-field integral computed how fast this entanglement *builds from vacuum*; the physical question is how fast the already-present entanglement *manifests as a decaying coherence*. Those are different questions, and the second one is answered by the modular machinery.

### 12.3.2 From the modular clock to a linear phase

Because $H_{\text{total}} = 0$, the physical state is *timeless*: there is no external clock (Chapter 11). Time must emerge from within, from the entanglement between the matter and its own gravitational field — the Page–Wootters mechanism. The generator of that emergent, internal time flow is the modular Hamiltonian $K$ of Chapters 7–9, and the whole theory turns on one identification, Axiom III made quantitative:

$$K \;=\; 2\pi\, H_{\text{phys}} \;+\; O(G^2). \tag{12.11}$$

This says: the modular flow that the algebra of accessible observables generates *is* physical time evolution, up to the Bisognano–Wichmann factor of $2\pi$ (the same $2\pi$ that set the Unruh temperature in Chapter 4). Under (12.11) the two branches of (12.10), separated by the physical energy gap $E_G$, are separated in *modular* energy by $\Delta K = 2\pi E_G$. The modular flow advances their relative phase at rate $\Delta K/\hbar = 2\pi E_G/\hbar$ per unit modular time.

A deterministic phase, by itself, is not decoherence — it is the reversible oscillation we met for electromagnetism in Chapter 9. What converts it into an irreversible decay is that the modular clock is a *quantum* clock: the emergent time it defines has an irreducible uncertainty, because $K$ and the modular time $T$ are conjugate. The KMS condition of Chapter 7 — the statement that the modular flow is thermal, at unit temperature in modular time or, equivalently, at $\beta_{\text{mod}} = 2\pi$ in the boost/proper-time parametrization the rest of the book uses — is precisely the statement that this clock carries thermal fluctuations. Feeding the constraint-supplied low-frequency weight (the $1/\nu^2$ tail that the *free* field lacked) into (12.2) produces the one cosine transform that is exactly linear,

$$\int_0^\infty \frac{1-\cos\nu t}{\nu^2}\,d\nu \;=\; \frac{\pi}{2}\,t, \tag{12.12}$$

so that $\Gamma(t) \to 2\Gamma_{\rm dec}\, t$ and the coherence decays exponentially:

$$|\rho_{LR}(t)| \;=\; \tfrac{1}{2}\, e^{-\Gamma_{\rm dec}\, t}, \qquad \Gamma_{\rm dec} \;=\; C\,\frac{E_{\rm G}}{\hbar} \;=\; C\,\frac{GM^2}{\hbar d}. \tag{12.13}$$

The energy scale $E_{\rm G}$ came from the field; the *rate* came from the constraint, through the identification (12.11) of modular time with physical time. The single power of $G$ enters through the constraint $\nabla^2\Phi = 4\pi G\rho$, which is linear in $G$ — a point we return to in §12.5.

### 12.3.3 The two statuses, kept apart

Everything in (12.13) rests on (12.11), and here we must be scrupulous, because this is the chapter where the theory's central conjecture becomes load-bearing.

**Two claims, two statuses — do not merge them.**

*The energy scale.* $E_{\rm G} = GM^2/d$ is the gravitational self-energy difference between the branches. [PROVEN] — classical gravity, Chapter 11, no conjecture.

*The rate.* $\Gamma_{\rm dec} = C\,E_{\rm G}/\hbar$ requires the operator identity $K = 2\pi H_{\rm phys}$. That identity is [CONJECTURED] as a full operator equation: it is [PROVEN] in expectation values and in every solvable model checked, but open as an operator statement, and the whole residue is isolated in one sharply-posed construction, R1 (Chapter 13). Therefore the rate is [CONJECTURED]: strongly motivated, reduced to a single open problem, but not proved.

This is not hedging for its own sake. The $2\pi$ of (12.11) is a theorem in the vacuum (Bisognano–Wichmann, Chapter 9); transporting it to the matter-loaded, constraint-satisfying state (12.10) is what remains open. An experiment that measured $\Gamma_{\rm dec}$ would test (12.11) at the operator level directly — which is why Chapter 15 matters as much as this one.

***Q 12.2 "The "quantum clock with timing uncertainty" story sounds like hand-waving dressed up. Is that really the derivation?"***

No — and the distinction matters. The clock-uncertainty picture is a *demoted heuristic*: it is a faithful way to *picture* why a timeless constraint turns a phase into a decay, but it is not the load-bearing calculation. The load-bearing object is the influence functional (12.2) with the constraint-supplied spectral weight, whose linear kernel (12.12) gives the exponential decay directly from the branch-energy gap read *linearly* (the reactive part of the influence phase), not from timing noise. In an exactly-solvable independent-boson model with a Hamiltonian-type constraint, one can watch the *same* reduced dynamics give a $G^1$-analogue rate for the constrained (dressed) initial state and a $G^2$-analogue rate for the unconstrained (bare) state — the order difference is exactly the reactive-versus-dissipative split of the influence phase, activated by the constraint. The clock language is a picture of that mechanism, not a substitute for it. Where the two appear to differ, the influence-functional statement is canonical.

## 12.4 The coefficient window $C \in [2/\pi, 1]$

The rate (12.13) carries an order-one coefficient $C$. We can do better than "order one": $C$ is boxed between two quantum speed limits, and the box is tight. Both ends are honest bounds; the natural value sits at the ceiling.

### 12.4.1 The floor: Margolus–Levitin

How fast can a quantum state become distinguishable from itself? The Margolus–Levitin theorem answers: a state of mean energy $E$ above its ground state cannot evolve to an orthogonal state in less time than

$$\tau_\perp \;\geq\; \frac{\pi\hbar}{2E}. \tag{12.14}$$

This is a speed limit — a *lower* bound on the time to orthogonalize, hence an *upper* bound on the orthogonalization rate. Apply it to our two branches, separated by the energy gap $E = E_\mathrm{G}$. The fastest they can become strictly distinguishable is $\tau_\perp = \pi\hbar/(2E_\mathrm{G})$, i.e. a rate

$$\Gamma_\perp \;\leq\; \frac{2}{\pi}\,\frac{E_\mathrm{G}}{\hbar} \qquad \Longrightarrow \qquad C_\perp \;=\; \frac{2}{\pi} \;=\; 0.637. \tag{12.15}$$

The branches cannot become strictly orthogonal faster than $C = 2/\pi$ allows. This fixes the floor.

### 12.4.2 The ceiling: Markovian dephasing

At the other end, the constrained dynamics in its Markovian (late-time, $t \gg d/c$) regime is an ordinary dephasing channel. The effective master equation for the matter is Lindblad,

$$\frac{d\rho}{dt} \;=\; -\frac{i}{\hbar}[H_\mathrm{eff}, \rho] \;-\; \frac{\Gamma}{2}\,[X, [X, \rho]], \qquad \Gamma = \frac{E_\mathrm{G}}{\hbar}, \tag{12.16}$$

with $X$ the which-path (position) distinguisher. The natural dephasing rate here is $\Gamma = E_\mathrm{G}/\hbar$, i.e. $C = 1$. This is the Mandelstam–Tamm scale (the energy variance across a sharp two-branch gap is $\Delta E \sim E_\mathrm{G}$) and it is exactly the Diósi–Penrose value — the rate that heuristic argument reached in the 1980s and that this framework now recovers from the constraint. It is the physically relevant rate for the fringe-visibility decay an interferometer measures.

### 12.4.3 The window, and its honest status

Both anchors flow from the *same* modular gap $E_\mathrm{G}$ under the single normalization $K = 2\pi H_\mathrm{phys}$; they differ only in which operational quantity one reads off — strict orthogonalization speed $(2/\pi)$ versus visibility-decay rate (1). The derivable statement is therefore

$$C \;\in\; \left[\tfrac{2}{\pi},\, 1\right] \;\approx\; [0.637,\, 1.000], \qquad \text{natural value } C = 1. \tag{12.17}$$

a window of width $\pi/2 \approx 1.6$. Its status: the window is a rigorous *bound* given the framework; $C$ is [CONJECTURED] to sit at the natural value 1 but is bounded, not uniquely pinned — the residual freedom is the clock-model normalization, an open point that Chapter 13 inherits. We do not claim a unique $C$; we claim a box and a preferred corner.

**Worked Example 12.2 — The decoherence time across the window**

The decoherence time is $\tau_{\text{dec}} = \hbar/(C\,E_{\text{G}})$. With $E_{\text{G}} = 6.7 \times 10^{-26}$ J at the benchmark:

$$\text{at } C = 1: \quad \tau_{\text{dec}} = \frac{\hbar}{E_{\text{G}}} = \frac{1.0546 \times 10^{-34}}{6.674 \times 10^{-26}} = 1.58 \times 10^{-9}\ \text{s} \approx 1.6\ \text{ns},$$
$$\text{at } C = \tfrac{2}{\pi}: \quad \tau_{\text{dec}} = \frac{\pi}{2}\,\frac{\hbar}{E_{\text{G}}} = \frac{\pi}{2}(1.58\ \text{ns}) = 2.48 \times 10^{-9}\ \text{s} \approx 2.5\ \text{ns}.$$

The whole coefficient uncertainty amounts to the difference between 1.6 and 2.5 nanoseconds — a factor $\pi/2$, not a factor of ten. The prediction is sharp enough to test.

▶ Run it: `TOOLS/verify-paper-numerics.py` `(cross-paper benchmark X: tau_dec at C=1; section N: C=2/pi)`

**Q 12.3** ***"So is $C = 1$ or not? You keep saying "natural value" but also "not pinned." Which is it?"***

Both statements are true and they are not in tension. What is *derived* is the box $[2/\pi, 1]$: the branches cannot orthogonalize faster than $2/\pi$ (Margolus–Levitin, a theorem), and the Markovian dephasing rate is 1 (the Diósi–Penrose value, the natural Lindblad normalization). What is *not* derived is a unique number inside the box, because the two ends are two different *observables* — strict orthogonality versus fringe-contrast decay — and which one a given experiment reports is a definitional choice about what is measured. For an interferometer measuring visibility, the relevant coefficient is $C = 1$; that is the natural and most defensible single value, and it is what we quote. But we do not dress it as a theorem. Pinning $C$ to 1 as a proved uniqueness requires closing the clock-model normalization (Chapter 13). Until then: box proved, corner preferred, uniqueness open. That is the whole of the honest statement.

## 12.5 $G^1$ versus $G^2$: the central resolution

We now confront the objection that has been raised against this theory more than any other. Standard perturbative quantum field theory — graviton emission and exchange, computed with Feynman diagrams — gives a gravitational decoherence rate scaling as $G^2$, not $G^1$. That calculation (Anastopoulos–Hu, Blencowe) is rigorous and correct. Our rate (12.13) scales as $G^1$. How can both be right?

They are both right because they compute *different physical quantities*. This is the crux, and it is worth stating plainly before any algebra.

### 12.5.1 Two questions, not one

*The $G^2$ calculation* starts from an *unentangled* initial state: a bare mass in superposition, tensored with the graviton vacuum, $(|L\rangle + |R\rangle) \otimes |0_{\text{grav}}\rangle$. The mass carries no gravitational dressing. The interaction $H_{\text{int}} \sim \sqrt{G}\,T\,h$ then *generates* entanglement over time through graviton exchange. The decoherence rate follows from the double commutator $\sim \langle[H_{\text{int}}, [H_{\text{int}}, \rho]]\rangle$, which brings in $H_{\text{int}}$ *twice* — two vertices, two factors of $\sqrt{G}$, hence $G^1$ from the coupling; the graviton propagator and the environment trace supply the rest, and the net rate is $G^2$. This answers: *at* *what rate does a bare mass become entangled with the graviton field?* It is the rate of entanglement *generation*.

*The $G^1$ calculation* starts from the *constraint-satisfying* state (12.10): the mass already dressed by its field, already entangled, because $\nabla^2\Phi = 4\pi G\rho$ leaves no choice. No graviton needs to be emitted; the which-path record already exists at $t = 0$. The rate follows from the single power of $G$ in the constraint. This answers: *given that the mass is already entangled with its field, at what rate does that entanglement manifest as decoherence?* It is the rate of entanglement *manifestation*.

> **The resolution in one line.** $G^2$ is the rate of entanglement *generation* from an unphysical, undressed initial state; $G^1$ is the rate of entanglement *manifestation* for the physical, constraint-satisfying state. The physical state is already entangled, so no graviton emission is needed — and the perturbative calculation, starting from the wrong (constraint-violating) initial state, simply does not see that pre-existing entanglement. Both are correct; they answer different questions; nature is in the constraint-satisfying state.

### 12.5.2 The hydrogen-atom analogy

The structure is familiar from a computation every physicist has done. The binding energy of hydrogen can be found two ways. *Perturbatively*, summing photon-exchange diagrams order by order in $\alpha$. Or *non-perturbatively*, solving the Schrödinger equation with the Coulomb potential — which is itself the *solution* of the electromagnetic constraint (Gauss's law). The Coulomb potential resums all the exchanges in one stroke; you never draw a diagram. The constraint-based gravitational calculation is the analogue of solving with the Coulomb potential: it uses the constraint solution directly rather than building the field up diagram by diagram.

There is one disanalogy, and it is the whole point. For hydrogen the two methods *agree*, because binding energy is an eigenvalue — a property of the Hamiltonian alone, independent of the initial state. Decoherence is not an eigenvalue; it is a property of the time evolution of a *specific* state. Start from different initial states — bare-and-unentangled versus dressed-and-entangled — and you get different decoherence rates from the same Hamiltonian. That is why the perturbative and constraint-based answers *disagree* here while they *agree* for hydrogen. The initial state is irrelevant to an eigenvalue and decisive for a decay.

### 12.5.3 How big is the gap?

The two rates differ by an enormous factor, so there is no ambiguity about which dominates in the laboratory.

**Worked Example 12.3 — The $G^2/G^1$ ratio and the $G^2$ timescale**

The suppression of the $G^2$ rate relative to the $G^1$ rate at the benchmark is the dimensionless gap

$$\frac{\Gamma_{G^1}}{\Gamma_{G^2}} = \left(\frac{m_{\mathrm{P}}}{M}\right)^2 \frac{d}{\ell_{\mathrm{P}}}. \tag{12.18}$$

For $M = 1\,\mu$g, $d = 1$ mm, with $m_P = 2.176 \times 10^{-8}$ kg:

$$\left(\frac{m_P}{M}\right)^2 = \left(\frac{2.176 \times 10^{-8}}{10^{-9}}\right)^2 = 4.74 \times 10^2, \qquad \frac{d}{\ell_P} = \frac{10^{-3}}{1.616 \times 10^{-35}} = 6.19 \times 10^{31},$$

$$\frac{\Gamma_{G^1}}{\Gamma_{G^2}} = (4.74 \times 10^2)(6.19 \times 10^{31}) = 2.9 \times 10^{34}.$$

So the $G^2$ process is about $3 \times 10^{34}$ times slower than the $G^1$ process. Since $G^1$ gives $\tau_{\text{dec}} = 1.6$ ns, the $G^2$ timescale is

$$\tau_{G^2} = (1.6 \times 10^{-9}\ \text{s})(2.9 \times 10^{34}) = 4.6 \times 10^{25}\ \text{s} = 1.5 \times 10^{18}\ \text{yr},$$

a hundred million times the age of the universe. The perturbative graviton-emission decoherence is, for any laboratory mass, utterly unobservable — which is precisely why Chapter 1 could dismiss quantized-metric effects as predicting "nothing observable" at the bench.

▶ Run it: `TOOLS/verify-paper-numerics.py (cross-paper benchmark X: G2/G1 gap ≈3e34; tau_G2 ≈1.5e18 yr)`

### 12.5.4 The QED control: why gravity, and only gravity

The obvious challenge to the $G^1$ mechanism is electromagnetism. An electron in superposition is dressed by its Coulomb field exactly as a mass is dressed by its gravitational field: $|\Psi\rangle = (|L\rangle\,|E_L\rangle + |R\rangle\,|E_R\rangle)/\sqrt{2}$, forced by Gauss's law. If constraint-entanglement caused decoherence at rate $E_{\text{self}}/\hbar$, electrons in double-slit experiments would decohere in picoseconds. They do not. So either the mechanism is wrong, or electromagnetism differs from gravity in exactly the place that matters. It is the latter, and the difference is the cleanest evidence that the constraint — not a computational trick — does the work.

Run the two computations side by side, as in Chapter 9. Through the overlap they are *identical*: both give a coherence $\gamma(t) = \langle E_R|e^{-iHt/\hbar}|E_L\rangle$. They diverge only at the last step, when the rate is extracted, and the divergence is structural:

| | **Gravity** | **Electromagnetism** |
|---|---|---|
| Constraint type | Hamiltonian: $H_{\text{total}} = 0$ | Spatial: $\nabla\cdot E = \rho/\varepsilon_0$ |
| Generates | time reparametrizations | $U(1)$ gauge transformations |
| Time | emergent (from the constraint) | background parameter |
| Total $H$ constrained? | yes ($H = 0$) | no ($H_{\text{QED}}$ free) |
| Modular = physical time? | yes (via $H = 0$) | no |
| Constraint entanglement gives | *decay* ($G^1$) | *phase shift only* |

The chain of §12.3 needs a *Hamiltonian* constraint. For gravity, $H_{\text{total}} = 0$ makes time emergent, forces the modular clock, and $K = 2\pi H_{\text{phys}}$ turns the branch phase into an irreversible decay:

$$\gamma_{\text{grav}}(t) \sim e^{-E_G t/\hbar} \qquad \text{(decay)}. \tag{12.19}$$

For electromagnetism the constraint is only *spatial* (Gauss's law generates internal gauge rotations, not time reparametrizations). $H_{\text{QED}}$ is not constrained to vanish; time is an ordinary background parameter; there is no emergent clock and no modular–physical identification. The Coulomb entanglement is real, but it produces a *reversible* phase:

$$\gamma_{\mathrm{EM}}(t) \;\sim\; e^{-iE_{\mathrm{self}}t/\hbar} \qquad \text{(unitary phase, } |\gamma_{\mathrm{EM}}| = 1). \tag{12.20}$$

The magnitude of the coherence is untouched; the fringes shift but do not fade. No universal $G^1$-type decoherence for electromagnetism.

**The control that isolates the mechanism.** Gravity and QED are identical through the overlap and diverge only at rate extraction, where gravity's Hamiltonian constraint ($H_{\mathrm{total}} = 0$) supplies an emergent clock and QED's spatial constraint does not. This is as clean a control as the subject offers: the *only* difference between the two calculations is the constraint type, and flipping it turns a decay into a phase. If the effect were a computational artifact it would appear in QED too. It does not. The constraint is doing the physics.

***Q 12.4 "Granting that the two calculations answer different questions — why is the $G^1$ "manifestation" rate the physical one? Why not the $G^2$ "generation" rate?"***

Because nature is in the constraint-satisfying state, not the bare one. The Wheeler–DeWitt constraint $\nabla^2\Phi = 4\pi G\rho$ holds in the world; physical states *are* dressed by their gravitational fields, exactly as physical charged states in QED are dressed by their Coulomb fields (a state violating Gauss's law is unphysical, and nobody computes with one). The $G^2$ calculation asks how fast an *unphysical*, undressed state would acquire its dressing — a legitimate question with a correct answer, but not the question about a real experiment. The real grain is already entangled with its field the instant it is in superposition; the physical rate is the rate at which *that* entanglement decoheres it, which is $G^1$. The status is what keeps us honest: this reasoning is sound, but it routes through $K = 2\pi H_{\mathrm{phys}}$, so the $G^1$ *rate* remains [CONJECTURED] while the $G^2$ generation rate is a settled perturbative fact.

## 12.6 The decay form and the Zeno crossover

We have a rate, (12.13), valid in the Markovian regime $t \gg d/c$. But the full decay is not a pure exponential, and the departure at short times is an experimental signature, so we should say what the shape actually is.

Return to the influence functional (12.2), now with the constraint-supplied weight. Two regimes, separated by the light-crossing time $d/c$:

*Zeno regime* ($t \ll d/c$). Before any gravitational mode has propagated across the separation, no mode can carry which-path information at the Markovian rate. Expanding $1 - \cos\nu t \approx (\nu t)^2/2$,

$$\Gamma(t) \;\approx\; M_2\, t^2, \qquad M_2 \;=\; \int d\nu\, \varrho(\nu)\, \nu^2 \;=\; \frac{\langle \Delta H^2\rangle}{\hbar^2}, \tag{12.21}$$

so $|\rho_{LR}(t)| = \frac{1}{2}\, e^{-M_2 t^2/2}$ — *Gaussian*, with *zero initial slope*. This is the universal quantum Zeno onset of Worked Example 2.3: nothing coupled to a finite-energy environment can decay linearly at $t = 0$, because the global evolution is unitary.

*Markovian regime* ($t \gg d/c$). The linear kernel (12.12) takes over and $|\rho_{LR}(t)| = \frac{1}{2}\, e^{-\Gamma_{\mathrm{dec}} t}$

— exponential, the Diósi–Penrose form as a *limit* of the constrained dynamics rather than an ansatz.

The crossover time follows from equating the two asymptotes, $\frac{1}{2}M_2t^2 = \Gamma_{\mathrm{dec}}t$:

$$\tau_{\mathrm{corr}} = \frac{2\Gamma_{\mathrm{dec}}}{M_2} \sim \frac{1}{\nu_{\mathrm{peak}}} \sim \frac{d}{c} = 3.3 \text{ ps} \quad (\text{at } d = 1 \text{ mm}), \tag{12.22}$$

the gravitational light-crossing time. Since $\tau_{\mathrm{corr}}/\tau_{\mathrm{dec}} = (d/c)\Gamma_{\mathrm{dec}} \approx 2 \times 10^{-3}$, the Zeno window is about a thousand times shorter than the decoherence time: for ordinary interferometry the measured decay is effectively exponential, and the Gaussian onset matters only for sub-picosecond-resolved protocols.

But the onset is not a mere technicality — it is a *discriminator*. A pure collapse model of the Diósi–Penrose type posits exponential decay from $t = 0$, with a nonzero initial slope. QGC predicts a zero-initial-slope, Zeno-protected window of duration $\sim d/c$ that such models lack. The clean test, from Chapter 2, is the ratio built from the decay exponent: a Gaussian onset $\Gamma(t) \propto t^2$ satisfies $\Gamma(2t)/\Gamma(t) = 4$, whereas a pure exponential gives $\Gamma(2t)/\Gamma(t) = 2$. Measuring that ratio in the onset separates QGC from collapse models. Chapter 15 builds the protocol.

## 12.7 Five arguments, one input

It is honest to close by naming what the whole edifice rests on. The $G^1$ rate has been reached, across the theory's development, by several apparently independent routes: the coherent-state modular argument (§12.3); the Page–Wootters clock picture (§12.3, now a demoted heuristic); the constrained influence functional (§12.6); the Type III → Type II algebraic transition of Chapter 9; and the hydrogen-atom non-perturbative analogy (§12.5). Five arguments, converging on the same answer.

But convergence can be an illusion if the arguments share a premise, and here they do. Every one of them uses the Wheeler–DeWitt/Hamiltonian constraint $H_{\mathrm{total}} = 0$ (Chapter 11) as its single physical input — the same constraint, entering in five costumes. The coherent-state argument needs it to identify $K$ with $H_{\mathrm{phys}}$; the clock picture needs it to make time emergent; the influence functional needs it to supply the $1/\nu^2$ IR weight that the free field lacks; the algebraic transition needs it to convert the algebra type; the analogy needs it to furnish the constraint solution that replaces the diagram sum. Strip the constraint out and all five collapse together — as the QED control of §12.5.4 shows, where its absence turns every one of them into a mere phase.

So the theory does not have five independent pillars under the rate; it has one, used five ways. We state this candidly because it is the theory's main epistemic caveat — and because it is a feature, not a buried flaw. A single, sharply-identified physical input is exactly what an experiment can attack: the rate is $\Gamma_{\mathrm{dec}} = C\,E_{\mathrm{G}}/\hbar$, $C \in [2/\pi, 1]$, natural $C = 1$, [CONJECTURED] through the one operator equation $K = 2\pi H_{\mathrm{phys}}$ whose residue is the open construction R1 of Chapter 13. If the constraint does what the theory says, the grain decoheres in 1.6 nanoseconds; if a levitated-nanosphere interferometer holds coherence a hundred times longer, the constraint does not, and the theory is dead. Chapter 13 takes the mathematics as far as it goes; Chapter 15 hands the rest to the laboratory.

## Chapter FAQ

**Q: If the free-field integral saturates at 14%, doesn't that mean gravity barely decoheres anything — contradicting the 1.6 ns headline?** The two numbers answer different questions. $\Gamma_{\text{sat}} = 0.154$ is the static overlap once the field equilibrates *if* the field were an independent bath switched on at $t = 0$; it is a one-time reduction and, honestly, if that were the whole story gravitational decoherence would be a minor 14% effect. The 1.6 ns comes from the *constrained* dynamics, where the field is welded to the mass at all times and the modular clock converts the branch-energy gap into a linear-in-$t$ decay. The saturation result is not wrong; it is the answer to a question (free-bath equilibration) that is not the physical one (constraint-enforced decoherence). Keeping them separate is the entire content of §12.2–12.3.

**Q: Real graviton emission is a $G^2$ effect and is textbook. Isn't a $G^1$ rate just wrong?** The $G^2$ result is correct — for the rate at which a *bare*, undressed mass becomes entangled with the graviton vacuum (entanglement generation). It requires two interaction vertices, hence two powers of $\sqrt{G}$. The $G^1$ rate is for a different object: the physical state is already dressed and entangled by the constraint, and $G^1$ is the rate at which *that* pre-existing entanglement decoheres (entanglement manifestation), carrying the single power of $G$ in $\nabla^2\Phi = 4\pi G\rho$. The two differ by $\sim 3 \times 10^{34}$ at the benchmark. Nature is in the dressed state, so $G^1$ is the physical rate — but it rests on $K = 2\pi H_{\text{phys}}$ and so stays [CONJECTURED].

**Q: Why doesn't the same argument decohere every charged particle through its Coulomb field?** Because electromagnetism has only a *spatial* constraint (Gauss's law), not a Hamiltonian one. Gauss's law generates internal $U(1)$ rotations, leaves $H_{\text{QED}}$ unconstrained, and keeps time a background parameter — so there is no emergent clock and no modular–physical identification. The Coulomb entanglement produces a reversible *phase*, not a decay: fringes shift but do not fade. Only a Hamiltonian constraint ($H_{\text{total}} = 0$, unique to dynamical geometry) makes time emergent and converts entanglement into irreversible decoherence (§12.5.4).

**Q: Is $C = 1$ the prediction, or is the prediction a range?** The rigorously derivable statement is the range $C \in [2/\pi, 1]$: the lower edge is the Margolus–Levitin speed limit on strict orthogonalization, the upper edge is the natural Markovian dephasing rate (the Diósi–Penrose value). For a visibility-decay measurement — what an interferometer reports — the relevant coefficient is the ceiling, $C = 1$, and that is the single value we quote. It is not proved unique; the residual freedom is the clock-model normalization (Chapter 13). So: quote $C = 1$, remember the box, do not claim the box has collapsed.

## Exercises

**Exercise 12.1.** The saturation exponent is $\Gamma_{\text{sat}} = \xi \ln(d/\ell_{\text{P}})$ with $\xi = GM^2/(\hbar c)$. (a) Recompute $\Gamma_{\text{sat}}$ for $M = 10\,\mu\text{g}$ at $d = 1$ mm, and for $M = 0.1\,\mu\text{g}$ at $d = 1$ mm (hold $\ln(d/\ell_{\text{P}}) = 73.2$ fixed). (b) By what factor does $\Gamma_{\text{sat}}$ change when $M$ increases tenfold at fixed $d$? (c) Use your answers to argue that $\Gamma_{\text{sat}}$ cannot be the universal constant $1/(2\pi)$.

**Exercise 12.2.** The Margolus–Levitin bound gives $\tau_\perp \geq \pi\hbar/(2E)$ for a state of mean energy $E$ above its ground state. (a) Taking $E = E_{\text{G}} = 6.7 \times 10^{-26}$ J, compute the minimum orthogonalization time $\tau_\perp$ at the benchmark. (b) Convert it to a maximum orthogonalization rate and read off the corresponding coefficient $C_\perp$ in $\Gamma = C_\perp E_G/\hbar$. (c) Compare $\tau_\perp$ with the $C = 1$ decoherence time 1.6 ns and comment on which is the "fastest" and which the "natural" timescale.

**Exercise 12.3.** The decay crosses over from Gaussian (Zeno) to exponential (Markovian) at $\tau_{\text{corr}} \approx d/c$. (a) Compute $\tau_{\text{corr}}$ for $d = 1$ mm and for $d = 100$ nm. (b) For each, form the ratio $\tau_{\text{corr}}/\tau_{\text{dec}}$ using $\tau_{\text{dec}} = \hbar d/(GM^2)$ at $M = 1\,\mu$g. (c) Show the ratio is independent of $d$, identify what it equals, and say what that implies about where to look for the $\Gamma(2t)/\Gamma(t) = 4$ signature.

## Solutions

*Solution (Exercise 12.1).* (a) $\xi = (M/m_P)^2$ scales as $M^2$; $\ln(d/\ell_P) = 73.2$ at $d = 1$ mm is unchanged. At $M = 1\,\mu$g, $\xi = 2.11 \times 10^{-3}$ and $\Gamma_{\text{sat}} = 0.154$. Scaling by $M^2$:

$$M = 10\,\mu\text{g}: \ \Gamma_{\text{sat}} = 100 \times 0.154 = 15.4; \qquad M = 0.1\,\mu\text{g}: \ \Gamma_{\text{sat}} = 0.01 \times 0.154 = 1.54 \times 10^{-3}.$$

(b) A tenfold increase in $M$ multiplies $\Gamma_{\text{sat}}$ by $10^2 = 100$. (c) A universal constant cannot change with $M$ at all, let alone by four orders of magnitude across a factor of 100 in mass. The value 0.154 passing near $1/(2\pi) = 0.159$ at the benchmark is a coincidence of the smooth function $\xi \ln(d/\ell_P)$ crossing 0.159 near $(1\,\mu\text{g}, \sim 1\,\text{cm})$; and a different mode-count gives 0.098. So $\Gamma_{\text{sat}}$ is a parameter- and cutoff-dependent overlap exponent, not a pinned constant — and not a rate.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch12 Gamma_sat scaling)`

*Solution (Exercise 12.2).* (a) $\tau_\perp = \pi\hbar/(2E_G) = (\pi/2)(\hbar/E_G) = (\pi/2)(1.58 \text{ ns}) = 2.48$ ns. (b) The maximum orthogonalization rate is $\Gamma_\perp = 1/\tau_\perp = (2/\pi)E_G/\hbar$, so $C_\perp = 2/\pi = 0.637$ — the floor of the window (12.17). (c) The $C = 1$ dephasing time is 1.6 ns; the Margolus–Levitin strict-orthogonalization time is 2.5 ns. Strict orthogonality is the *slower* process (it demands the branches become fully distinguishable), so its bound gives the smaller rate, $C = 2/\pi$. Fringe-visibility dephasing is faster, $C = 1$ — the natural value we quote. The factor between them is exactly $\pi/2$, the whole width of the coefficient window.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch12 Margolus-Levitin tau_perp, C=2/pi)`

*Solution (Exercise 12.3).* (a) $\tau_{\text{corr}} = d/c$: for $d = 10^{-3}$ m, $\tau_{\text{corr}} = 10^{-3}/(3 \times 10^8) = 3.3$ ps; for $d = 10^{-7}$ m, $\tau_{\text{corr}} = 10^{-7}/(3 \times 10^8) = 3.3 \times 10^{-16}$ s $= 0.33$ fs. (b) With $\tau_{\text{dec}} = \hbar d/(GM^2)$ at $M = 1\,\mu$g: for $d = 1$ mm, $\tau_{\text{dec}} = 1.6$ ns and $\tau_{\text{corr}}/\tau_{\text{dec}} = 3.3 \times 10^{-12}/1.6 \times 10^{-9} = 2.1 \times 10^{-3}$; for $d = 100$ nm, $E_G = GM^2/d = (6.674 \times 10^{-11})(10^{-18})/10^{-7} = 6.67 \times 10^{-22}$ J, $\tau_{\text{dec}} = \hbar/E_G = 1.58 \times 10^{-13}$ s, and $\tau_{\text{corr}}/\tau_{\text{dec}} = 3.3 \times 10^{-16}/1.58 \times 10^{-13} = 2.1 \times 10^{-3}$ — the *same* ratio. (c) Algebraically, $\tau_{\text{corr}}/\tau_{\text{dec}} = (d/c)(GM^2)/(\hbar d) = GM^2/(\hbar c) = \xi$, the dimensionless coupling, independent of $d$: it equals $\xi = 2.1 \times 10^{-3}$ at $M = 1\,\mu$g regardless of separation. So the Zeno window is always a fixed small fraction $\sim \xi$ of the decoherence time at fixed mass; to widen it *relative* to $\tau_{\text{dec}}$ one must raise $M$ (increase $\xi$), not shrink $d$. The $\Gamma(2t)/\Gamma(t) = 4$ signature lives in that window, so heavier masses — not just wider or narrower splits — make the Gaussian onset easier to resolve.

▶ Run it: `TOOLS/verify-paper-numerics.py (section N: ch12 crossover d/c, ratio = xi)`

**What to carry forward**

1. The free-field influence integral *saturates* (Theorem 12.1, [PROVEN]): the graviton bath is super-Ohmic ($\varrho(0^+) = 0$), so $\Gamma(t)$ climbs to a static value $\Gamma_{\rm sat} = \xi \ln(d/\ell_{\rm P}) \approx 0.154$ and stops. This is a one-time 14% loss of the squared overlap, *not* a rate; its nearness to $1/2\pi$ is coincidence, and it scales as $M^2$ with a log cutoff.

2. The linear-in-$t$ rate comes from the Wheeler–DeWitt constraint through the modular identity $K = 2\pi H_{\rm phys}$: $\Gamma_{\rm dec} = C\, E_{\rm G}/\hbar$ with $C \in [2/\pi, 1]$ (Margolus–Levitin floor, Mandelstam–Tamm/Diósi–Penrose ceiling), natural $C = 1$, giving $\tau_{\rm dec} = 1.6$ ns ($C = 1$) to 2.5 ns ($C = 2/\pi$). The energy scale $E_{\rm G}$ is [PROVEN]; the *rate* is [CONJECTURED] (rests on $K = 2\pi H_{\rm phys}$, reduced to R1, Chapter 13); $C$ is bounded, not uniquely pinned.

3. $G^1$ versus $G^2$ is not a contradiction but two questions: $G^2$ is entanglement *generation* from a bare state (perturbative QFT, $\sim 3 \times 10^{34}$ times slower, $\tau_{G^2} \approx 1.5 \times 10^{18}$ yr); $G^1$ is entanglement *manifestation* for the physical constraint-satisfying state. Electromagnetism, computed identically, gives only a reversible phase — its constraint is spatial, not Hamiltonian — which is the clean control isolating the mechanism. All five $G^1$ arguments reduce to the single input $H_{\rm total} = 0$.

# The R1 Frontier

13

**Synopsis.** This is the honest frontier of the book. Everything in Part III — the rate $\Gamma_{\text{dec}} = C\,E_{\text{G}}/\hbar$, the nanosecond on the cover — rests on a single operator equation, $K = 2\pi H_{\text{phys}}$ (Chapter 7): modular time is physical time. That equation is [PROVEN] wherever we can check it by hand — in every expectation value, in every solvable model, exactly in flat-space Rindler. What is *not* proven is the leap from those checks to the full operator statement for a genuine, dynamical, interacting gravitational causal diamond. That leap is not a fog of unknowns: it is *one* well-posed mathematical construction, which the theory's internal review named *R1*. This chapter takes R1 apart in the open. We watch the type $\text{III}_1$ obstruction forbid the very entropy Axiom I needs; we watch the crossed product repair it with a theorem; we follow the finite diamond, and the graviton, as far as rigor honestly reaches; and we name — to sixty-four decimal places — exactly how much is left and why it is a rigor gap, not a size gap.

**You will be able to.** state precisely what separates the central conjecture from a theorem; explain why a local region has no entropy and how the crossed product supplies one; reproduce the IR-safety cancellation ($R$ drops out of the rate); and read the R1 ledger — what is [PROVEN], what is [DERIVED] at CFT-proxy standing, and what remains [CONJECTURED] — without mistaking any one label for another.

**Prerequisites.** Chapter 6 (algebras of observables, the type $\text{III}_1$ character of a local region — no trace), Chapter 7 (the crossed product $\text{III} \to \text{II}_\infty$, the Hislop–Longo diamond weight, the CHM map, modular flow), Chapter 9 (Axiom III, $K = 2\pi H_{\text{phys}}$), and Chapter 12 (the rate that depends on all of it).

## 13.1 What exactly is being claimed — and what is not

Let us be scrupulous about the object at issue, because the whole point of this chapter is that its status has been widely, and understandably, overstated — including, at times, in earlier drafts of this very theory.

The rate of gravitational decoherence, derived across Chapters 11 and 12, has the form

$$\Gamma_{\text{dec}} = C\,\frac{E_{\text{G}}}{\hbar}, \qquad E_{\text{G}} = \frac{GM^2}{d}, \qquad C \in \left[\tfrac{2}{\pi},\, 1\right], \tag{13.1}$$

and it is a rate — a number of nats of coherence lost per second — only because a certain identity holds between two clocks. On one side is *modular time*: the flow $\sigma_t$ that Tomita–Takesaki theory attaches to any region's algebra and state (Chapter 7), an intrinsic, purely information-theoretic notion of "time" with no reference to any metre stick. On the other is *physical time*: evolution by the honest Hamiltonian $H_{\text{phys}}$ an observer inside the region would read off a laboratory clock. The generator of modular flow is the modular Hamiltonian $K$, and the claim — Axiom III of Chapter 9, the heart of the correspondence — is that the two coincide up to the Bisognano–Wichmann factor of $2\pi$:

$$K \;=\; 2\pi\, H_{\rm phys} \;+\; O(G^2). \tag{13.2}$$

If (13.2) holds *as an operator equation* — as an equality of two operators on the physical Hilbert space, not merely of their expectation values — then the modular gap between the two branches of a superposition is $2\pi E_{\rm G}/\hbar$, the modular flow is a genuine time translation an observer can follow, and (13.1) is a bona fide rate. The energy scale $E_{\rm G}$ is [PROVEN] independently (Chapter 11); the rate rides entirely on (13.2).

Here is the exact status, and every word in it is load-bearing:

> $K = 2\pi H_{\rm phys}$ **is** [PROVEN] **in expectation values and in every solvable model** — the thermal (Gibbs) case of Chapter 7, the Rindler wedge where Bisognano–Wichmann makes it *exact*, every Gaussian model we can diagonalise. **It is** [CONJECTURED] **as a full operator equation** for a dynamical, interacting gravitational causal diamond. The entire distance between "proven" and "conjectured" is one sharply-posed construction, R1.

**Q 13.1** ***"So is the central equation proven or not? You keep saying both."***

Both, precisely because the two statements are genuinely different mathematical objects, and conflating them is exactly the error this chapter exists to prevent. An *expectation-value* identity says $\langle\psi|\,K\,|\psi\rangle = 2\pi\,\langle\psi|\,H_{\rm phys}\,|\psi\rangle$ for the states $|\psi\rangle$ we care about — this we can verify, and have, to machine precision. An *operator* identity says $K$ and $2\pi H_{\rm phys}$ act *identically on every state*, agreeing in every matrix element, every commutator, every correlation function — a strictly stronger claim, and the one the rate-extraction argument of Chapter 12 actually invokes when it treats $\sigma_t$ as physical evolution. The gap between the two is not philosophy; it is the difference between knowing a function's average and knowing the function. In a solvable model you can check the stronger statement by diagonalising everything; in an interacting field theory you cannot, and what is missing is a proof — named, bounded, and, as we will see, physically minuscule. Calling it [PROVEN] outright would be the overclaim the theory's own reviewers flagged and corrected. Calling it hopeless would ignore that it is one construction away, with the residue quantified. The honest word is [CONJECTURED], with a very short, very sharp list of what remains.

The plan of the chapter is that list, walked end to end. Section 13.2 states the obstruction that makes the whole enterprise nontrivial: a local region has no entropy at all, so Axiom I cannot even be written down naively. Section 13.3 repairs it with the crossed product — a theorem, not a hope. Section 13.4 follows the finite causal diamond and finds the rate is IR-safe (a real, non-trivial check). Section 13.5 carries the construction to the graviton itself. Section 13.6 states the two remaining interacting values, M7 and M8, as the sharp residue — and is candid about what "a controlled-model value but not a proof" means. Section 13.7 gives the ledger, the grade B+, and the honest fork: what would close R1, and what would falsify it instead.

## 13.2 The obstruction: a local region has no entropy

Recall from Chapter 6 the fact that makes quantum gravity mathematically unlike anything in a finite-dimensional Hilbert space. The algebra of observables localized in a bounded region of spacetime — a causal diamond $\mathcal{D}_R$, the double cone that is the causal completion of a ball of radius $R$ — is, in the vacuum of a quantum field theory, a von Neumann factor of *type* $III_1$. We built the classification in Chapter 6; here is the one consequence that matters.

> A type $\mathrm{III}_1$ factor admits *no trace*. There is no linear functional Tr on it with $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ and $\mathrm{Tr}(\mathbb{1})$ finite. Consequently there are no density matrices $\rho$ representing the region's state, no $\mathrm{Tr}(\rho \ln \rho)$, and *no von Neumann entropy at all.*

This is not a technicality to be regularised away; it is the honest structure of the vacuum. We met its physical shadow already in Chapter 4: the entanglement entropy of a region is genuinely infinite, because the vacuum is entangled with itself across the boundary at every length scale down to zero distance. The type $\mathrm{III}_1$ character is that infinity made precise. Every would-be density matrix is smeared into oblivion by the short-distance entanglement; there is no finite $\rho$ to take the log of.

Now look at what Axiom I asks for. Axiom I (Chapter 9) is the statement that the generalized entropy

$$S_{\text{gen}} = \frac{A}{4\ell_{\text{P}}^2} + S_{\text{ext}} \tag{13.3}$$

is a physical, finite, well-behaved quantity — indeed the quantity whose stationarity gives Einstein's equations in Chapter 10. But the second term, $S_{\text{ext}}$, is exactly the von Neumann entropy of the region's state, and we have just said that quantity does not exist. The naive reading of Axiom I is asking for the entropy of a state that has no entropy.

**Q 13.2** ***"If the entropy does not exist, isn't Axiom I simply ill-posed? What exactly is missing?"***

What is missing is a legitimate way to *define* the entropy, and the honest situation is that the naive definition fails and a better one exists but must be constructed. The failure is real: you cannot take $\mathrm{Tr}(\rho \ln \rho)$ of a state with no $\rho$. The fix is also real, and it is the whole content of the next section: you enlarge the algebra by adjoining the observer's own clock, which converts type $\mathrm{III}_1$ into type $\mathrm{II}_\infty$, and type II algebras *do* have a trace, hence a *renormalized* entropy — the divergent short-distance piece subtracted off in a canonical way — which is finite and is exactly $S_{\text{gen}}$. So Axiom I is not ill-posed; it is well-posed only after the crossed-product construction, which is one of the things R1 must supply and, for the kinematic case, *does* supply rigorously. The precise scope of "does supply" is the next section's business, and it is the part where honesty pays: the theorem is airtight for the static case; the dynamical diamond is where the open question lives.

## 13.3 The repair: adjoin the clock, gain a trace

The resolution is one of the genuine triumphs of the last few years of mathematical physics, and it is a *theorem*, so we state it as one and are careful about its scope.

The move (Chapter 7) is to stop describing the region by its bare algebra $\mathcal{A}(\mathcal{D}_R)$ and instead describe it together with an observer who carries a clock. Mathematically, one adjoins to the algebra the generator of its own modular flow — the observer's clock is conjugate to modular time — and forms the *crossed product*

$$\widehat{\mathcal{A}} \;=\; \mathcal{A}(\mathcal{D}_R) \rtimes_\sigma \mathbb{R}. \tag{13.4}$$

Takesaki's structure theorem then does something remarkable and completely general.

**Theorem 13.1** (Crossed product supplies a trace)**.** *For any von Neumann algebra of type III and any modular flow $\sigma_t$, the crossed product (13.4) is a factor of type $II_\infty$. A type $II_\infty$ factor carries a semifinite trace* tr*; hence $\widehat{\mathcal{A}}$ admits density matrices $\hat\rho$, a finite renormalized von Neumann entropy $S = -\,\mathrm{tr}(\hat\rho \ln \hat\rho)$, and a well-defined notion of "rate" (a $d/dt$ of a trace).* [PROVEN] *— Takesaki (1973), unconditional; it assumes nothing about the geometry.*

The physics of the theorem is worth one sentence of interpretation, in the Einstein manner. *Entropy is not a property of a region; it is a property of a region together with an observer who can time it.* The bare region has no entropy because it has no clock to measure change against; give it a clock — adjoin the modular generator — and the trace, the density matrix, and the entropy all snap into existence. The observer's clock is not decoration; it is what makes the ledger of Axiom I balance.

This is the Chandrasekaran–Longo–Penington–Witten (CLPW, 2023) mechanism, and it is essential to state its established scope precisely, because this is exactly where the frontier begins.

> **What CLPW established rigorously.** For the *de Sitter static patch* — a horizon whose modular flow is a genuine *isometry* (a Killing time translation of the background) — the crossed product (13.4) is type $\mathrm{II}_\infty$ with a trace, the generalized entropy $S_{\text{gen}}$ is its finite von Neumann entropy, and the whole structure is a theorem. [PROVEN] for that case. The construction requires the modular flow to be an *isometry* of the background — a symmetry of the actual geometry.

That last clause is the hinge of the entire chapter. The de Sitter static patch has an isometric modular flow, and CLPW is a theorem there. The *finite causal diamond* — the region an actual laboratory experiment occupies — does not: as we are about to see, its modular flow is *conformal*-Killing, not Killing. It preserves the geometry only up to a position-dependent rescaling. The crossed-product construction for a conformal-Killing flow with interacting matter is the case CLPW did not solve, and it is precisely R1. The mathematics that works perfectly for the symmetric case must be extended to the case that is only *almost* symmetric — and that extension is the open problem, stated exactly.

## 13.4 The finite diamond, and a real check: IR-safety

We now specialise to the object an experiment actually inhabits: a finite causal diamond $\mathcal{D}_R$, the causal completion of a ball of radius $R$. Chapter 7 gave us its modular data explicitly, and we lean on that result rather than re-deriving it.

For the vacuum of a conformal field on the diamond, the modular Hamiltonian is *local* and *geometric* — the Hislop–Longo result:

$$K \;=\; 2\pi \int_{r<R} d^3x\, f(r)\, T_{00}(x), \qquad f(r) \;=\; \frac{R^2 - r^2}{2R}. \tag{13.5}$$

The weight $f(r)$ is the fingerprint of the whole construction: it is the local inverse-temperature profile of the diamond. At the centre it takes the value

$$f(0) \;=\; \frac{R}{2}, \qquad \text{so} \qquad T_{\text{loc}}(0) \;=\; \frac{1}{2\pi f(0)} \;=\; \frac{1}{\pi R}, \tag{13.6}$$

the local Tolman temperature a central observer reads. The flow this generates is a conformal-Killing flow of Minkowski space (it moves the diamond's causal structure to itself but rescales distances), and Chapter 7's CHM map $\hat{g} = \Omega^{-2}\eta$ straightens it into an exact Killing flow — a constant-speed time translation — on the static cylinder $\mathbb{R} \times H^3$, hyperbolic space of curvature radius $R/2$. On that cylinder the observer at the diamond's centre follows the flow at a uniform rate; she has, at last, a clock.

Now comes a check that could have failed and did not — which is what makes it worth trusting. The diamond radius $R$ is entirely arbitrary. It is the observer's own choice of how large a region to call "the experiment"; nothing physical can depend on it. In particular the *decoherence rate must not depend on* $R$. Let us see that it does not.

**Worked Example 13.1 — The diamond size cancels out**

Put the two branches of the superposition near the centre, at $r_0 = aR$ with $a$ a fixed fraction (so $a = 0$ for a centred source). Two quantities carry the branch information, and we track the $R$-dependence of each.

*The modular gap.* The two branches differ in modular energy by $\Delta K = 2\pi f(r_0)\, E_{\text{G}}/\hbar$, and by (13.5) the weight at the source is

$$f(r_0) \;=\; \frac{R^2 - (aR)^2}{2R} \;=\; \frac{R}{2}\,(1 - a^2) \;\propto\; R^1.$$

The gap grows linearly with $R$: a bigger diamond, a bigger modular gap.

*The observer's clock.* But the modular gap is measured in *modular* time, and to get a physical rate we must convert to the observer's *proper* time. The intrinsic clock of Chapter 7 runs at $d\tau/ds = 2\pi f(0) = \pi R$ (the constant cylinder norm makes it uniform) — also $\propto R^1$. A bigger diamond stretches the clock by exactly the same factor.

*The ratio.* The physical phase rate is the modular gap divided by the clock rate, and the two powers of $R$ cancel:

$$\frac{d\Phi}{d\tau} \;=\; \frac{\Delta K}{d\tau/ds} \;=\; \frac{2\pi f(0)\, E_{\text{G}}/\hbar}{2\pi f(0)} \;=\; \frac{E_{\text{G}}}{\hbar}, \tag{13.7}$$

independent of $R$. Restoring the off-centre source, the assembled rate is $\Gamma_{\text{dec}} = 2N(1 - a^2)\, E_{\text{G}}/\hbar$ with $N$ the (single) open normalization: the Tolman factor $(1 - a^2)$ is itself $R$-independent. The arbitrary IR regulator drops out completely.

▶ Run it: `TOOLS/verify_diamond_rate_ir_safety.py` (`R`-independence across $\sim 6$ decades in $R$; rate $= 2N(1 - a^2)E_{\text{G}}/\hbar$)

**Q 13.3** ***"Isn't that cancellation rigged? Both quantities were built from the same $f(0)$, so of course the $R$'s cancel — that proves nothing."***

This is the sharpest possible objection and it deserves the sharpest possible answer, because the theory's own adversarial review raised exactly it. You are right that, *once* you know the modular gap and the clock are both governed by the *same* weight $f$, their ratio is manifestly $R$-independent — the cancellation is then a near-tautology, and by itself would "verify" nothing a construction could fail. The physical content is one level up: it is the claim that the branch gap and the observer's clock are *both* controlled by the one Hislop–Longo weight (13.5) — that the same geometric object sets the energy splitting the branches feel and the rate the observer's clock ticks. That coincidence is not automatic; it is the statement that the modular flow really is the physical flow, i.e. it is (13.2) again, viewed through the diamond. The honest reading of Worked Example 13.1 is therefore: *if* the central equation holds, the rate is IR-safe, and this is a real and necessary consistency of the construction — a bigger arbitrary box cannot change a laboratory number. It is not, and we do not claim it to be, an independent proof of (13.2). Labelling matters here: the cancellation is [PROVEN]; what it is a cancellation *of* inherits the central equation's [CONJECTURED] status. Keeping those two apart is the discipline of this chapter.

There is one more thing the diamond teaches, and it sharpens where type II is actually needed. The IR-safe phase rate (13.7) is a *phase* — an overlap exponent — and a phase lives happily in the bare type III algebra; it needs no trace. Type II, the crossed product, the trace, the entropy — all of that machinery is required only to turn the phase into a *rate*: to say how many nats of coherence are lost per second, you need the $d/dt$ of a trace, and the trace is exactly what type III forbids and the crossed product restores. The static overlap of Chapter 11 gives the energy scale; the constraint plus the crossed-product trace give the rate. This is the same division of labour we flagged in Chapter 12, seen now from the algebraic side.

## 13.5 Carrying it to the graviton

Everything so far — Hislop–Longo, CHM, the IR cancellation — was stated for a *conformal* matter field, because that is where the conformal map that straightens the diamond flow is available. But the field that actually carries the branch information is the gravitational field itself: the linearized graviton, a massless spin-2 field whose action is *not* conformally invariant. The conformal map seemed unavailable to it. This was the graviton residue of R1 — the worry that "the CHM conformal map is structurally unavailable to the spin-2 field."

It is dissolved at the level of the *flow*, and the key realisation is that modular flow is intrinsic to the pair (algebra, state) — it never asked the action to be conformally invariant; it only needs the algebra and the vacuum. Two established results and one anchor combine.

**Theorem 13.2** (Graviton diamond flow is geometric)**.** *The gauge-invariant algebra of the linearized graviton in a ball $B_R$ factorises (Benedetti–Casini) into two scalar towers labelled by the two helicities $P = \pm$ and by angular momentum $\ell \geq 2$:*

$$\mathcal{A}_{\mathrm{grav}}(B_R) \;=\; \bigotimes_{P=\pm} \; \bigotimes_{\ell\geq 2} \; \bigotimes_{|m|\leq\ell} \mathcal{A}_\ell^{(m,P)}((0,R)), \tag{13.8}$$

*each factor a reduced scalar mode. Each tower inherits the local $f(r)$-weighted modular Hamilto-*

*nian* (13.5) *(Huerta–van der Velde), so the graviton modular flow* is *the CHM conformal-Killing flow restricted to the towers.* [PROVEN AT CFT PROXY].

Three facts assemble the theorem, and each carries its own honesty.

*The tower content.* The $\ell = 0$ and $\ell = 1$ metric perturbations carry no radiative degrees of freedom — they are the ADM constraint and gauge sector ($10 - 4 - 4 = 2$ propagating helicities). So only $\ell \geq 2$ survives, with two helicities per $(\ell, m)$: the mode count in the tower at angular momentum $\ell$ is $\sum_{|m|\leq\ell} \sum_{P=\pm} 1 = 2(2\ell + 1)$, a number the graviton-modular-flow harness checks directly.

*The first-principles anchor.* Why should the graviton towers carry the *same* $f(r)$ weight as a scalar? Because at $M \to 0$ the graviton's own radial equations degenerate to the scalar one. The Regge–Wheeler (odd-parity) and Zerilli (even-parity) master potentials of black-hole perturbation theory, and the scalar potential, all collapse at $M = 0$ to the same centrifugal barrier $\ell(\ell + 1)/r^2$ — using the identity $(\ell - 1)(\ell + 2) + 2 = \ell(\ell + 1)$ — with solutions the Riccati–Bessel functions $r\, j_\ell(\omega r)$. This degeneration is verified symbolically and on a lattice to machine precision: it is the concrete reason the spin-2 field inherits the scalar's geometric flow.

*Block-diagonality.* The SO(3)×parity-invariant vacuum has a covariance that is block-diagonal in $(\ell, m, P)$, so by Araki's theorem the modular operator is a direct sum over the towers — the flow does not mix them, and it suffices to know it tower by tower.

Now the candor, stated as plainly as the result. Theorem 13.2 *advances* R1; it does *not* close it.

> **What the graviton reduction does and does not do.** It removes the graviton obstruction: the propagating graviton flows geometrically, with the same weight $f(r)$ as the matter proxy, so the diamond's total modular flow is clock-like after all — the fear that the spin-2 sector had a non-geometric, non-clock-like flow that would spoil (13.2) for the total algebra is answered. [PROVEN AT CFT PROXY]. It does *not* supply the intrinsic CLPW observer for the interacting theory, does *not* fix the anomaly/trace normalization, and does *not* settle the edge-mode/center bookkeeping at $\partial B_R$. Those remain open. The central equation (13.2) therefore stays [CONJECTURED].

"CFT proxy" is the phrase carrying the caveat, and it means exactly this: the construction is rigorous when the matter is a conformal field, for which the CHM map is exact; the physical graviton is transcribed into that framework tower by tower, inheriting its standing, but the transcription is not itself a first-principles proof for the non-conformal interacting theory. It is a very strong [DERIVED], resting on [PROVEN] inputs — and it is not a theorem for the real field. That distinction is the whole reason this chapter is titled "frontier" and not "closure."

## 13.6 The sharp residue: two interacting values, M7 and M8

After the crossed product, the diamond, the IR cancellation, and the graviton reduction, what is actually left? Not a vague mist. The open content of R1 condenses, at the interacting level, to exactly *two values.* Naming them precisely — and being honest about what "controlled-model value but not a proof" means — is the most important paragraph in the book's ledger.

### 13.6.1 M8: the modular kernel, and why $O(G^2)$ is astronomically small

The first value governs how much the interacting corrections perturb the modular Hamiltonian away from $2\pi H_{\rm phys}$. Write the state as a reference plus a perturbation, $\rho = \rho_0 + \delta\rho$, and pass to the modular eigenbasis. First-order modular perturbation theory (Chapter 7) gives the change in $K$ as the image of $\delta\rho$ under a *universal kernel*:

$$(\delta K)_{ij} \;=\; -\,h(\omega_{ij})\,\widetilde{\delta\rho}_{ij}, \qquad h(\omega) \;=\; \frac{\omega}{2\sinh(\omega/2)}, \qquad h(0) = 1, \tag{13.9}$$

where $\omega_{ij}$ is the modular-energy difference between eigenstates $i$ and $j$. The kernel $h$ is positive-definite (its Fourier transform is $\frac{\pi}{2}\,\mathrm{sech}^2(\pi s) \geq 0$, so Bochner's theorem applies), which means the map $\delta\rho \mapsto \delta K$ is a *complete contraction* with *contraction constant exactly* 1: the correction to $K$ is never larger than the correction to the state that caused it, and off-resonant couplings are suppressed exponentially, $h \sim |\omega|\,e^{-|\omega|/2}$.

What does this buy the rate? The dominant $O(G^2)$ channel — the graviton-self-interaction squeezing of the two branches — contributes to the rate only at second order because it is a purely anomalous (particle-number changing) term whose KMS expectation vanishes. Composing the benchmark numbers gives the size of the leading correction to the rate:

$$\left|\frac{\delta\Gamma}{\Gamma}\right|_{O(G^2)} \;\lesssim\; \left(\frac{\ell_{\rm P}}{d}\right)^2 + c\,\epsilon_1^2 \;\approx\; 2.6\times10^{-64} \qquad (1\,\mu{\rm g},\; 1\,{\rm mm}), \tag{13.10}$$

the pair-production channel $(\ell_{\rm P}/d)^2$ dominating. Read the magnitude: the leading interacting correction to the decoherence rate is of order $10^{-64}$. The $G^1$-versus-$G^2$ signal — the thing an experiment will test (Chapter 15) — is clean to some sixty-four decimal places against this channel.

### 13.6.2 M7: the split constant

The second value governs the operator-norm error incurred when the diamond's algebra is split into the two nearly-independent sub-regions the branches occupy — the rigorous bookkeeping the type II construction needs. The proxy is the corridor mutual information $I(A\!:\!B)$ between the sub-regions; Pinsker's inequality bounds the product-state approximation error by $\|\rho_{AB} - \rho_A\otimes\rho_B\|_1 \leq \sqrt{I/2}$. In the gapped Gaussian model (the gap being the $H^3$ curvature gap of the CHM cylinder), the mutual information decays exponentially with the split distance $\delta$:

$$I(\delta) \;=\; C_{\rm split}\,e^{-2\kappa\delta}, \qquad C_{\rm split} \approx 0.060. \tag{13.11}$$

The constant $C_{\rm split} \approx 0.060$ is an explicit number, computed in closed form from a two-mode lock ($\nu_A - \frac{1}{2} = g^2/16w^4$ with $g = e^{-\kappa\delta}$) and confirmed on a lattice. The rate $2\kappa$ is twice the curvature gap because mutual information scales as correlation squared.

### 13.6.3 What "controlled-model value but not a proof" means

Both $C_{\rm split} \approx 0.060$ and the $O(G^2)$ residual $\approx 2.6\times10^{-64}$ are *numbers* — definite, reproducible, harness-checked. So why is R1 still open?

**Q 13.4 *"Why should I believe these controlled-model values? You have explicit numbers with error bars — isn't that a proof?"***

It is a proof *of the model*, and that is exactly the gap. A "controlled model" here means a Gaussian — free-field, or two-mode — system in which every operator can be diagonalised and every quantity computed in closed form. In that model, $C_{\text{split}} = 0.060$ is a theorem: you can check it by hand, and we have. What a Gaussian model *cannot* do is prove the corresponding statement in the full *interacting* field theory, where the operators do not commute into a tidy normal form and the closed-form diagonalisation is unavailable. The value is a rigorously-computed *representative* of what the interacting answer should be — strong evidence, a controlled approximation, a number an experiment could not plausibly contradict — but it is not a field-theoretic proof that the interacting constant *equals* the Gaussian one. The distinction is the same one as in Q 13.1: knowing the answer in every case you can solve is not the same as a proof for the case you cannot. That is why M7 and M8 are the [CONJECTURED] content: they are the two places where the construction has an explicit controlled-model value and lacks the interacting proof. Note the crucial feature of (13.10), though — the residue is not merely bounded, it is bounded to be *physically negligible*, $O(10^{-64})$. R1 is a *rigor* gap, not a *size* gap: nobody doubts the number is tiny; what is missing is the theorem that it is tiny.

There is a subtlety worth flagging, because it is the technical heart of M8 and it shows the reasoning is genuinely delicate. For a *unitary* family of states, $\delta\rho = [G, \rho_0]$, the algebraic identity $h(\omega) \cdot 2\sinh(\omega/2) = \omega$ makes the kernel factor cancel exactly: $\delta K = [G, K_0]$, with *no* contraction suppression. The squeezing channel is such a unitary direction, so its operator norm is genuinely $O(\sqrt{G})$ in the coupling — it is only because the term is purely anomalous (its KMS expectation vanishes) that its contribution to the rate onsets at $O(G^2)$ and lands at (13.10). The kernel gives the residual's *structure*; it does not by itself give the small number — the vanishing KMS expectation does. Getting this right is what separates the honest bound from a hopeful one.

## 13.7 The ledger, the grade, and the honest fork

We can now write down, with every label exact, what R1 is and is not.

| **R1 ingredient** | **Status** | **Source** |
|---|---|---|
| Type $\mathrm{III}_1$ obstruction is real (no trace, flat space) | [ESTABLISHED] | local-field support |
| $\mathcal{A} \rtimes_\sigma \mathbb{R}$ is type $\mathrm{II}_\infty$ + trace | [PROVEN] | Takesaki (unconditional) |
| CLPW crossed product for the isometric (static-patch) flow | [PROVEN] | CLPW 2023 |
| Diamond flow = CHM Killing flow on $\mathbb{R} \times H^3$ | [PROVEN] | CHM; symbolic check |
| Geometric modular flow, weight $f(r)$, $f(0) = R/2$ | [PROVEN] | Hislop–Longo; Huerta–vdV |
| IR-safe phase rate $E_G/\hbar$, $R$-independent | [PROVEN] | Worked Ex. 13.1 |
| Graviton $\to$ two scalar $\ell \geq 2$ towers, $2(2\ell + 1)$ modes | [PROVEN AT CFT PROXY] | Benedetti–Casini; RW/Zerilli |
| Composition (one $f$, one $2\pi$, one flow) | [PROVEN] | synthesis check (5/5) |
| M8 kernel $h(\omega)$, contraction constant 1 | [PROVEN] | first-order modular theory |
| M7 split value $C_{\text{split}} \approx 0.060$ (Gaussian) | [DERIVED] | controlled model; lattice |
| M8 $O(G^2)$ residual $\approx 2.6 \times 10^{-64}$ (controlled) | [DERIVED] | controlled model |
| Operator equation $K = 2\pi H_{\text{phys}}$ | [CONJECTURED] | R1 closure (M7, M8) |
| Absolute rate $\Gamma_{\text{dec}} = C\, E_G/\hbar$, $C \in [2/\pi, 1]$ | [CONJECTURED] | inherits from the above |

The theory's internal adversarial review grades the whole programme **B+**. The reasoning is exactly the ledger above:

> **Why B+.** The energy scale $E_G = GM^2/d$ is [PROVEN] (classical gravity, rigorous, Chapter 11). The $G^1$ *scaling* is established. But the operator equation $K = 2\pi H_{\text{phys}}$ and hence the rate $\Gamma_{\text{dec}} = C\, E_G/\hbar$ are [CONJECTURED] — proven in expectation values and solvable models, open as full operator statements, and reduced to the single well-posed construction R1. One sharp falsifiable prediction; the internal consistency suite passes at $< 10^{-15}$ relative error; the residual gap localized to R1. Not A (that would need R1 closed); not lower (the physics is intact and the prediction unchanged). B+ is the grade of a theory that knows exactly what it has not yet proven.

And now the fork, stated as candidly as Einstein flags where his own deliberations go meaningless.

*What would close R1.* Two things, and only two: a field-theoretic proof that the interacting

M7 split constant equals its Gaussian value, and a field-theoretic proof that the M8 residual is $O(G^2)$ as the controlled model says (equivalently, a construction of the crossed product for the conformal-Killing flow with interacting matter, together with the intrinsic CLPW observer, the anomaly normalization, and the edge-mode bookkeeping). Close those and $K = 2\pi H_{\mathrm{phys}}$ becomes [PROVEN], the rate extraction of Chapter 12 is legitimized, the coefficient selection $C = 1$ is fixed, and the grade lifts.

*What would falsify it instead.* An experiment. If a levitated-nanosphere interferometer (Chapter 15) measures the $G^2$ generation rate rather than the $G^1$ manifestation rate — coherence times $\sim 10^{34}$ times longer than (13.1) predicts — the conjecture is dead, R1 or no R1. Or R1 itself might turn out ill-posed under scrutiny, which would be its own kind of falsification. The mathematics and the laboratory are both live arbiters, and the theory is honest enough to name both.

Step back and look at the shape of what is left. The residue is $O(G^2) \approx 10^{-64}$ — so far below any conceivable measurement that, *physically*, the conjecture is as good as certain. Yet mathematically the residue is entirely real: a theorem is a theorem, and no smallness of a number substitutes for a proof that the number is small. This is the exact posture of a theory at its frontier: astonishingly close to certainty in physics, one sharp construction away from certainty in mathematics, and candid to the last decimal about the difference. Few approaches to quantum gravity can say precisely, in one named construction and two numbered values, what stands between them and a theorem. This one can. That precision is not a weakness of the frontier; it is the frontier's whole value.

## Chapter FAQ

**Q: In one sentence — is the theory proven?** No: its central operator equation $K = 2\pi H_{\mathrm{phys}}$, and therefore its decoherence rate, are [CONJECTURED] — [PROVEN] in every expectation value and solvable model, open as a full operator equation for the interacting gravitational diamond, and reduced to one construction (R1) whose remaining content is two controlled-model values (M7, M8) and a physically negligible $O(G^2)$ residue. The energy scale is [PROVEN]; the prediction is falsifiable; the grade is B+.

**Q: Why does a local region really have no entropy — isn't that just a divergence you could regulate?** The divergence is physical, not a regularisation artefact (Chapter 2 Q, Chapter 4): the vacuum is entangled with itself across the boundary at every scale down to zero distance, so the entanglement entropy is genuinely infinite, and the algebra is type $\mathrm{III}_1$ — which *by classification* has no trace and no density matrix. Any cutoff that makes the entropy finite imports arbitrariness. The crossed product subtracts the infinity canonically, by adjoining the observer's clock, leaving the finite renormalized $S_{\mathrm{gen}}$. Only relative and renormalized entropies are physical — Axiom I's fine print.

**Q: If CLPW already proved the crossed-product construction, what is left for R1?** CLPW proved it for a modular flow that is a true *isometry* of the background — the de Sitter static patch. The finite laboratory diamond has a *conformal*-Killing flow (an isometry only up to a position-dependent rescaling), and the crossed product for a conformal-Killing flow with interacting matter is not in the literature. R1 is precisely that extension: a theorem for the almost-symmetric case, given a theorem for the symmetric one.

**Q: The $O(G^2)$ residual is $10^{-64}$. Why not just declare victory?** Because $10^{-64}$ is a *size*, and what is missing is a *proof*. The number is computed in a controlled (Gaussian) model where everything diagonalises; the interacting field theory is where the closed-form diagonalisation fails, and the controlled value, however tiny and however robust, is not a theorem for the interacting case. R1 is a rigor gap, not a size gap. Declaring victory on a size would be exactly the overclaim the honest labelling of this chapter refuses.

## Exercises

**Exercise 13.1.** The M8 kernel is $h(\omega) = \omega/[2\sinh(\omega/2)]$. (a) Show $h(0) = 1$ by expanding sinh. (b) Prove the algebraic identity $h(\omega) \cdot 2\sinh(\omega/2) = \omega$, and use it to show that for a *unitary* family of states, $\delta\rho = [G, \rho_0]$, the first-order modular change is $\delta K = [G, K_0]$ *exactly* — with no kernel suppression. (c) The Fourier transform of $h$ is $\widehat{h}(s) = \frac{\pi}{2}\,\mathrm{sech}^2(\pi s)$. Verify $\int_{-\infty}^{\infty} \widehat{h}(s)\,ds = 1$ and explain why this equals $h(0)$, and why $\widehat{h} \geq 0$ forces $h(\omega) \leq 1$ (the contraction constant).

**Exercise 13.2.** A source sits at $r_0 = aR$ inside a diamond of radius $R$. (a) Using the Hislop–Longo weight $f(r) = (R^2 - r^2)/2R$, compute $f(r_0)$ and confirm it scales as $R^1$. (b) The observer clock runs at $d\tau/ds = 2\pi f(0) = \pi R$. Form the phase rate $\Delta K/(d\tau/ds)$ and show it equals $(1-a^2)\,E_\mathrm{G}/\hbar$, independent of $R$. (c) At what fractional radius $a$ is the rate halved relative to a centred source? What is the physical meaning of the vanishing rate as $a \to 1$ (source at the horizon)?

## Solutions

*Solution (Exercise 13.1).* (a) $2\sinh(\omega/2) = \omega + \omega^3/24 + \cdots$, so $h(\omega) = \omega/(\omega + \omega^3/24 + \cdots) \to 1$ as $\omega \to 0$: $h(0) = 1$. (b) Directly, $h(\omega) \cdot 2\sinh(\omega/2) = \frac{\omega}{2\sinh(\omega/2)} \cdot 2\sinh(\omega/2) = \omega$. For a unitary family $\rho(\theta) = e^{-i\theta G}\rho_0 e^{i\theta G}$, one has $\delta\rho = -i[G, \rho_0]$, whose modular-basis entry is $\widetilde{\delta\rho}_{ij} = -i\,G_{ij}\,2\sinh(\omega_{ij}/2)$ (the $\rho_0^{\pm 1/2}$ factors contribute eigenvalues $e^{\mp k/2}$). Then $(\delta K)_{ij} = -h(\omega_{ij})\widetilde{\delta\rho}_{ij} = +i\,G_{ij}\,\omega_{ij} = [G, K_0]_{ij}$: the kernel exactly cancels the $2\sinh$. No suppression — so the squeezing channel is $O(\sqrt{G})$ in norm, and its smallness in the rate comes from its vanishing KMS expectation, not from the kernel. (c) $\int \frac{\pi}{2}\,\mathrm{sech}^2(\pi s)\,ds = \frac{\pi}{2} \cdot \frac{1}{\pi}[\tanh(\pi s)]_{-\infty}^{\infty} = \frac{1}{2} \cdot 2 = 1$. This equals $h(0)$ because $h(0) = \int \widehat{h}$ (the value of a function at the origin is the integral of its transform); and $\widehat{h} \geq 0$ (positive-definiteness, Bochner) means $h$ is a characteristic function, so $|h(\omega)| \leq h(0) = 1$ — the contraction constant is exactly 1.

▶ Run it: `TOOLS/verify_bilocal_value_kernel.py (three-form equivalence; Bochner/Schur contraction; unitary-family exactness)`

*Solution (Exercise 13.2).* (a) $f(r_0) = (R^2 - a^2R^2)/2R = (R/2)(1-a^2)$, manifestly $\propto R^1$. (b) $\Delta K = 2\pi f(r_0)\,E_\mathrm{G}/\hbar = \pi R(1-a^2)\,E_\mathrm{G}/\hbar$, and $d\tau/ds = \pi R$, so $\Delta K/(d\tau/ds) = (1-a^2)\,E_\mathrm{G}/\hbar$ — the two powers of $R$ cancel, leaving an $R$-independent rate carrying only the Tolman factor $(1-a^2)$. (c) The rate is halved when $1 - a^2 = \frac{1}{2}$, i.e. $a = 1/\sqrt{2} \approx 0.71$. As $a \to 1$ the source approaches the diamond's edge, which the CHM map sends to the horizon at infinite modular time; the local Tolman factor $(1-a^2) \to 0$, so the modular flow — and with it the decoherence rate — freezes. Physically: a superposition pressed against the horizon of its own causal diamond dephases ever more slowly, the redshift of modular time. (The overall normalization $N$ multiplying the rate is the separate clock-model normalization and does not affect this $R$- and $a$-dependence.)

▶ RUN IT: `TOOLS/verify_diamond_rate_ir_safety.py` (rate $= 2N(1-a^2)E_G/\hbar$; $R$-independence)

## What to carry forward

1. The entire rate rests on the operator equation $K = 2\pi H_{\text{phys}}$, which is [PROVEN] in expectation values and every solvable model but [CONJECTURED] as a full operator statement for the interacting gravitational diamond. That single gap is one well-posed construction, R1. A local region's algebra is type III$_1$ — no trace, no entropy — so Axiom I is well-posed only *after* the crossed product (Takesaki, [PROVEN]) adjoins the observer's clock and supplies the renormalized $S_{\text{gen}}$.
2. The finite diamond flows by the geometric Hislop–Longo weight $f(r) = (R^2 - r^2)/2R$ ($f(0) = R/2$, $T_{\text{loc}}(0) = 1/\pi R$); the rate is IR-safe — the arbitrary diamond size $R$ cancels between the modular gap and the observer's clock ([PROVEN], but a consistency of the central equation, not an independent proof of it). The graviton reduces to two scalar $\ell \geq 2$ towers ($2(2\ell+1)$ modes each) carrying the same $f(r)$ flow ([PROVEN AT CFT PROXY]), dissolving the graviton residue — advancing R1, not closing it.
3. The open content is two controlled-model values: the split constant $C_{\text{split}} \approx 0.060$ (M7) and the $O(G^2)$ residual $\approx 2.6 \times 10^{-64}$ (M8) — explicit numbers in Gaussian models, not field-theoretic proofs for the interacting theory. R1 is a *rigor* gap, not a *size* gap: the residue is astonishingly small physically ($\sim 10^{-64}$) yet mathematically real. Grade **B+**. Closing M7 and M8 lifts it; an experiment measuring the $G^2$ rate falsifies it (Chapter 15).

# Part IV

# The Universe and the Laboratory

# 14 The Universe as the Laboratory

**Synopsis.** The three axioms of Part III were forged on a bench-top: a grain, a beam splitter, a millimetre. This chapter changes only the size of the apparatus. We apply the *same* horizon clock — the modular temperature of Chapter 7, the Unruh temperature of Chapter 4 — to the one horizon nobody can leave, the cosmological horizon of Chapter 3. With no new free parameter it hands us the acceleration scale $a_0 = cH_0/2\pi \approx 1.08 \times 10^{-10}$ m/s$^2$ that galactic rotation curves have whispered for forty years (right to about ten percent, and honestly so); it fixes the *form* of dark energy, $\rho_{\rm DE} \propto c^2H^2/G$ (a cosmological constant, $w = -1$), while leaving its *coefficient* to be read off the sky; and it tames a number that looked catastrophic — a naive cosmic decoherence rate of $10^{103}$ Hz turns out to be a Minkowski-formula artifact, replaced by the finite, sub-causal $\Gamma_{\rm total} = g(1)H_0/4\pi \approx 7 \times 10^{-20}$ Hz. We close on a striking consistency check — "the lock" — in which all of cosmology's input cancels.

**You will be able to.** derive $a_0 = cH_0/2\pi$ from entanglement equilibrium at the cosmological horizon and compare it honestly with data; state precisely which part of $\rho_{\rm DE} = \alpha c^2H^2/G$ is derived and which is fitted; compute the de Sitter form factor $g(x) = 1 - \frac{2}{\pi}\,\mathrm{Si}(x)$ and the cosmic decoherence rate; and verify the cancellation $a_0/(c\,\Gamma_{\rm total}) = 2/g(1)$.

**Prerequisites.** Chapters 3 (FRW, de Sitter, the Hubble horizon $c/H$), 4 (Unruh temperature, Bunch–Davies modes), 5 (the generalized entropy $S_{\rm gen}$), 7 (the horizon clock, temperature $H/2\pi$), 9–10 (Axioms I+II and Einstein's equations with $\Lambda$), and 11–12 (the decoherence functional). No new machinery is built here; only pointed elsewhere.

## 14.1 The horizon clock, one last time

Everything in this chapter comes from a single sentence we have already earned, transported to the largest horizon there is.

Recall the shape of the argument. In Chapter 4 an observer accelerating uniformly through the vacuum found it warm: the Unruh temperature (4.27), $T_{\rm U} = \hbar a/2\pi c k_{\rm B}$. In Chapter 7 we saw *why* the number is $2\pi$ and nothing else — it is the single Euclidean revolution, the angle you turn through to return to where you began (7.15) — and we learned to read $2\pi$ off any horizon at all: whatever generates the flow into the hidden region is a Hamiltonian, and the state is thermal with respect to it at inverse temperature $2\pi$ in the natural (dimensionless) flow time.

Now take the horizon of Chapter 3. If the late universe is dominated by a cosmological constant, the Friedmann equation makes $H$ constant and space expands exponentially, $a(t) = a_0e^{Ht}$ (3.34): de Sitter space. It has a horizon at the physical distance where recession outruns light,

$$R_{\rm H} = \frac{c}{H}, \tag{14.1}$$

the Hubble horizon (3.35). Galaxies past $R_{\mathrm{H}}$ are causally cut off from us exactly as the region behind a black-hole horizon is cut off — and, like every horizon in this book, it carries a temperature. Feeding the de Sitter surface gravity into the same Bisognano–Wichmann $2\pi$ that gave us Unruh, the cosmological horizon is thermal at

$$T_{\mathrm{dS}} = \frac{\hbar H}{2\pi\, k_{\mathrm{B}}}, \qquad \text{equivalently a clock frequency} \qquad \omega_H = \frac{H}{2\pi}. \tag{14.2}$$

*The universe ticks.* An observer at rest in the cosmic rest frame sits inside a thermal bath at temperature $T_{\mathrm{dS}}$, and the clock buried in that bath runs at the frequency $\omega_H = H/2\pi$ — one Hubble rate, divided by the same $2\pi$ revolution. This is the Gibbons–Hawking temperature of de Sitter space, and (14.2) is the *only* cosmological input this chapter uses. Everything below — the MOND scale, dark energy's form, the cosmic decoherence rate, the lock — is squeezed out of this one clock.

How cold is it? Insert $H = H_0$ for the present universe. We take $H_0 = 2.268 \times 10^{-18}$ s$^{-1}$ throughout, which is the standard 70 km s$^{-1}$ Mpc$^{-1}$ (a Mpc is $3.086 \times 10^{22}$ m; the conversion is Exercise 14.1). Then

$$T_{\mathrm{dS}} = \frac{(1.0546 \times 10^{-34})(2.268 \times 10^{-18})}{2\pi\,(1.381 \times 10^{-23})} = 2.7 \times 10^{-30}\ \mathrm{K}. \tag{14.3}$$

Thirty orders of magnitude below the coldest laboratory, colder than the cosmic microwave background by thirty powers of ten. You will never feel this bath directly. Its consequences, though, are neither small nor subtle, and the first of them has been sitting in every galaxy's rotation curve since the 1970s.

**Q 14.1** ***"The Unruh temperature needed a definite acceleration $a$. What sets the "acceleration" of an observer sitting still in the cosmos?"***

The de Sitter horizon has a surface gravity of its own, $\kappa = cH$, playing the role that proper acceleration played for Rindler and that $c^4/4GM$ played for Schwarzschild (Chapter 3). A static observer in de Sitter is, geometrically, accelerating relative to the horizon — she must fire an engine to avoid being carried off by the expansion — and the temperature she reads is set by that surface gravity through the universal $\hbar\kappa/2\pi c k_{\mathrm{B}}$. The $2\pi$ is not fitted here any more than it was for Unruh; it is the same single revolution of Chapter 7, now around the tip of the cosmological wedge. One horizon, one clock, one $2\pi$.

## 14.2 MOND from the horizon clock

### 14.2.1 The forty-year anomaly

Spiral galaxies rotate wrong. Newton, fed the visible mass, predicts orbital speeds that fall off as $v \sim r^{-1/2}$ once you pass the bright interior — the same law that makes Neptune orbit slower than Mercury. Instead the measured speed goes *flat*: out past the last visible star, $v(r)$ stops falling and holds constant for as far as the neutral hydrogen can be traced. Either there is a great deal of unseen mass arranged just so (dark matter), or Newton's law itself bends when the acceleration gets small enough. Milgrom's 1983 observation was that the bend, if real, happens at a *universal acceleration*: rotation curves depart from Newton not at a fixed radius or a fixed mass, but wherever the Newtonian acceleration drops below one and the same scale,

$$a_0^{\rm obs} \;\approx\; 1.2\times 10^{-10}\ {\rm m/s^2}. \tag{14.4}$$

This number is measured, not fitted galaxy by galaxy: one constant fixes the transition in every rotation curve ever plotted. Where does a fundamental acceleration that small come from? Notice its size: $a_0^{\rm obs}$ is, to a factor of a few, exactly $cH_0$ — the speed of light times the inverse age of the universe. That coincidence is the thread we now pull.

### 14.2.2 Deriving the scale

The derivation runs on Axioms I and II applied at the cosmological horizon, exactly as they were applied at a Rindler horizon in Chapter 10. The physical picture first, in plain words, before any symbol: a mass at rest is bound energy *removed* from the de Sitter thermal bath; removing energy from a bath lowers its entropy; and the surrounding geometry must find room in its own entropy budget to carry that deficit. Where the budget suffices, small curvature adjustments restore equilibrium and ordinary gravity holds. Where it fails — at low enough acceleration — the deficit can only be carried non-locally, and the effective gravity changes. The scale at which the budget runs out is $a_0$.

*Derivation. Step 1 — the entropy a mass displaces.* A mass $M$ at rest is energy $E = Mc^2$ withdrawn from the bath at temperature $T_{\rm dS}$. By the Clausius relation $dS = dE/T$, it lowers the bath entropy by

$$\Delta S(M) \;=\; \frac{Mc^2}{T_{\rm dS}} \;=\; \frac{2\pi k_{\rm B}\, Mc^2}{\hbar H_0}, \tag{14.5}$$

using (14.2). Because $T_{\rm dS}$ is so absurdly cold, even a modest mass makes an enormous deficit; this deficit is what the geometry must accommodate.

*Step 2 — the budget the geometry can offer.* By Axiom I the de Sitter volume carries entropy at a fixed density $s_\Lambda = 3c^2H_0k_{\rm B}/(4G\hbar)$ (this is the horizon entropy $S_{\rm gen} = A_H/4\ell_{\rm P}^2$ of Chapter 5, spread through the enclosed volume; we take it as input from there). Inside a sphere of radius $r$ the available entropy is

$$S_{\rm avail}(r) \;=\; s_\Lambda\cdot\tfrac{4}{3}\pi r^3 \;=\; \frac{\pi c^2 H_0 k_{\rm B}\, r^3}{G\hbar}, \tag{14.6}$$

growing as $r^3$. Local equilibrium (Axiom II, $\delta S_{\rm gen} = 0$) requires the budget to cover the deficit, $S_{\rm avail}(r) \geq \Delta S(M)$.

*Step 3 — where the budget runs out.* Equality $S_{\rm avail}(r_c) = \Delta S(M)$ pins the critical radius. Every constant cancels:

$$\frac{2\pi k_{\rm B} Mc^2}{\hbar H_0} = \frac{\pi c^2 H_0 k_{\rm B}\, r_c^3}{G\hbar} \quad\Longrightarrow\quad r_c = \Big(\frac{2GM}{H_0^2}\Big)^{1/3}. \tag{14.7}$$

The Newtonian acceleration $GM/r_c^2$ evaluated here still depends on $M$, so it is not yet universal — the entropy-budget argument establishes that the relevant scale is set by $cH$, but the universal *value* comes from the bath itself.

*Step 4 — the universal acceleration.* The intrinsic energy of the de Sitter bath is $E_{\rm dS} = k_{\rm B}T_{\rm dS} = \hbar H_0/2\pi$. Converting it to an acceleration with the two Planck scales $m_{\rm P} = \sqrt{\hbar c/G}$, $\ell_{\rm P} = \sqrt{\hbar G/c^3}$ of Chapter 1 — an acceleration is an energy per (mass×length) — gives, with the constants collapsing,

$$a_0 = \frac{E_{\rm dS}}{m_{\rm P}\,\ell_{\rm P}} = \frac{\hbar H_0/2\pi}{\sqrt{\hbar c/G}\,\sqrt{\hbar G/c^3}} = \frac{cH_0}{2\pi}. \tag{14.8}$$

$$\boxed{a_0 = \frac{c\,H_0}{2\pi}} \tag{14.9}$$

*The MOND acceleration is the speed of light times the horizon clock frequency* $\omega_H = H_0/2\pi$*.* It contains no adjustable constant: the sole external number is the measured $H_0$, and nothing whatever was tuned to galactic data. Two logically distinct steps went in, and honesty requires separating them (this is the Einstein move of naming what is forced versus what is a convention): the entropy budget forces the scale to be $cH$, and the factor $\frac{1}{2\pi}$ is the Gibbons–Hawking temperature assignment (14.2) — a dimensional identification of the same standing as Bekenstein's separation of the black-hole area law from its coefficient, not an independent derivation of the $2\pi$. [DERIVED] (with that fine print).

**Worked Example 14.1 — The MOND scale from the Hubble rate**

Insert $c = 2.998 \times 10^8$ m/s and $H_0 = 2.268 \times 10^{-18}$ s$^{-1}$ into (14.9):

$$a_0 = \frac{(2.998\times10^8)(2.268\times10^{-18})}{2\pi} = \frac{6.80\times10^{-10}}{6.283} = 1.08\times10^{-10}\ \mathrm{m/s^2}. \tag{14.10}$$

Compare the observed value (14.4), $a_0^{\rm obs} \approx 1.2\times10^{-10}$ m/s$^2$: the ratio is $a_0/a_0^{\rm obs} \approx 0.90$. A parameter-free number computed from $c$ and $H_0$ alone lands within about 10% of a quantity measured in galaxies — objects that know nothing of the cosmological horizon. This is the headline cosmological prediction of the framework.

▶ Run it: `TOOLS/verify-paper-numerics.py (check: mond_scale,` $a_0 = cH_0/2\pi$`)`

Once the scale is fixed, the phenomenology follows by the standard MOND logic (Chapter 10's equilibrium supplies the interpolation in the low-acceleration regime). Below $a_0$ the effective acceleration becomes $g_{\rm eff} = \sqrt{g_{\rm N}\,a_0}$ instead of $g_{\rm N}$, so far from a mass $M$,

$$g_{\rm eff} = \frac{\sqrt{GMa_0}}{r} \implies \frac{v^2}{r} = g_{\rm eff} \propto \frac{1}{r} \implies v = \text{const}, \tag{14.11}$$

flat rotation curves, and the asymptotic speed obeys the baryonic Tully–Fisher law $v^4 = GMa_0 = GM\,cH_0/2\pi$ — normalization set by cosmology, no free parameter.

**Q 14.2** ***"Isn't MOND just a fit? You have one number $a_0$ and you tuned it to rotation curves."***

That is exactly what MOND *was* as Milgrom proposed it: a phenomenological fit, one constant $a_0$ read off the data, no derivation. The content here is precisely that we did *not* tune it. Equation (14.9) computes $a_0$ from $c$ and $H_0$ — two numbers measured with no reference to any galaxy — and the answer agrees with the fitted value to ten percent. A fit that you can derive from unrelated inputs stops being a fit and becomes a prediction. The honest caveat is the ten percent itself (and the $2\pi$ convention noted above): we predict

1.08, the sky says 1.2. That is a genuine near-miss, not an exact hit, and §14.2.3 says so plainly rather than rounding it away.

### 14.2.3 The honest comparison, including the tension

Two things must be said out loud, in the main text, because a parameter-free prediction earns its keep only if its misses are reported as loudly as its hits.

*First, the present-day value is close but not exact.* We predict $a_0 = 1.08 \times 10^{-10}$; the data give $a_0^{\rm obs} = (1.2 \pm 0.2) \times 10^{-10}$. The central prediction is low by about 10% — inside the observational error bar, but only just, and a future tightening of $a_0^{\rm obs}$ could turn the 10% into a real discrepancy. We do not claim an exact match; we claim a parameter-free number that is close.

*Second, the theory predicts that $a_0$ evolves, and the one place we can check it, it is off.* Equation (14.9) was derived at the cosmological horizon, whose clock frequency is $H$, not $H_0$. Run the same argument at redshift $z$, where the Hubble rate was $H(z) > H_0$, and it predicts

$$a_0(z) \;=\; \frac{c\,H(z)}{2\pi}, \tag{14.12}$$

a MOND scale that was *larger* in the past. This is a sharp, falsifiable departure from Milgrom's original constant $a_0$, and it is testable with high-redshift rotation curves (JWST kinematics, Gaia). The current status is mixed and we report it honestly: an evolving $a_0$ of roughly this form is favoured over a strictly constant one at $z \sim 1$, which is a point for (14.12) — but the measured $a_0(z{\sim}1)$ sits about 24% *above* the value (14.12) predicts. The direction (evolution, and upward with $z$) matches; the magnitude does not, yet. This is a live tension, not a confirmation, and the next few years of high-$z$ kinematics will decide it. [DERIVED] (the scale and its evolution), with an open $\sim$24% tension in the one high-redshift data point.

**What is and is not claimed for MOND here.** *Claimed:* the acceleration scale $a_0 = cH_0/2\pi$ is derived from de Sitter thermodynamics with no fitted constant, agrees with the observed scale to $\sim$10%, and predicts a specific redshift evolution $a_0(z) = cH(z)/2\pi$. *Not claimed:* that the framework explains the full success of the dark-matter paradigm (clusters, the CMB acoustic peaks, structure formation), where pure MOND is known to struggle; that the $2\pi$ coefficient is independently derived rather than assigned by Gibbons–Hawking; or that the $z \sim 1$ magnitude is yet consistent. This is a parameter-free prediction that is close and testable, no more and no less.

## 14.3 Dark energy: the form is ours, the number is the sky's

The same horizon, the same two axioms, deliver the accelerated expansion itself — but here the honest ledger reads differently from MOND, and keeping the difference straight is the whole point of the section.

Apply Axiom I to the cosmological horizon: it carries entropy $S_{\rm gen} = A_H/4\ell_{\rm P}^2$ with area $A_H = 4\pi R_{\rm H}^2 = 4\pi(c/H)^2$. Apply Axiom II — equilibrium, the saturation of the generalized second law — to the horizon plus the matter it encloses,

$$\frac{d}{dt}\Big(\frac{A_H}{4\ell_{\mathrm{P}}^2} + S_{\mathrm{matter}}\Big) \;=\; 0. \tag{14.13}$$

Combined with the Friedmann equation that relates $H$ to the energy density (Chapter 10 derived Einstein's equations, $\Lambda$ included, from exactly this equilibrium), the horizon's entropy budget acts as a vacuum energy whose density scales as

$$\boxed{\rho_{\mathrm{DE}} \;=\; \alpha\,\frac{c^2H^2}{G}} \tag{14.14}$$

with $\alpha$ a dimensionless number of order $10^{-1}$. Read the *structure* of (14.14) first, because the structure is the prediction. A density $\propto H^2/G$ tracking the horizon is, in a universe approaching de Sitter, essentially constant: as the matter dilutes and $H \to$ const, $\rho_{\mathrm{DE}} \to$ const, i.e. a cosmological constant with equation of state $w = -1$. *The horizon's entropy budget behaves exactly like* $\Lambda$. That $\rho_{\mathrm{DE}} \propto c^2H^2/G$ with $w = -1$ is what QGC derives. [DERIVED] (the functional form).

### Worked Example 14.2 — Where $\alpha$ comes from — and where it does not

The coefficient is *not* predicted; it is read off the measured dark-energy fraction. Writing the critical density $\rho_{\mathrm{crit}} = 3H^2c^2/8\pi G$ and $\rho_{\mathrm{DE}} = \Omega_\Lambda\rho_{\mathrm{crit}}$, equation (14.14) gives

$$\alpha \;=\; \frac{3\,\Omega_\Lambda}{8\pi}. \tag{14.15}$$

Insert the observed $\Omega_\Lambda \approx 0.70$:

$$\alpha = \frac{3(0.70)}{8\pi} = \frac{2.10}{25.13} = 0.084. \tag{14.16}$$

(The value $\alpha \approx 0.082$ quoted in the research papers uses the Planck best fit $\Omega_\Lambda \approx 0.685$; the two differ only through which sky survey one reads.) So $\alpha$ is *fitted* to make $\Omega_\Lambda$ come out right — the number comes from the sky. What the framework supplies is the box (14.14): the shape, the $w = -1$, the tracking of the horizon. The structure is derived; the coefficient is measured. [MOTIVATED] (the value of $\alpha$; fitted to $\Omega_\Lambda$).

▶ RUN IT: `TOOLS/verify-paper-numerics.py` (section N: ch14 $\alpha = 3\Omega_\Lambda/8\pi$)

There is one further honest point, and it is a positive one. The value of $\alpha$ is *fitted today*, but the framework does contain a genuine dynamical landmark: the GSL/event-horizon equilibrium has a future de Sitter attractor, $\Omega_\Lambda \to 1$, at which $\alpha \to 3/8\pi \approx 0.119$. That attractor value is a prediction of the asymptotic state, not a fit; it is simply not where we sit now ($\Omega_\Lambda \approx 0.7$, not 1). We flag it as a structural feature and decline to overclaim: the tempting near-coincidence $\alpha \approx 1/4\pi$ was investigated and found to have no independent dynamical meaning, so we do not promote it.

**Q 14.3 *"Why is $\alpha$ not predicted, when $a_0$ was? Both came from the same horizon and the same axioms."***

Because they are different *kinds* of quantity, and the asymmetry is instructive. The MOND scale $a_0$ is set by the horizon *temperature* alone: one clock frequency, $\omega_H = H_0/2\pi$, times $c$ — a rate, and rates the horizon clock fixes outright. Dark energy's $\alpha$ is a *normalization of a density*: it depends not just on the horizon's temperature but on how the horizon entropy is apportioned against the enclosed matter as the universe evolves — an $\mathcal{O}(1)$ bookkeeping factor that the equilibrium condition constrains to the right functional form but does not, at the present epoch, pin to a unique value. In Einstein's language (Chapter 1's discussion of $\phi(v)$): a constraint can force the *shape* of an answer without determining every constant multiplying it. The shape, $c^2H^2/G$ with $w = -1$, is forced. The constant $\alpha$ is, for now, data. Saying otherwise would be dishonest, and the framework's discipline is to label the fit a fit.

## 14.4 Cosmic decoherence and the Minkowski artifact

We now turn the decoherence machinery of Chapters 11–12 on the universe itself, and meet a number that at first looks like a disaster — then dissolves into a lesson about using the right propagator.

### 14.4.1 The absurd naive rate

The bench-top result was $\Gamma = E_{\mathrm{G}}/\hbar$ with $E_{\mathrm{G}} = GM^2/d$, giving $\tau_{\mathrm{dec}} = \hbar d/GM^2$ (Chapter 12). What if we feed it the cosmos — the mass inside the horizon, $M \sim M_U \sim 10^{53}$ kg, separated by the horizon scale, $d \sim R_{\mathrm{H}} = c/H_0$? Then

$$\Gamma_{\mathrm{naive}} \sim \frac{GM_U^2}{\hbar\, R_{\mathrm{H}}} = \frac{c^5}{4G\hbar H_0} \sim 10^{103}\ \mathrm{Hz}. \tag{14.17}$$

A decoherence rate of $10^{103}$ per second — $10^{60}$ times the inverse Planck time. Taken at face value the universe would decohere infinitely faster than anything can causally happen, which is nonsense: the causal bound at horizon scale is $\Gamma \le c/R_{\mathrm{H}} = H_0 \sim 10^{-18}$ Hz. Something is badly wrong with (14.17).

What is wrong is the *propagator*. The formula $\Gamma = GM^2/\hbar d$ was derived on Minkowski space, where the graviton field is built from plane waves and *every* mode, down to arbitrarily long wavelength, contributes to the which-path record. On a Minkowski background that is (with self-energy subtraction) mathematically legitimate. On an *expanding* background it is simply the wrong calculation: modes longer than the horizon do not behave like plane waves at all.

### 14.4.2 The de Sitter recomputation

Redo the overlap of Chapter 11 on the de Sitter background (3.34) with the graviton in the Bunch–Davies vacuum (Chapter 4's de Sitter modes). One physical fact changes everything. Bunch–Davies modes split by wavelength: *sub-horizon* modes ($k_{\mathrm{phys}} \gg H$) oscillate like flat-space waves and carry which-path information as before; *super-horizon* modes ($k_{\mathrm{phys}} \ll H$) are *frozen* — their amplitude locks in and stops oscillating, so they no longer generate time-dependent decoherence. The Hubble scale thus provides an *automatic infrared cutoff* at $k_H = aH/c = 1/R_{\mathrm{H}}$

that Minkowski entirely lacked. This is the resolution in one sentence: the catastrophic long-wavelength modes of (14.17) are frozen out by the expansion.

Carrying the calculation through (the same coherent-state overlap, now with the horizon cutoff and proper — not coordinate — separation), the decoherence exponent factorizes into the Minkowski answer times a dimensionless form factor:

$$\Gamma_{\rm dS}(t,d,H) \;=\; \frac{G\,\Delta m^2}{\hbar\,d}\;g\Big(\frac{Hd}{c}\Big)\;t, \tag{14.18}$$

where $\Delta m$ is the mass difference between the two branches and the form factor, obtained in closed form by solving the sourced de Sitter mode equation, is

$$\boxed{g(x) \;=\; 1-\frac{2}{\pi}\,\mathrm{Si}(x)} \qquad \mathrm{Si}(x)=\int_0^x \frac{\sin u}{u}\,du. \tag{14.19}$$

Here $x = Hd/c$ is the separation measured in Hubble radii, and Si is the sine integral of Chapter 2's toolkit. Read the three limits; they carry the whole physics:

- $g(0) = 1 - \frac{2}{\pi}\,\mathrm{Si}(0) = 1$: at sub-horizon separations ($Hd/c \to 0$) the form factor is unity and (14.18) *recovers the Minkowski Chapter 12 rate exactly.* This is the consistency check: had it failed, the calculation would be wrong.
- $g(1) = 1 - \frac{2}{\pi}\,\mathrm{Si}(1)$: at the horizon scale, a finite $\mathcal{O}(1)$ number, computed below.
- $g(x) \to 0$ as $x \gg 1$: super-horizon configurations do not decohere. Mass distributions separated by more than $R_{\rm H}$ are causally disconnected and cannot record which-path information about each other. The de Sitter horizon draws the line.

**Worked Example 14.3 — The horizon form factor $g(1)$**

The sine integral at unit argument is $\mathrm{Si}(1) = 0.946083$ (a standard tabulated value; expand $\sin u/u = 1-u^2/6+u^4/120-\cdots$ and integrate term by term to check it: $1-\frac{1}{18}+\frac{1}{600}-\cdots = 0.9461$). Then

$$g(1) = 1-\frac{2}{\pi}\,\mathrm{Si}(1) = 1-\frac{2}{\pi}(0.946083) = 1-0.6023 = 0.398. \tag{14.20}$$

So at the horizon scale the de Sitter decoherence is suppressed to about 40% of the naive Minkowski value. (This value $g(1) = 0.398$ replaces an earlier estimate of 0.92 that was arithmetically wrong; the closed form (14.19) settles it.)

▶ RUN IT: `TOOLS/verify-paper-numerics.py` (check: `cosmic_decoherence`, $g(1) = 1 - \frac{2}{\pi}\,\mathrm{Si}(1)$)

### 14.4.3 The cosmic rate

Now the payoff. The which-path information about the cosmic wavefunction is carried mode by mode across the horizon, and the branch granularity — the smallest mass difference $\Delta m$ that counts as a distinct cosmic branch — is set by the de Sitter thermal mass, $\Delta m_{\rm dS} = \hbar H/2\pi c^2$ (the mass equivalent of one quantum of the horizon temperature; derived from the Wheeler–DeWitt branch structure, not assumed). Each such thermal-mass mode is itself a horizon-scale object, so it carries the form factor $g(1)$, and its rate is $g(1)\,t_P^2 H^3/4\pi^2$ — fantastically small, about

$10^{-141}$ Hz per mode. But there are $S_{\rm dS} \sim A_H/4\ell_{\rm P}^2 \sim 10^{122}$ modes on the horizon, all decohering in parallel, and the total is their sum:

$$\boxed{\Gamma_{\rm total} \;=\; S_{\rm dS}\cdot g(1)\,\frac{t_P^2 H^3}{4\pi^2} \;=\; g(1)\,\frac{H}{4\pi}} \tag{14.21}$$

the horizon-mode count $S_{\rm dS} = 1/(t_P H)^2$ cancelling the two Planck factors cleanly. *The universe decoheres at $g(1)$ times the horizon clock frequency divided by two* — $\Gamma_{\rm total} = g(1)\,H/4\pi \approx H/(10\pi)$. Numerically, with $H = H_0$,

$$\Gamma_{\rm total} = 0.398 \cdot \frac{2.268\times 10^{-18}}{4\pi} = 0.398\cdot(1.805\times 10^{-19}) = 7.2\times 10^{-20}\ \text{Hz}. \tag{14.22}$$

Seven parts in $10^{20}$ per second: about $0.03\,H_0$, a factor $4\pi/g(1) \approx 32$ *below* the causal bound $H$, not above it. The $10^{103}$ Hz of (14.17) is gone — exposed as a Minkowski-formula artifact, not a physical prediction. On de Sitter the answer is finite, causal, and utterly unobservable: a horizon decoheres roughly once every $4\pi/g(1) \approx 32$ Hubble times. [DERIVED] (controlled approximation: the linearized coherent-state overlap on a fixed de Sitter background, with graviton self-interaction and secular-growth corrections bounded far below this in the relevant regime).

**Q 14.4** ***"If the naive $10^{103}$ Hz was just the wrong propagator, was there ever a real problem — or did you invent a crisis to solve?"***

The crisis was real in the sense that it is the first thing anyone does: take the lab formula, plug in cosmic numbers, get an absurdity. Many would read that absurdity as "the theory predicts nonsense at cosmic scale" and stop. The point of (14.18)–(14.21) is that the absurdity is entirely an artifact of importing the Minkowski propagator into an expanding universe — the same category error that would afflict *any* decoherence calculation done that way, QGC or not. The de Sitter recomputation is not a patch bolted on to rescue a bad number; it is what the theory actually says when you use its own axioms on the correct background. That it lands sub-causal, with no tuning, is the non-trivial part. The one genuinely open input is the branch granularity $\Delta m$ — we take $\Delta m_{\rm dS}$; a different choice would rescale the number — but no reasonable choice reopens the $10^{103}$ catastrophe, because the horizon cutoff has already removed it structurally.

## 14.5 The lock

Here is a check that has no right to be as clean as it is. We have two independent cosmological outputs of the horizon clock: the MOND scale $a_0 = cH_0/2\pi$ from entanglement equilibrium (§14.2), and the cosmic decoherence rate $\Gamma_{\rm total} = g(1)H_0/4\pi$ from the influence functional (§14.4). They were computed by completely different routes — one thermodynamic, one dynamical. Form their natural dimensionless combination, an acceleration divided by (speed × rate):

$$\frac{a_0}{c\,\Gamma_{\rm total}} \;=\; \frac{cH_0/2\pi}{c\cdot g(1)H_0/4\pi} \;=\; \frac{H_0/2\pi}{g(1)H_0/4\pi} \;=\; \frac{2}{g(1)}. \tag{14.23}$$

$$\boxed{\frac{a_0}{c\,\Gamma_{\text{total}}} = \frac{2}{g(1)} = \frac{2}{0.398} \approx 5.03} \tag{14.24}$$

*Every trace of $H_0$ has cancelled.* The ratio of the MOND scale to the cosmic decoherence rate is a pure number — $2/g(1) \approx 5.03$ — with no cosmological input surviving. The value of the Hubble rate, the one measured quantity feeding both sides, drops out completely, leaving only the sine-integral constant that defines the de Sitter form factor.

Why should this be a nontrivial check rather than an accident? Because the two sides come from physically different calculations that share exactly *one* ingredient: the horizon clock frequency $\omega_H = H_0/2\pi$. It appears in $a_0$ directly (times $c$), and in $\Gamma_{\text{total}}$ through both the temperature that sets the thermal mass and the horizon mode count. When the same clock feeds two outputs and you take their ratio, the clock cancels — *provided* both outputs really were built from that one clock and nothing else. The cancellation in (14.23) is a certificate that they were: it is the framework checking its own bookkeeping.

Be precise about its status, though, because it is easy to oversell. The lock is *not* a new prediction of an observable — $2/g(1)$ is a pure number you cannot go out and measure independently of $a_0$ and $\Gamma_{\text{total}}$ themselves. It is an *internal consistency relation*: a demonstration that the MOND sector and the cosmic-decoherence sector are the same physics seen twice, not two unrelated constructions that happen to live in the same chapter. That both trace to a single $\omega_H$ is the unifying claim of Part IV, and (14.23) is its sharpest expression. [DERIVED] (a consistency identity, not an independent observable).

## 14.6 The arrow of time, briefly

One more consequence of the horizon budget deserves a mention, kept short because its status is more provisional than the rest.

The generalized second law (Axiom II, in its inequality form) says $S_{\text{gen}}$ never decreases. On a de Sitter background the horizon entropy $A_H/4\ell_P^2$ has a finite maximum — the full de Sitter value $S_{\text{dS}} \sim 10^{122}$ — which the universe is monotonically approaching as it expands toward the attractor $\Omega_\Lambda \to 1$. That maximum is a *ceiling*, and a ceiling in the future implies a *floor* in the past: for $S_{\text{gen}}$ to have been rising toward it for the whole history of the universe, it must have started far below, in a low-entropy state. The cosmological horizon and the GSL thus supply, at least in outline, the ingredient the "past hypothesis" usually has to assume by hand — a low-entropy boundary condition — by making the high-entropy de Sitter state the generic future and the low-entropy state the necessary past. The entropy-capacity identity $S_{\text{gen}} \leq S_{\text{dS}}$ orients the arrow: time runs from the floor toward the ceiling. [MOTIVATED] (this connects the framework's entropy law to the arrow of time; it is a structural observation, not a derivation of the initial condition, and we do not claim to have explained why the early universe was special — only to have located where such an explanation would have to live).

## 14.7 The cosmological ledger

Part IV used the *same three axioms* as the bench-top decoherence of Part III — Generalized Entropy, Entanglement Equilibrium, Modular Time — and the same one horizon clock $\omega_H = H/2\pi$. Two arenas, one framework. Here is the honest scorecard.

**MOND scale $a_0 = cH_0/2\pi$.** [DERIVED], parameter-free, $\approx 10\%$ from the observed value, with a testable redshift evolution $a_0(z) = cH(z)/2\pi$ — favoured over a constant at $z \sim 1$ but $\sim$24% above the data point there. Falsifiable with Gaia/JWST high-$z$ kinematics.
**Dark energy $\rho_{\rm DE} = \alpha c^2 H^2/G$.** Functional form (a $w = -1$ cosmological constant tracking the horizon) [DERIVED]; coefficient $\alpha \approx 0.082$ [MOTIVATED] — fitted to $\Omega_\Lambda \approx 0.7$, not predicted. The structure comes from QGC; the number comes from data. Future attractor $\alpha \to 3/8\pi$ is a genuine landmark.
**Cosmic decoherence $\Gamma_{\rm total} = g(1)H_0/4\pi \approx 7 \times 10^{-20}$ Hz.** [DERIVED] (controlled approximation); the naive $10^{103}$ Hz was a Minkowski-formula artifact, cured structurally by the horizon cutoff. Sub-causal, unobservably small.
**The lock $a_0/(c\,\Gamma_{\rm total}) = 2/g(1) \approx 5.03$.** [DERIVED], an internal consistency identity in which all cosmological input cancels — a certificate that both sectors are the one horizon clock, not an independent observable.

Only the first entry is a prediction an experimentalist can chase, and it is the one that matters: $a_0 = cH_0/2\pi$ ties galactic dynamics to the cosmological horizon with no adjustable constant, and its redshift evolution is a live test running now. The rest is structure and consistency — important for showing the framework hangs together, silent as laboratory predictions.

Which returns us to the bench. The cosmological sector is safe and mostly unobservable; the decisive tests of QGC are the ones with a beam splitter and a microgram grain. Chapter 15 takes us there: the five-sigma protocol, the mass–separation–time window where the uncoolable gravitational floor of Chapter 1 first shows through the removable noise, and the list of results — in the laboratory, not the sky — that would kill the theory outright.

## Chapter FAQ

**Q: The whole chapter hangs on the de Sitter temperature $\hbar H/2\pi k_{\rm B}$. What if the late universe is not exactly de Sitter?** It is not, and the argument tolerates that. The late universe is *approaching* de Sitter (the attractor $\Omega_\Lambda \to 1$), so $H$ is slowly varying rather than strictly constant. The horizon temperature and clock frequency then use the instantaneous $H$, which is precisely what makes $a_0(z) = cH(z)/2\pi$ evolve. Nothing requires exact de Sitter; it requires that the horizon have a well-defined temperature, which any slowly-evolving FRW horizon does. The corrections from $\dot{H} \neq 0$ are of order the slow-roll parameter, small today.

**Q: Is the $2\pi$ in $a_0$, in $T_{\rm dS}$, and in the Unruh temperature the *same* $2\pi$?** Yes — that is the content of Chapter 7. Every one of them is the single Euclidean revolution: the angle you rotate through to return to the tip of a horizon wedge. Unruh's $2\pi$ (Rindler wedge), the de Sitter $2\pi$ (cosmological wedge), the black-hole $2\pi$ (Schwarzschild wedge), and hence the $2\pi$ in $a_0$ are one geometric fact wearing four costumes. This is why the coefficient is never fitted and never anything but $2\pi$.

**Q: If cosmic decoherence is $10^{-20}$ Hz, is it ever relevant to anything?** Not observationally, and that is the honest answer. It is many orders of magnitude slower than the bench-top $\tau_{\rm dec} = 1.6$ ns of Chapter 1. Its importance is entirely conceptual: it shows the framework does *not* predict a cosmic catastrophe (the naive $10^{103}$ Hz would have been fatal), and it demonstrates that the same decoherence functional, applied consistently on the correct background, stays causal at every scale. A theory that predicted observable cosmic decoherence would already be dead; QGC predicts an unobservable, sub-causal rate, which is exactly what survival requires.

**Q: MOND without dark matter — does this chapter claim the dark-matter paradigm is wrong?** No. It claims one specific thing: the MOND *acceleration scale* is $cH_0/2\pi$, derivable from the horizon clock. Where MOND phenomenology succeeds (galactic rotation curves, the baryonic Tully–Fisher relation), that scale is the framework's parameter-free input. Where MOND struggles (galaxy clusters, the CMB acoustic peaks, large-scale structure), this chapter takes no position and offers no rescue. The derivation of $a_0$ is agnostic about how much of the missing-mass problem MOND versus dark matter ultimately explains; it only fixes the scale at which the transition, wherever it operates, occurs.

## Exercises

**Exercise 14.1.** Convert $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ to inverse seconds, using $1\,\mathrm{Mpc} = 3.086\times10^{22}$ m, and confirm the value $H_0 = 2.268 \times 10^{-18}\ \mathrm{s^{-1}}$ used throughout. Then recompute $a_0 = cH_0/2\pi$ and check (14.10). What $a_0$ would the older "67 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$" value give, and how does the change compare with the 10% gap to $a_0^{\mathrm{obs}}$?

**Exercise 14.2.** (a) Using the series $\sin u/u = \sum_{n\geq0}(-1)^n u^{2n}/(2n+1)!$, show that $\mathrm{Si}(x) = \sum_{n\geq0}(-1)^n x^{2n+1}/[(2n+1)(2n+1)!]$ and hence that the form factor has the small-$x$ onset $g(x) = 1 - \frac{2}{\pi}x + \mathcal{O}(x^3)$. (b) Evaluate Si(1) to three terms and reproduce $g(1) = 0.398$. (c) What does the linear onset $g(x) \approx 1 - \frac{2}{\pi}x$ say physically about how decoherence turns off as a superposition's separation grows toward the horizon?

**Exercise 14.3.** Verify the lock (14.23) numerically from your own values: compute $a_0 = cH_0/2\pi$ and $\Gamma_{\mathrm{total}} = g(1)H_0/4\pi$ with $H_0 = 2.268 \times 10^{-18}$ and $g(1) = 0.398$, form $a_0/(c\,\Gamma_{\mathrm{total}})$, and confirm you get $2/g(1) \approx 5.03$ independently of the numerical value of $H_0$ you chose. Then explain, in one sentence, why choosing a *different* $H_0$ leaves the ratio unchanged.

## Solutions

*Solution (Exercise 14.1).* $H_0 = 70 \times 10^3\ \mathrm{m/s} \div (3.086 \times 10^{22}\ \mathrm{m}) = 2.268 \times 10^{-18}\ \mathrm{s^{-1}}$, confirming the value used. Then $a_0 = (2.998\times10^8)(2.268\times10^{-18})/2\pi = 1.08\times10^{-10}\ \mathrm{m/s^2}$, matching (14.10). With 67 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ ($H_0 = 2.171 \times 10^{-18}$), $a_0 = 1.04 \times 10^{-10}$ — a 4% shift, comparable to the uncertainty in $H_0$ itself and smaller than the $\sim$10% gap to $a_0^{\mathrm{obs}} = 1.2 \times 10^{-10}$. The prediction's near-miss is not an artifact of the Hubble tension: it survives the full range of measured $H_0$.

▶ RUN IT: `TOOLS/verify-paper-numerics.py` (section N: ch14 $a_0$)

*Solution (Exercise 14.2).* (a) Integrating $\sin u/u = 1 - u^2/6 + u^4/120 - \cdots$ term by term from 0 to $x$ gives $\mathrm{Si}(x) = x - x^3/18 + x^5/600 - \cdots$; the leading term $\mathrm{Si}(x) \approx x$ yields $g(x) = 1 - \frac{2}{\pi}\,\mathrm{Si}(x) = 1 - \frac{2}{\pi}x + \mathcal{O}(x^3)$. (b) $\mathrm{Si}(1) = 1 - 1/18 + 1/600 - \cdots = 1 - 0.0556 + 0.00167 - \cdots = 0.9461$, so $g(1) = 1 - (2/\pi)(0.9461) = 1 - 0.6023 = 0.398$. (c) The linear onset says decoherence switches off *gradually and immediately* as the separation grows: at $Hd/c = x$ the rate is already reduced by a fraction $\frac{2}{\pi}x$ below its Minkowski value. Physically, the modes that would have carried which-path information at separation $d$ start crossing the horizon and freezing as $d \to R_{\mathrm{H}}$, and the fraction frozen grows linearly at first — there is no sharp threshold, only a smooth handover to the super-horizon regime where $g \to 0$.

▶ RUN IT: `TOOLS/verify-paper-numerics.py` (section N: ch14 $g(1)$)

*Solution (Exercise 14.3).* $a_0 = 1.08 \times 10^{-10}$ m/s$^2$ and $\Gamma_{\text{total}} = 0.398 \times 2.268 \times 10^{-18}/(4\pi) = 7.2 \times 10^{-20}$ Hz. Then $a_0/(c\,\Gamma_{\text{total}}) = 1.08 \times 10^{-10}/[(2.998 \times 10^8)(7.2 \times 10^{-20})] = 1.08 \times 10^{-10}/(2.16 \times 10^{-11}) = 5.03 = 2/g(1)$. Choosing any other $H_0$ leaves this unchanged because $a_0 \propto H_0$ and $\Gamma_{\text{total}} \propto H_0$ identically, so the single power of $H_0$ cancels in the ratio, leaving only $2/g(1)$ — which is the whole content of the lock.

▶ RUN IT: `TOOLS/verify-paper-numerics.py` (section N: ch14 lock $= 2/g(1)$)

**What to carry forward**

1. The same horizon clock $\omega_H = H/2\pi$ (the de Sitter temperature $T_{\text{dS}} = \hbar H/2\pi k_{\text{B}}$) that gave the Unruh and modular temperatures, applied to the cosmological horizon, is the *only* cosmological input — and from it alone the MOND acceleration $a_0 = cH_0/2\pi \approx 1.08 \times 10^{-10}$ m/s$^2$ falls out with no free parameter, agreeing with the observed scale to $\sim$10% and predicting the testable evolution $a_0(z) = cH(z)/2\pi$ (favoured at $z \sim 1$ but $\sim$24% above the data there).
2. Dark energy's *form* $\rho_{\text{DE}} = \alpha c^2 H^2/G$ (a $w = -1$ cosmological constant) is [DERIVED] from horizon GSL saturation; its coefficient $\alpha \approx 0.082$ is [MOTIVATED] — fitted to $\Omega_\Lambda \approx 0.7$, not predicted. The structure is QGC's; the number is the sky's.
3. The naive cosmic decoherence rate $10^{103}$ Hz is a Minkowski-formula artifact; the de Sitter recomputation gives $\Gamma_{\text{total}} = g(1)H_0/4\pi \approx 7 \times 10^{-20}$ Hz with $g(x) = 1 - \frac{2}{\pi}\,\text{Si}(x)$, $g(1) = 0.398$ — finite, sub-causal, unobservable. And the lock $a_0/(c\,\Gamma_{\text{total}}) = 2/g(1) \approx 5.03$ cancels all cosmological input: a consistency certificate that both sectors are one horizon clock. The decisive tests remain in the laboratory (Chapter 15).

# 15

# Falsification

**Synopsis.** A theory earns its keep at the bench. Everything in Part III has pointed at one number — the rate at which a superposed mass loses coherence to its own gravitational field — and this chapter hands that number to an experimentalist as a protocol she can cost and run. We show that the gravitational rate is an *uncoolable floor*, separate it from ordinary decoherence by three fingerprints, and work out, every figure from the constants, the five-sigma envelope: a 1.5 femtogram silica sphere split by a micrometre decoheres in about 0.59 seconds, and roughly 37 interferometer runs distinguish that from "nothing happened." We give the entanglement-witness test and its clean magic-angle null, the sharp discriminators between this theory and spontaneous-collapse models, and five concrete results — each with a timeline — that would kill the theory outright. We end with the honest ledger and the one limitation that matters most: the theory predicts no number that was not, in some form, available before it. Its contribution is unification, and its virtue is that it names its own executioner.

**You will be able to.** state the three laboratory signatures that separate gravitational from environmental decoherence; compute the five-sigma run count from a visibility deficit and shot noise; read the BMV square-law and its magic-angle zero; list the five kill-switches and the discriminators against Diósi–Penrose collapse; recite the status of every headline claim.

**Prerequisites.** Chapters 1 and 2 (visibility, decoherence, the environment overlap); the rate $\Gamma_{\text{dec}} = C\,E_{\text{G}}/\hbar$ with $C \in [2/\pi, 1]$ and the Gaussian onset of Chapter 12; the honest B+ status and the open construction R1 of Chapter 13; the cosmological scale $a_0 = cH_0/2\pi$ of Chapter 14.

## 15.1 The floor that cannot be cooled

Every decoherence source we met in Chapter 2 is, in the end, a choice. Gas molecules can be pumped away; blackbody photons can be frozen out; vibrations can be isolated. The visibility law $\mathcal{V} = |\langle E_L|E_R\rangle|$ of (2.12) makes the point exactly: the environment records which-path information at a rate set by how often it looks and how much it learns, and both are contingent on the apparatus. Push the vacuum harder, lower the walls' temperature, and the environmental timescale $\tau_{\text{env}}$ of (2.14) climbs without limit, in principle to infinity.

Gravity is the exception, and the whole experimental case rests on why. The mass in superposition sources a gravitational field, and Chapter 11 showed that a constraint of the theory — the vanishing of the total Hamiltonian — ties that field rigidly to where the mass sits, with no gauge freedom left to hide the difference. The two branches carry two genuinely distinct field configurations, $|E_L\rangle$ and $|E_R\rangle$ are two coherent states of that field, and their overlap decays at a rate fixed by the branches' gravitational self-energy difference alone:

$$\Gamma_{\text{grav}} = C\,\frac{E_G}{\hbar} = C\,\frac{GM^2}{\hbar\, d}, \qquad C \in \left[\tfrac{2}{\pi}, 1\right], \qquad \tau_{\text{grav}} = \frac{1}{\Gamma_{\text{grav}}} = \frac{1}{C}\,\frac{\hbar d}{GM^2}. \tag{15.1}$$

The only quantities on the right are the mass, the separation, and constants of nature. No vacuum, no cryostat, no vibration isolation appears — because the field the record lives in is sourced by the very mass whose superposition we are trying to protect. You cannot pump away an object's own gravity. This is the sense in which (15.1) is a *floor*: beneath all the removable noise sits a rate that engineering cannot lower at fixed $M$ and $d$. The experimental programme of this chapter is the hunt for that floor, in the one corner of the mass ladder where it rises above everything else before the object gets too heavy to superpose at all.

***Q 15.1 "How can you possibly separate a gravitational effect from ordinary decoherence? Every real experiment is drowning in gas collisions and photon scattering."***

By its *shape*, not its size. The removable sources and the gravitational floor obey different laws — in mass, in separation, and in the very time profile of the decay — and the next section lays out three fingerprints, any one of which distinguishes them. The strategy is not to eliminate environmental decoherence (impossible) but to push it, source by source, until its timescale $\tau_{\text{env}}$ exceeds the gravitational $\tau_{\text{grav}}$, so that the *first* thing the fringes feel as they fade is the floor. Exercise 2.2 of Chapter 2 already showed the target: get $\tau_{\text{env}} \gg \tau_{\text{grav}}$, and the residual decay carries gravity's signature or it carries nothing — and "nothing" is itself a falsification, as §15.6 makes precise.

## 15.2 Three fingerprints

To read gravity off a decay curve we need laws it obeys that no removable source shares. There are three, and they are independent: an experiment that confirmed even one would be strong evidence, and confirming all three would be decisive.

### 15.2.1 Fingerprint one: the mass law

The gravitational rate (15.1) scales as $\Gamma_{\text{grav}} \propto M^2$. That alone does not distinguish it — collisional and blackbody decoherence also grow with mass. What distinguishes it is the *combination* with separation and with material, and we must be honest about a subtlety in the mass law itself.

Equation (15.1) treats the mass as a point. A real sphere is not a point: when the superposition separation $d$ is smaller than the sphere's own radius $R$, the two branches overlap and the self-energy difference is reduced. The honest self-energy of a uniform sphere of density $\rho$, radius $R = (3M/4\pi\rho)^{1/3}$, split by $\Delta x$, is

$$E_G(\Delta x) = \frac{GM^2}{R}\, f(s), \quad s = \frac{\Delta x}{R}, \quad f(s) = \begin{cases} \frac{1}{2}s^2 - \frac{3}{16}s^3 + \frac{1}{160}s^5, & s \le 2, \\ \frac{6}{5} - \dfrac{R}{\Delta x}, & s \ge 2, \end{cases} \tag{15.2}$$

which rises quadratically as the branches begin to separate and, once they fully clear ($s \ge 2$), approaches the plateau $\frac{6}{5}\, GM^2/R$ with a residual $-GM^2/\Delta x$ correction — the magnitude of that separation-dependent piece is the point-mass scale of Chapter 11, the convention in which the headline $\tau_{\text{dec}} = \hbar d/GM^2$ is quoted. In the *saturated* regime that the near-term experiments actually occupy — the branches barely clearing each other, $\Delta x \gtrsim 2R$ — the self-energy sits at $E_{\text{G}} \approx \frac{6}{5}\,GM^2/R$, and since $R \propto M^{1/3}$ this gives a finite-size floor

$$\Gamma^{\text{sat}}_{\text{grav}} \;\propto\; \frac{M^2}{R} \;\propto\; \rho^{1/3}\,M^{5/3}, \qquad \tau^{\text{sat}}_{\text{grav}} \;\propto\; \rho^{-1/3}\,M^{-5/3}. \tag{15.3}$$

So the mass law an experiment measures is not the naive $M^{-2}$ of the point formula but $M^{-5/3}$ once finite size dominates — a distinction worth stating plainly, because an experimentalist fitting a scaling exponent needs to know which regime she is in. The material enters only through the density $\rho$ (a mild $\rho^{1/3}$), which is itself a test: gravity couples to mass-energy alone, so at fixed $M$ and $\Delta x$ the rate is *composition-independent* beyond that density factor — unlike collisional decoherence, which depends on polarizability and cross-section.

### 15.2.2 Fingerprint two: the inverse separation law

This is the decisive one, because it runs backwards. Every environmental source decoheres a *wider* superposition *faster*: a bigger which-path displacement is easier for a gas molecule or a photon to resolve. Blackbody scattering grows as $d^2$, collisional decoherence as $d^2$; wider is always worse. Gravity does the opposite. In the point regime (15.1), $\Gamma_{\text{grav}} \propto 1/d$: pull the branches apart and the gravitational rate goes *down*, the coherence lives *longer*.

> **The sign that only gravity carries.** Environmental decoherence rate grows with separation ($\propto d^2$). Gravitational decoherence rate *falls* with separation ($\propto 1/d$ for a point, monotonically decreasing in the finite-size formula (15.2) once $\Delta x > R$). A measured decoherence rate that *decreases* as the superposition is widened, at fixed mass, cannot be any known environmental effect. It is the cleanest qualitative fingerprint in the whole programme.

The reason traces straight back to Chapter 11: the gravitational self-energy that distinguishes the branches is a *binding* energy, largest when the two mass configurations are closest and their fields most different per unit displacement, smaller when they are far apart. Chapter 1 flagged this as the feature that would surprise you until the constraint explained it; here it becomes an experimental handle no background can mimic.

### 15.2.3 Fingerprint three: the Gaussian onset

The first two fingerprints are about how the rate depends on the setup. The third is about the *shape of the decay in time*, and it separates this theory not from gas but from its closest rival, spontaneous collapse.

Chapter 2 proved, in the solvable dephasing model (2.15), that any decoherence generated by unitary evolution of the system plus its environment begins *quadratically* in time: the visibility opens as

$$\mathcal{V}(t) \;=\; \exp\!\Big[-\tfrac{1}{2}\,\Gamma^2\,t^2 + O(t^3)\Big] \qquad \text{(short times)}, \tag{15.4}$$

not as $e^{-\Gamma t}$. Nothing can decay linearly at $t = 0$ if the global evolution is unitary, and the gravitational mechanism of Part III is unitary at the level of system-plus-field — so it inherits this Gaussian onset. A genuine collapse model, by contrast, adds a stochastic term *by hand* that makes the visibility decay exponentially from the very first instant. The two curves share a headline timescale but differ in their short-time curvature, and Chapter 12 built this difference into an operational discriminator: form the ratio $\Gamma(2t)/\Gamma(t)$ of the decay exponent at two times. For a Gaussian onset $\Gamma(t) \propto t^2$ the ratio is 4; for exponential decay $\Gamma(t) \propto t$ it is 2. Measure that ratio at early times and you read the mechanism off directly, with no absolute calibration of the rate required.

**Worked Example 15.1 — Reading the mechanism from a curvature ratio**

Suppose an interferometer measures visibility at two hold times, $t$ and $2t$, both well inside the coherence window, and reports the accumulated decay exponent $\Gamma(t) = -2\ln\mathcal{V}(t)$ (Chapter 12's convention, $\mathcal{V} = e^{-\Gamma/2}$). Two hypotheses:

$$\text{Gaussian (QGC):}\quad \Gamma(t) = \lambda^2 t^2, \qquad \text{Exponential (collapse):}\quad \Gamma(t) = \lambda t.$$

The ratio at doubled time is

$$\frac{\Gamma(2t)}{\Gamma(t)} = \frac{\lambda^2(2t)^2}{\lambda^2 t^2} = 4 \qquad \text{versus} \qquad \frac{\lambda\,(2t)}{\lambda\, t} = 2.$$

A measured ratio of 4 says the decoherence was generated by unitary entanglement with a field — QGC's mechanism. A ratio of 2 says the visibility was decaying exponentially from $t = 0$ — a stochastic collapse. The test needs only *relative* visibilities at two times; the absolute rate, the mass calibration, and the coefficient $C$ all cancel. This is the sharpest model-independent discriminator the programme offers, and it does not even require reaching five-sigma on the rate itself.

▶ Run it: `TOOLS/verify_falsification_5sigma.py` `(envelope row supplies the` $\Gamma$ `scale; onset ratio is model structure)`

**Q 15.2 *“Isn’t this just Diósi–Penrose with extra steps? The timescale $\hbar d/GM^2$ is theirs, not yours.”***

The timescale is theirs, and Chapter 1 said so plainly: it is the Diósi–Penrose scale, proposed on heuristic grounds decades before this framework existed. What differs is everything around the number. Diósi–Penrose is a *collapse* model: it modifies quantum mechanics, adding a stochastic term that localizes the wavefunction and, as a side effect, injects energy — the object heats up. QGC adds nothing to quantum mechanics; the evolution stays exactly linear and unitary, the coherence is not destroyed but relocated into the gravitational field (Chapter 2's lesson, verbatim), and there is no anomalous heating. So the two agree on the headline $\hbar d/GM^2$ and disagree on (i) the decay shape — Gaussian onset versus exponential, the ratio test above; (ii) the coefficient — $C \in [2/\pi, 1]$ versus Diósi–Penrose's specific value; and (iii) energy conservation — exact versus violated. §15.5 lays out all three as experimental tests. The honest summary: QGC did not invent the number, it re-derived it from an axiom set that also yields the cosmology of Chapter 14, and kept quantum mechanics intact while doing so.

## 15.3 The five-sigma envelope, every number worked

We now cost the experiment. The question is concrete: what mass, what separation, what hold time, and how many runs suffice to distinguish "gravitational decoherence at the predicted rate" from "no gravitational decoherence" at five standard deviations? Everything below recomputes from $G$, $\hbar$, $k_{\mathrm{B}}$ and the uniform-sphere self-energy (15.2); the verification script that locks these numbers is named at the end.

### 15.3.1 The statistics of a visibility deficit

The two hypotheses make different predictions for the fringe visibility after a hold time $t$. Under the gravitational floor, $\mathcal{V}(t) = e^{-t/\tau_{\mathrm{grav}}} < 1$; under the null hypothesis — the perturbative $G^2$ prediction, whose rate is smaller by the factor $\sim 3 \times 10^{34}$ of Chapter 12 — the visibility stays effectively $\mathcal{V} \approx 1$ over any laboratory time. The single-run discriminant is therefore the *deficit*

$$\delta \;=\; 1 - \mathcal{V}(t) \;=\; 1 - e^{-t/\tau_{\mathrm{grav}}}. \tag{15.5}$$

A visibility estimate from $N$ runs carries shot noise: fringe contrast is estimated from photon or detection counts, and the standard error on $\mathcal{V}$ scales as $1/\sqrt{N}$ (a Bernoulli count, order unity per run in the idealized single-quantum limit). To resolve the deficit at five sigma we need the signal to exceed five times the noise,

$$\delta \;\geq\; \frac{5}{\sqrt{N}} \qquad \Longrightarrow \qquad \boxed{N_{5\sigma} \;=\; \frac{25}{\delta^2} \;=\; \frac{25}{(1-\mathcal{V})^2}.} \tag{15.6}$$

This is an idealized floor: a real per-run contrast $c < 1$ multiplies it by $1/c^2$, so (15.6) is the best case, the number of runs against which every practical estimate is compared. Its shape already tells the strategy: you want the deficit $\delta$ as large as possible, which means holding for a time comparable to $\tau_{\mathrm{grav}}$, which means choosing the mass so that $\tau_{\mathrm{grav}}$ is neither absurdly short (the object decoheres before you can measure) nor absurdly long (the deficit is unresolvable). That optimization is the envelope.

### 15.3.2 The envelope point

**Worked Example 15.2 — The 1.5 femtogram silica sphere**

Take a uniform silica sphere, $\rho = 2200$ kg/m$^3$, of mass $M = 1.5 \times 10^{-15}$ kg (1.5 femtograms), split by $\Delta x = 1\ \mu\mathrm{m} = 10^{-6}$ m, held coherent for $t = 1$ s. First its radius:

$$R = \left(\frac{3M}{4\pi\rho}\right)^{1/3} = \left(\frac{3(1.5 \times 10^{-15})}{4\pi(2200)}\right)^{1/3} = 5.46 \times 10^{-7}\ \mathrm{m}, \qquad s = \frac{\Delta x}{R} = 1.83.$$

Since $s < 2$ we are just below saturation; the self-energy from (15.2) with $f(1.83) = \frac{1}{2}(1.83)^2 - \frac{3}{16}(1.83)^3 + \frac{1}{160}(1.83)^5 = 0.654$ is

$$E_{\mathrm{G}} = \frac{GM^2}{R} f(s) = \frac{(6.67430 \times 10^{-11})(1.5 \times 10^{-15})^2}{5.46 \times 10^{-7}} (0.654) = 1.80 \times 10^{-34}\ \mathrm{J}.$$

The gravitational timescale (taking the natural $C = 1$) is

$$\tau_{\mathrm{grav}} = \frac{\hbar}{E_{\mathrm{G}}} = \frac{1.054571817 \times 10^{-34}}{1.80 \times 10^{-34}} = 0.586 \text{ s} \;\approx\; 0.59 \text{ s}.$$

After one second of holding, the visibility has fallen to

$$\mathcal{V}(1 \text{ s}) = e^{-1/0.586} = e^{-1.707} = 0.182, \qquad \delta = 1 - 0.182 = 0.818.$$

That is a large deficit — the fringes have lost four-fifths of their contrast — so the run count from (15.6) is small:

$$N_{5\sigma} = \frac{25}{(0.818)^2} = \frac{25}{0.669} = 37 \text{ runs}.$$

Thirty-seven interferometer runs, at the idealized shot-noise floor, separate "gravity decohered the sphere" from "nothing happened" at five sigma. This is a fundable protocol, not a thought experiment.

▶ RUN IT: `TOOLS/verify_falsification_5sigma.py` (CHECK 1-2: $\tau_{\mathrm{grav}} \approx 0.59$ s, $\mathcal{V} \approx 0.18$, $N_{5\sigma} \approx 37$)

The one demand the envelope makes of the apparatus is that *environmental* decoherence stay slower than the floor over the one-second hold. Setting the residual-gas rate $\Gamma_{\mathrm{gas}} = n\,\sigma\,\bar{v}$ (cross-section $\sigma = \pi R^2$, number density $n = P/k_{\mathrm{B}}T$, mean speed $\bar{v} = \sqrt{8k_{\mathrm{B}}T/\pi m_{\mathrm{gas}}}$) below $\Gamma_{\mathrm{grav}} = 1/\tau_{\mathrm{grav}}$ fixes a maximum pressure

$$P_{\max} = \frac{k_{\mathrm{B}}T}{\pi R^2\,\bar{v}\,\tau_{\mathrm{grav}}} \;\approx\; 7 \times 10^{-15} \text{ mbar} \quad (T = 4 \text{ K},\ \text{helium background}), \tag{15.7}$$

a cryogenic extreme-high-vacuum requirement — demanding, at the edge of current levitated-optomechanics capability, but not fantastical. This is the MAQRO-class regime, and it is why the honest timeline for the decisive measurement is years, not decades: roughly 2028–2035.

**Worked Example 15.3 — Why the window is narrow: the threshold mass**

The envelope is not arbitrary — it sits just above a sharp threshold. In the saturated regime, $\tau_{\mathrm{grav}} \propto M^{-5/3}$ by (15.3), so lighter spheres decohere far more slowly. Demanding $\tau_{\mathrm{grav}} = 1$ s for a silica sphere in the saturated ($\Delta x \gg R$) limit, $E_{\mathrm{G}} = \frac{6}{5}GM^2/R$, gives

$$M_* = \left[\frac{\hbar}{\frac{6}{5}\,G\,t_{\mathrm{target}}}\left(\frac{3}{4\pi\rho}\right)^{1/3}\right]^{3/5} = 7.6 \times 10^{-16} \text{ kg}.$$

Below $M_*$ the visibility deficit collapses: at $M = 10^{-18}$ kg (a current Vienna-class levitated sphere) the point-formula timescale is astronomically long, $\mathcal{V} \approx 1$, and (15.6) demands $N_{5\sigma} > 10^{12}$ runs — out of reach. Above $M_*$ the deficit turns on sharply. The decisive mass window is a factor of a few wide, centered a whisker above $M_* \approx 8 \times 10^{-16}$ kg: heavy enough to gravitate measurably, light enough to hold coherent for a second. That the window exists at all, at an accessible mass, is the entire experimental opportunity of this book.

▶ RUN IT: `TOOLS/verify_falsification_5sigma.py` (CHECK 5, 7: $M_* \approx 7.6 \times 10^{-16}$ kg; $\tau_{\rm grav} \propto M^{-5/3}$)

| **Platform** | $M$ **[kg]** | $\tau_{\rm grav}$ **[s]** | $\mathcal{V}(1\,\mathrm{s})$ | $N_{5\sigma}$ |
|---|---|---|---|---|
| Envelope (silica, $\Delta x = 1\,\mu$m) | $1.5 \times 10^{-15}$ | 0.59 | 0.18 | $\approx 37$ |
| Marginal fallback ($\Delta x = 0.5\,\mu$m) | $1.0 \times 10^{-15}$ | 2.2 | 0.64 | $\approx 190$ |
| MAQRO-class (fully decoheres) | $1.0 \times 10^{-14}$ | $\ll 1$ | $\approx 0$ | 25 (floor) |
| Vienna, present ($\Delta x = 10\,$nm) | $1.0 \times 10^{-18}$ | huge | $\approx 1$ | $> 10^{12}$ |

The table reads as a ladder: below threshold the test is impossible; at the envelope it costs tens of runs; well above threshold the sphere decoheres completely and the run count hits the absolute floor $N_{5\sigma} = 25$ (the deficit saturates at $\delta \to 1$). The sweet spot — decidable but not trivial — is the envelope row.

**Q 15.3** ***"What if the experiment sees nothing — the fringes survive the full second at full contrast?"***

Then the theory is dead, and that is the point of building it this way. A null result at the envelope — coherence maintained at $\mathcal{V} \approx 1$ when (15.1) predicts $\mathcal{V} \approx 0.18$, with the environment demonstrably controlled below $P_{\rm max}$ — means the gravitational rate is not $G^1$. It would mean the perturbative $G^2$ prediction is right, the coherence time is longer by the factor $3 \times 10^{34}$ of Chapter 12, and the central conjecture $K = 2\pi H_{\rm phys}$ of Chapter 13 is false as an operator equation. This is kill-switch (i) of §15.6. A theory whose central claim can be extinguished by thirty-seven runs of a fundable experiment is doing exactly what a scientific theory should. The honest self-assessment of this framework puts the probability that the null wins at roughly one-in-two; we say so, and we wait for the data.

## 15.4 The entanglement witness and the magic angle

The envelope test of §15.3 watches a single mass lose coherence. A second, complementary experiment watches *two* masses become entangled, and it witnesses something sharper: that gravity itself carries quantum information.

### 15.4.1 Gravitationally-induced entanglement

Place two masses in adjacent interferometers, each in a superposition of two paths, close enough that they interact gravitationally but with no other coupling between them. This is the Bose–Marletto–Vedral (BMV), or QGEM, proposal. The gravitational interaction accumulates a relative phase that depends on which pair of paths the two masses take — the paths that bring them closest wind up more phase than the paths that keep them apart. If gravity is a quantum channel, that path-dependent phase *entangles* the two interferometers: the joint state becomes non-separable, and an entanglement witness measured at the end returns a non-zero value.

The logic is a clean piece of quantum information: a *local operation* (each mass evolving in its own interferometer) plus a *classical channel* between them cannot create entanglement. If the two masses end up entangled, and the only thing connecting them was gravity, then gravity is not a classical channel — it carried quantum information. Measuring the witness is therefore a test of whether the gravitational field must be treated quantum-mechanically at all, which is precisely the question Chapter 1 opened with.

### 15.4.2 The square law and the mass trend

In the saturated (QGEM) regime, the QGC prediction for the residual visibility at the witness time — the visibility that survives the source masses' *own* self-decoherence — follows a square law in the geometric ratio $d/\Delta$ ($d$ the inter-mass separation, $\Delta$ the single-mass wavefunction extent), with a Legendre-$P_2(\cos\theta)$ angular structure:

$$V_{\text{par}} = \exp\Big[-\tfrac{3\pi}{5}\Big(\tfrac{d}{\Delta}\Big)^2\tfrac{d}{R}\Big], \qquad V_{\text{perp}} = V_{\text{par}}^2, \tag{15.8}$$

for the parallel ($\theta = 0$) and perpendicular ($\theta = \pi/2$) arm geometries. The residual mass trend is the striking part: $-\ln V \propto 1/R \propto M^{-1/3}$ — *lighter* masses show *more* suppression. This is opposite in sign to the $M^2$ scaling of Diósi–Penrose and CSL collapse models, and so is a clean discriminant: a mass scan of the witness that finds suppression *growing* toward small mass rules out the collapse family and is consistent with QGC's finite-size self-decoherence.

### 15.4.3 The magic-angle null

The angular structure hides a special geometry. The leading (dipole) contribution to the gravitational phase carries a factor $P_2(\cos\theta) = \frac{1}{2}(3\cos^2\theta - 1)$, which *vanishes* at the magic angle

$$\cos^2\theta_m = \tfrac{1}{3}. \tag{15.9}$$

At exactly this angle the dominant dipole phase is a true zero — not small, not suppressed, but identically absent — and the surviving signal comes only from the higher (quadrupole and beyond) multipoles:

$$V_{\text{magic}} = \exp\Big[-\tfrac{36\pi}{49}\Big(\tfrac{d}{\Delta}\Big)^3 - \tfrac{9\pi}{7}\Big(\tfrac{d}{\Delta}\Big)^5\Big], \tag{15.10}$$

a cubic-plus-quintic instead of the square law of (15.8). A true zero is worth more to an experimentalist than any peak, because it is *background-free*: systematic effects that would fake a gravitational signal generically do not respect the magic angle, so scanning through $\theta_m$ and watching the dipole phase pass through zero isolates the genuine gravitational contribution from everything that merely mimics it. The magic angle is QGC's cleanest single-geometry signature.

**Worked Example 15.4 — Reading the magic angle from a geometry scan**

Fix the mass and sweep the geometric ratio $u \equiv d/\Delta$ at three arm orientations: parallel, perpendicular, magic. From (15.8) and (15.10), at small $u$ the visibilities behave as

$$-\ln V_{\text{par}} \propto u^2 \cdot \tfrac{d}{R}, \qquad -\ln V_{\text{perp}} = 2(-\ln V_{\text{par}}), \qquad -\ln V_{\text{magic}} \propto u^3,$$

so the perpendicular-to-parallel ratio is exactly 2, geometry-fixed and $u$-independent — a parameter-free QGC prediction — while the magic-angle suppression is one power of $u$ weaker, quenched by the dipole null. Converting to Bures information distance ($d_B^2 = 2 - \sqrt{2(1+V)}$) sharpens this into a crossover: the magic-angle and parallel curves cross at $u_* = \sqrt{7/18} \approx 0.62$, a specific dimensionless number a geometry scan at fixed mass can hit. Collapse models, which have no $\theta$-dependence at all, predict none of this angular structure. *Scope*: the crossover location is a leading-order result; the exact-phase magic-angle formula carries a cubic correction that shifts it by a few percent and removes the magic-versus-parallel crossing entirely — the robust, background-free feature is the dipole *null* at $\cos^2\theta_m = 1/3$, not the precise crossover value.

▶ RUN IT: `TOOLS/verify_falsification_5sigma.py (CHECK 6: BMV/QGEM saturates; magic angle from bmv-prediction)`

**Q 15.4 *"If the source masses self-decohere in milliseconds at the QGEM design point, isn't the entanglement gone before you can measure it?"***

At the nominal design point ($\Delta \sim d/2$) yes — the same gravitational self-decoherence of §15.3 swamps the entanglement build-up, and all geometries give $V \approx 0$: the witness dies in milliseconds while the protocol needs seconds. This is not a bug, it is the theory being consistent with itself: the very effect that makes the single-mass test work makes the naive two-mass test hard. The discrimination window requires *widely separated* wavefunction extents, $\Delta \gg d$, where the self-decoherence is weak and the entanglement phase survives. QGC therefore predicts that the standard BMV expectation — entanglement observable at the design point — is *wrong*, because it neglects the source masses' own gravitational self-decoherence. That disagreement is itself a test: whether the witness survives at the design point or only in the $\Delta \gg d$ regime distinguishes QGC from a picture in which gravity entangles but masses do not self-decohere.

## 15.5 QGC versus collapse: the discriminators

The nearest competitor to QGC is not perturbative QFT — which predicts essentially no effect — but the family of spontaneous-collapse models, chiefly Diósi–Penrose (DP), which predict the *same* headline timescale $\hbar d/GM^2$. Because they agree on the number that a first, crude experiment would measure, the discrimination must be sharper. The modular-physical-correspondence (MPC) mechanism of Part III — QGC's route to the rate — differs from DP in three measurable ways, and an experiment that reaches the floor can test all three.

**QGC (MPC) versus Diósi–Penrose (collapse).**

1. *Decay shape.* QGC: Gaussian onset, $-\ln\mathcal{V} \propto t^2$ at short times, ratio $\Gamma(2t)/\Gamma(t) = 4$ (Worked Example 15.1). DP: exponential from $t = 0$, ratio 2.
2. *Coefficient.* QGC: $C \in [2/\pi, 1]$, natural value $C = 1$, a *window* with a physical floor (Margolus–Levitin). DP: a single specific coefficient with no window.
3. *Energy and linearity.* QGC keeps quantum mechanics exactly linear and unitary — no spontaneous localization, no added noise, and crucially *no anomalous heating*: energy is

> conserved. DP is a genuine collapse: it localizes the wavefunction and thereby injects energy, predicting a small but in-principle-measurable heating and, in some variants, spontaneous radiation.

The third discriminator is the most consequential and the most model-free. Collapse models must violate energy conservation — localizing a spread-out wavefunction costs energy, which the stochastic term supplies from nowhere — and this shows up as anomalous heating of the test mass or, in charged variants, spontaneous photon emission that experiments already bound. QGC predicts *none* of this: the coherence is not destroyed but transferred to the gravitational field, an exactly unitary process that conserves energy. An experiment that reaches the decoherence floor and finds the predicted loss of coherence *without* accompanying heating supports QGC over DP; one that finds heating tracking the decoherence supports collapse. Because the heating test does not require resolving the Gaussian onset — only monitoring the object's energy budget — it may be the first discriminator within reach.

## 15.6 The kill-switches

A theory is only as good as the results that would destroy it. Here are five, each a concrete measurement, each of which would falsify QGC outright. They are stated so an experimentalist can aim at them.

**K1.** *Coherence survives the floor.* At the envelope point (Worked Example 15.2), with environmental decoherence demonstrably below $P_{\max}$, the visibility stays at $\mathcal{V} \approx 1$ well beyond $\tau_{\text{grav}} \approx 0.59$ s — say to $100\,\tau_{\text{grav}}$. This is a $G^2$ timescale where QGC demands $G^1$, and it kills the rate directly. Thirty-seven runs decide it.

**K2.** *The rate scales as $G^2$.* A mass-and-separation scan finds the decoherence rate scaling as $G^2$ — equivalently, times longer by the factor $\sim 3 \times 10^{34}$ of Chapter 12 — rather than the $G^1$ of (15.1). This refutes the mechanism, not just the coefficient: perturbative QFT wins, and the constraint-enforced $G^1$ route of Chapter 11 is wrong.

**K3.** *The onset is exponential, not Gaussian.* The short-time decay measured at two hold times gives the ratio $\Gamma(2t)/\Gamma(t) = 2$, not 4 (Worked Example 15.1). Exponential decay from $t = 0$ is the signature of a stochastic collapse term added by hand — it contradicts QGC's unitary mechanism regardless of the rate.

**K4.** *Spontaneous heating appears.* The test mass heats, or radiates, in step with its loss of coherence, consistent with a collapse model's energy injection. QGC conserves energy exactly; measurable heating tied to decoherence falsifies it in favour of collapse (§15.5).

**K5.** *The cosmic scale evolves wrong.* The MOND acceleration scale, read from high-redshift rotation curves or strong lensing, evolves inconsistently with $a_0(z) = c\,H(z)/2\pi$ (Chapter 14). Since the same axiom set fixes both the laboratory rate and the cosmic scale, a cosmological failure impeaches the whole framework, not merely its gravitational-decoherence sector.

> **Timeline.** K1–K4 live in the levitated-optomechanics / matter-wave-interferometry programme, whose decidable window for the $G^1$-versus-$G^2$ discrimination is roughly 2028–2035

(cryogenic extreme-high-vacuum, MAQRO-class platforms). K5 is already partly testable: present high-$z$ rotation-curve and lensing data constrain $a_0(z)$ now, with DESI/Euclid tightening the dark-energy sector ($w = -1$) over the same years.

**Q 15.5** ***"Five kill-switches sounds like a lot of ways to be wrong. Isn't that just a theory hedging its bets?"***

The opposite. A theory with *no* kill-switch is the one hedging — it can absorb any result. QGC's five are not five escape routes but five independent throats to cut, and any single one of them, at five sigma, ends the theory. That they exist, at accessible masses and on a decadal timeline, is the strongest thing this book can say in its own favour. Contrast the usual situation in quantum gravity, where the characteristic scale is fifteen orders of magnitude beyond any conceivable experiment (Chapter 1's Planck wall): there, falsification is not even in principle available. Here it is a funding proposal.

## 15.7 The honest ledger

We end where scientific integrity demands: with an unflinching accounting of what is established and what is not. Every headline claim of this book, with its status.

| Claim | Status | Limiting step |
|---|---|---|
| Energy scale $E_G = GM^2/d$; Margolus–Levitin bound | [PROVEN] | classical gravity / theorem |
| Axiom I/II theorems (BH entropy, Unruh, Einstein eq., Page curve) | [DERIVED] | area law (Axiom I) as input |
| $a_0 = cH_0/2\pi$ (MOND scale) | [DERIVED] | parameter-free; but $\sim 10\%$ level; galactic-dynamics interpolation phenomenological |
| Dark energy $\rho_{DE} = \alpha c^2 H^2/G$ | [DERIVED] | coefficient $\alpha$ fitted, not predicted |
| $G^1$ scaling (not $G^2$) | [DERIVED] | linearized gravity, coherent-state |
| Coefficient $C \in [2/\pi, 1]$ in $\Gamma_{dec} = C\, E_G/\hbar$ | [DERIVED] | window only; not uniquely pinned |
| Rate $= C\, E_G/\hbar$; operator identity $K = 2\pi H_{phys}$ | [CONJECTURED] | reduced to R1 (Chapter 13) |
| $C = 1$ (natural, Diósi–Penrose reading) | [CONJECTURED] | clock-model normalization |
| Born rule | [ASSUMED] | as in all of quantum mechanics |

The proven content is real and load-bearing: the energy scale $E_G = GM^2/d$ is classical gravity, and the Margolus–Levitin bound that turns an energy into a minimum time is a theorem.

The derived content — the axiom-based thermodynamics of Part III, the parameter-free MOND scale of Chapter 14 — is rigorous within its stated inputs. The conjectural content is one thing: the operator identity $K = 2\pi H_{\text{phys}}$ that upgrades the energy scale into an actual decoherence *rate*. It is proven in expectation values and in every solvable model, open as a full operator equation, and the entire residual uncertainty is isolated in the single well-posed construction R1 of Chapter 13. That is why the grade is **B+**: a proven scale, a conjectured rate, a clean experimental arbiter, and a labelled path to A.

### 15.7.1 The single biggest limitation

Now the candour that this book has insisted on from its first page, stated without softening.

> **No prediction of QGC is unavailable from a pre-existing theory.** The gravitational decoherence scale is Diósi–Penrose (1987/1996). The MOND phenomenology is Milgrom (1983). Holographic dark energy is Li (2004). The Page curve predates the framework. Every testable number QGC issues was, in some form, on the table before QGC existed.

This is the honest core of the whole enterprise, and it must not be dressed up. QGC's contribution is not a brand-new number that only it could produce; it is *unification* — the demonstration that one axiom set yields both the laboratory decoherence scale and the cosmic acceleration scale, with quantum mechanics kept exactly linear and with sharp discriminators against the collapse models it superficially resembles. Unification is a weaker claim than novel prediction, and this book presents it as such. The internal audit confirms it: the one structurally novel quantity the framework contains — the lock $a_0/(c\,\Gamma_{\text{tot}})$ between the cosmic acceleration scale and the cosmic decoherence rate — is internally rigid but not externally testable, because both sides are built from independently measured cosmic quantities. And the "five converging arguments" for $G^1$ reduce, on inspection, to a single physical input: the Wheeler–DeWitt constraint of Chapter 11. Five framings, one fact.

Why insist on this so loudly? Because it is the most scientifically load-bearing feature of the book. A framework that quietly claimed novelty it did not have would be worse than useless — it would be misleading. A framework that says, in the open, "here is the number, here is who found it first, here is what I add (a common root and a sharp test), and here is the experiment that will kill me if I am wrong" is doing science. The honesty is not a disclaimer appended to the theory; it is part of the theory's content.

**Q 15.6** ***"If QGC predicts nothing new, why should anyone care whether the experiment confirms it?"***

Because a confirmed *unification* is worth more than an isolated number. That the Diósi–Penrose scale and the MOND scale fall out of the same three axioms — one for the laboratory, one for the cosmos — is a nontrivial claim about the structure of physics, and the experiment tests *that structure* through its sharpest consequence. If the floor is found, at the predicted rate, with the Gaussian onset and without heating, then the unification is not a coincidence of formulas but a fact about nature, and the axioms of Part III describe something real. If it is not found, the unification was a beautiful mistake, and we will have learned that too — at the cost of thirty-seven runs. Either way the money is well spent. That is the difference between a theory that touches the ground and one that cannot.

## 15.8 What to measure, in order

We close the book the way Einstein closed his 1905 electrodynamics paper: with the theory cashed out as a numbered list an experimentalist can act on, in priority order.

1. **The floor, at the envelope point.** A levitated silica sphere, $M \approx 1.5 \times 10^{-15}$ kg, split by $\Delta x \approx 1\ \mu$m, held for $\sim 1$ s at cryogenic extreme-high-vacuum ($P < 7 \times 10^{-15}$ mbar). Measure the visibility. QGC predicts $\mathcal{V} \approx 0.18$; the null predicts $\mathcal{V} \approx 1$. Roughly 37 runs decide it at five sigma. *This is the experiment.* (Worked Example 15.2; kill-switches K1, K2.)

2. **The decay shape.** At the same platform, measure visibility at two hold times and form $\Gamma(2t)/\Gamma(t)$. QGC (Gaussian onset) gives 4; collapse gives 2. Model-free; needs no absolute calibration. (Worked Example 15.1; kill-switch K3.)

3. **The energy budget.** Monitor the sphere's heating as it decoheres. QGC predicts no anomalous heating (exact energy conservation); collapse predicts heating tracking the decoherence. Possibly the first discriminator within reach. (§15.5; kill-switch K4.)

4. **The separation law.** Vary $\Delta x$ at fixed $M$. QGC predicts the rate *falls* as the split widens (the inverse-separation fingerprint); every environmental source predicts it rises. A qualitative sign flip no background can fake. (§15.2.)

5. **The entanglement witness and the magic angle.** Two-mass BMV/QGEM in the $\Delta \gg d$ regime, scanning through the magic angle $\cos^2\theta_m = 1/3$ where the dipole phase is a true zero. Background-free confirmation that gravity carries quantum information; the $M^{-1/3}$ mass trend distinguishes QGC from collapse. (§15.4; Worked Example 15.4.)

6. **The cosmic cross-check.** Test $a_0(z) = c\,H(z)/2\pi$ against high-redshift rotation curves and strong lensing; watch $w = -1$ in DESI/Euclid. A cosmological failure impeaches the shared axiom set. (Chapter 14; kill-switch K5.)

And the verdict. This book began, in Chapter 1, with an impasse: two flawless rulebooks, quantum mechanics and general relativity, falling jointly silent on one bench-top question — where does a mass in two places gravitate? It ends with that question turned into a protocol: a femtogram sphere, a beam splitter, a cryostat, and thirty-seven runs. The theory that lives at the seam is graded honestly at B+ — a proven energy scale, a conjectured rate, one open construction, and no prediction that predates it except the unification itself. What it offers in exchange for that modesty is rare: it hands the experimentalist its own executioner, with a date (2028–2035) and an address (a levitated nanosphere at cryogenic vacuum). Few theories of gravity can say as much. Most cannot be reached at all. This one can be reached, and reached soon, and that — not any single number — is the wager of this book, now placed on the laboratory bench where it belongs.

## Chapter FAQ

**Q: The envelope needs $7 \times 10^{-15}$ mbar and a one-second coherent superposition of a femtogram sphere. Is that realistic, or is "fundable" doing a lot of work?** It is at the edge, not beyond it. Levitated-optomechanics groups already trap and cool nanospheres in this mass range; the demanding parts are the coherent $\mu$m-scale splitting and the cryogenic extreme-high-vacuum, both of which are active engineering targets rather than new physics. The honest timeline is 2028–2035, and the marginal-fallback row of the table (§15.3: $M = 10^{-15}$ kg, $\Delta x = 0.5\ \mu$m, $\sim 190$ runs) gives a decidable test even if the envelope point proves too hard. The programme degrades gracefully; it does not hinge on a single heroic parameter.

**Q: You keep contrasting QGC with Diósi–Penrose, but both give $\hbar d/GM^2$. If an experiment measures that timescale, hasn't it confirmed *both*?** It has confirmed the *scale*, which both share, and which was the Diósi–Penrose contribution historically. To separate the two you need the finer discriminators of §15.5: the Gaussian-versus-exponential onset (the $\Gamma(2t)/\Gamma(t)$ ratio, 4 versus 2), the coefficient window ($C \in [2/\pi, 1]$ versus DP's single value), and above all the presence or absence of anomalous heating (unitary QGC conserves energy; collapse does not). A first experiment confirms the scale; a careful one adjudicates the mechanism.

**Q: What if the experiment is inconclusive — systematics dominate, and neither hypothesis is cleanly resolved?** That is the most probable frustrating outcome, and the framework's own honest assessment weights it substantially. The defence is redundancy: three independent fingerprints (§15.2), plus the magic-angle null, plus the cosmic cross-check, so that no single systematic can fake all of them. The inverse-separation sign flip and the magic-angle zero are especially robust precisely because they are *qualitative* — a systematic that grows with $\Delta x$ cannot mimic a rate that falls with $\Delta x$, and a background that ignores the magic angle cannot fake a dipole null. Inconclusive on the rate does not mean inconclusive on the mechanism.

**Q: The ledger lists the rate as [CONJECTURED]. Doesn't that make the whole falsification programme premature — shouldn't the theory be proven before it is tested?** No — and this inverts how physics actually works. The conjecture is exactly what the experiment tests. If the rate were already proven, the measurement would be a formality; because it is conjectured, reduced to the single construction R1 (Chapter 13), the measurement is decisive. A theory proves what it can and then hands the rest to nature. The energy scale is proven by mathematics; the rate will be decided by a levitated sphere. Both are proper parts of doing science, and the honest labelling of which is which is the point of the ledger.

## Exercises

**Exercise 15.1.** An interferometer holds the envelope sphere ($\tau_{\text{grav}} = 0.59$ s) but for only $t = 0.3$ s per run, with a realistic per-run fringe contrast $c = 0.5$. (a) Compute the visibility $\mathcal{V}(0.3\text{ s})$ and the deficit $\delta$. (b) Compute the idealized run count $N_{5\sigma} = 25/\delta^2$, then correct it for the contrast, $N = N_{5\sigma}/c^2$. (c) Compare with the 37 runs of Worked Example 15.2 and comment on why holding longer and improving contrast both matter.

**Exercise 15.2.** (a) Using the saturated finite-size scaling $\tau_{\text{grav}} \propto M^{-5/3}$ (Eq. (15.3)), by what factor does $\tau_{\text{grav}}$ change if the sphere mass is doubled from the envelope value? (b) Contrast with the naive point-mass scaling $\tau_{\text{grav}} \propto M^{-2}$: by what factor would *that* predict? (c) An experimentalist measures the decoherence-time exponent and finds $\tau_{\text{grav}} \propto M^{-1.7\pm0.1}$. Which regime is she in, and what does this tell her about the ratio $\Delta x/R$ of her superpositions?

## Solutions

*Solution (Exercise 15.1).* (a) $\mathcal{V}(0.3) = e^{-0.3/0.586} = e^{-0.512} = 0.599$, so the deficit is $\delta = 1 - 0.599 = 0.401$ — smaller than the 0.818 of the full-second hold, because less time has passed. (b) Idealized: $N_{5\sigma} = 25/(0.401)^2 = 25/0.161 = 155$ runs. Corrected for contrast $c = 0.5$: $N = 155/(0.5)^2 = 155/0.25 = 622$ runs. (c) Against the 37 runs of the ideal full-second, unit-contrast case, this is a factor $\sim 17$ more runs — a factor $\sim 4$ from the shorter hold (smaller deficit) and a factor 4 from the halved contrast. Both levers matter multiplicatively: the deficit enters as $\delta^{-2}$ and the contrast as $c^{-2}$, so an experiment that both holds to the full $\tau_{\text{grav}}$ *and* maximizes fringe contrast is exponentially cheaper in runs. This is why the envelope point — chosen so that one second is comparable to $\tau_{\text{grav}}$, giving a near-maximal deficit — is the sweet spot.

▶ RUN IT: `TOOLS/verify_falsification_5sigma.py (section N: ch15 run-count checks)`

*Solution (Exercise 15.2).* (a) Doubling $M$ with $\tau_{\text{grav}} \propto M^{-5/3}$ multiplies $\tau_{\text{grav}}$ by $2^{-5/3} = 0.315$: the heavier sphere decoheres about 3.2 times *faster* (shorter $\tau_{\text{grav}}$). (b) The naive point-mass law $\tau_{\text{grav}} \propto M^{-2}$ would give $2^{-2} = 0.25$, a factor 4 faster — steeper than the finite-size law. The gap between $2^{-5/3}$ and $2^{-2}$ is exactly the finite-size correction: because a heavier sphere is also *bigger* ($R \propto M^{1/3}$), its self-energy grows less steeply than $M^2$, softening the exponent from 2 to 5/3. (c) A measured exponent $-1.7 \pm 0.1$ is consistent with $-5/3 = -1.67$ and inconsistent with $-2$: she is in the *saturated* finite-size regime, with $\Delta x \lesssim 2R$ (the branches barely clearing each other). This is itself a fingerprint — the exponent encodes the geometry of the superposition, and finding 5/3 rather than 2 confirms she is probing the object's own gravitational self-energy rather than a point-source idealization.

▶ RUN IT: `TOOLS/verify_falsification_5sigma.py (CHECK 7:` $\tau_{\text{grav}} \propto M^{-5/3}$ `saturated scaling)`

**What to carry forward**

1. Gravitational decoherence, $\Gamma_{\text{grav}} = C\,GM^2/\hbar d$, is an *uncoolable floor* — sourced by the mass's own field, removable by no vacuum or cryostat — with three laboratory fingerprints: the finite-size mass law ($\tau_{\text{grav}} \propto M^{-5/3}$, not the naive $M^{-2}$), the *inverse* separation law (rate falls as the split widens, opposite to all environmental decoherence), and the Gaussian onset ($\Gamma(2t)/\Gamma(t) = 4$, not 2).

2. The five-sigma envelope is fundable and worked from the constants: a 1.5 fg silica sphere split by 1 $\mu$m decoheres in $\tau_{\text{grav}} \approx 0.59$ s, so $\mathcal{V}(1\,\text{s}) \approx 0.18$ and $N_{5\sigma} = 25/(1-\mathcal{V})^2 \approx 37$ runs distinguish it from the no-effect null (cryogenic XHV, 2028–2035). The BMV magic-angle null ($\cos^2\theta_m = 1/3$) gives a background-free witness that gravity carries quantum information.

3. Five kill-switches would end the theory (coherence past $\tau_{\text{grav}}$; $G^2$ scaling; exponential onset; spontaneous heating; $a_0(z) \neq cH(z)/2\pi$). The honest ledger: $E_{\text{G}}$ [PROVEN], the rate/$K = 2\pi H_{\text{phys}}$ [CONJECTURED] (reduced to R1), MOND [DERIVED] at $\sim 10\%$, dark energy $\alpha$ fitted, Born rule [ASSUMED] — grade **B+**. The single biggest limitation, stated openly: no QGC prediction is unavailable from pre-existing theory; the contribution is unification, and naming its own executioner is the most load-bearing feature of the book.

# Part V

# Appendices

# A

# Notation and Glossary

*Every symbol and term of art in this volume, with the place it is first defined. Canonical macro source: `VERIFIED/papers/common/qgc-notation.sty`. A status tag on a row records the honest standing of the object it names; the single-page summary of all such tags is Appendix E.*

## Conventions

Natural logarithms throughout; entropies in nats (1 bit $= \ln 2$ nats), $k_B$ restored only at laboratory temperatures. Metric signature $(-,+,+,+)$ (from Chapter 3 on). Operators wear hats only where confusion threatens ($\hat{T}_{\mu\nu}$, $\hat{\rho}$ vs. classical counterparts); states are $|\psi\rangle$ (pure) or $\rho$ (general). SI units with all constants visible on each formula's face — $G$, $\hbar$, $c$ are never set to one except in clearly-quarantined power-counting arguments. The Einstein summation convention (repeated upper–lower indices summed 0 to 3) is in force from Chapter 3; Greek indices $\mu, \nu, \ldots$ run over spacetime, Latin $i, j, \ldots$ over space. Two symbols carry a deliberate, standard clash of letters, disambiguated by context: $\alpha$ is both a coherent-state amplitude and a Bogoliubov coefficient (Chapter 4); $K$ is the null-congruence's role and the modular Hamiltonian (Chapter 7).

## Symbols

| Symbol | First defined | Meaning |
|---|---|---|
| *Constants and scales* | | |
| $G,\ \hbar,\ c$ | Ch. 1, WE 1.2 | the three input constants ($6.67430 \times 10^{-11}$, $1.054571817 \times 10^{-34}$, 299792458 SI) |
| $k_B$ | Ch. 4 | Boltzmann constant, restored at laboratory temperatures |
| $m_P,\ \ell_P,\ t_P$ | (1.10)–(1.11) | Planck mass $\sqrt{\hbar c/G}$, length $\sqrt{\hbar G/c^3}$, time $\ell_P/c$ |
| $E_G$ | (1.14) | gravitational self-energy of a mass split, $GM^2/d$ [PROVEN] as the seam's energy scale |
| $\tau_{dec}$ | (1.12) | gravitational decoherence time $\hbar d/GM^2 = \hbar/E_G$; the rate is [CONJECTURED] |
| *Geometry and curvature (Ch. 3)* | | |
| $g_{\mu\nu},\ \eta_{\mu\nu}$ | (1.1), (3.7) | spacetime metric; flat (Minkowski) metric diag$(-1,+1,+1,+1)$ |

*continued on next page*

| Symbol | First defined | Meaning |
|---|---|---|
| $ds^2,\ \tau$ | (3.7) | line element; proper time, $-c^2 d\tau^2 = ds^2$ |
| $\Gamma^\rho{}_{\mu\nu}$ | (3.10) | Christoffel symbols; the removable (frame-dependent) part of gravity — "$\Gamma$ is how gravity looks" |
| $\nabla_\mu$ | (3.12) | covariant derivative |
| $R^\rho{}_{\sigma\mu\nu}$ | (3.13) | Riemann tensor; the invariant (uncancellable) field — "$R$ is what gravity is" |
| $R_{\mu\nu},\ R$ | (3.14) | Ricci tensor and scalar (contractions of Riemann) |
| $G_{\mu\nu}$ | (3.15) | Einstein tensor $R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}$; $\nabla^\mu G_{\mu\nu} = 0$ is the contracted Bianchi identity (3.16) [PROVEN] |
| $T_{\mu\nu}$ | (3.20) | stress–energy tensor; $\nabla^\mu T_{\mu\nu} = 0$ (conservation) is the law the identity enforces |
| $\Lambda$ | (3.20) | cosmological constant; vacuum energy density |
| $\kappa$ | (3.26) | surface gravity of a horizon ($\kappa = a$ Rindler; $c^4/4GM$ Schwarzschild) |
| $r_s$ | (3.28) | Schwarzschild radius $2GM/c^2$ |
| $a(t),\ H$ | (3.31), (3.32) | FRW scale factor; Hubble rate $\dot{a}/a$ |
| $D_H = c/H$ | (3.35) | de Sitter (Hubble) horizon; the cosmological analogue of $r_s$ |
| $\theta,\ \sigma_{\mu\nu},\ \omega_{\mu\nu}$ | (3.39) | expansion, shear, twist of a null congruence; Raychaudhuri focuses light — the bridge to Ch. 10 |
| *Quantum fields (Ch. 4)* | | |
| $a,\ a^\dagger$ | (4.1) | annihilation/creation (ladder) operators, $[a, a^\dagger] = \mathbb{1}$; one pair per field mode |
| $\|0\rangle,\ N = a^\dagger a$ | (4.6) | field vacuum ($a_{\vec{k}}\|0\rangle = 0$); number operator |
| $\|\alpha\rangle,\ D(\alpha)$ | (4.9) | coherent state $D(\alpha)\|0\rangle$; displacement operator; $a\|\alpha\rangle = \alpha\|\alpha\rangle$, overlap $\|\langle\alpha\|\beta\rangle\|^2 = e^{-\|\alpha-\beta\|^2}$ |
| $\alpha,\ \beta$ (Bogoliubov) | (4.18) | mode-mixing coefficients, $\|\alpha\|^2 - \|\beta\|^2 = 1$; count $\langle 0_a\|N_b\|0_a\rangle = \|\beta\|^2$ |
| $T_{\mathrm{U}}$ | (4.27) | Unruh temperature $\hbar a/2\pi c k_{\mathrm{B}}$; the vacuum is warm to an accelerated observer [DERIVED] |
| $h_{\mu\nu}$ | (4.28) | graviton (spin-2) field in $g = \eta + \sqrt{32\pi G}\, h$; the Newtonian field is its coherent state (4.30) |
| *Entropy and horizons (Ch. 5)* | | |
| $A,\ S_{\mathrm{BH}}$ | Ch. 5 | horizon area; black-hole entropy |

*continued on next page*

| Symbol | First defined | Meaning |
|---|---|---|
| $A/4\ell_P^2$ | Ch. 5 | Bekenstein–Hawking entropy (nats); the 1/4 is forced by the first law [DERIVED] |
| $T_H$ | Ch. 5 | Hawking temperature $\hbar\kappa/2\pi c k_B$ |
| $S_{gen}$ | Ch. 5, (9.2) | generalized entropy $A/4\ell_P^2 + S_{ext}$; the two-term split is scheme-dependent [ASSUMED] |
| Bekenstein bd. | Ch. 5 | $S \leq 2\pi k_B RE/\hbar c$; a matter information capacity with no $G$ |
| holographic bd. | Ch. 5 | $S_{max} = k_B A/4\ell_P^2$; a region's information scales with area [MOTIVATED] |
| *Operator algebras (Ch. 6)* | | |
| $\mathcal{M}, \mathcal{M}'$ | Ch. 6 | von Neumann algebra of observables; its commutant |
| $B(\mathcal{H})$ | Ch. 6 | bounded operators on $\mathcal{H}$ |
| $\mathcal{Z}(\mathcal{M})$ | Ch. 6 | centre $\mathcal{M} \cap \mathcal{M}'$; trivial iff $\mathcal{M}$ is a factor |
| type I / II / III | Ch. 6 | Murray–von Neumann factor types; a spacetime region is type $\mathrm{III}_1$ (no trace, no density matrix, infinite entanglement) |
| $\pi, \|\Omega\rangle$ | Ch. 6 | GNS representation and cyclic vector, built from (algebra, state) alone — purification where $\rho$ does not exist |
| *Modular theory (Ch. 7)* | | |
| $S = J\Delta^{1/2}$ | Ch. 7 | Tomita operator; polar split into conjugation and modulus |
| $\Delta, J$ | Ch. 7 | modular operator; modular conjugation |
| $K = -\ln\Delta$ | Ch. 7 | modular Hamiltonian; exists in type III where no $H$ does |
| $\sigma_t$ | Ch. 7 | modular flow (automorphism); the state's own time — the state is the clock |
| $\beta_{mod} = 2\pi$ | Ch. 7 | dimensionless modular period; the Unruh $2\pi$, always this one |
| KMS | Ch. 7 | imaginary-time periodicity = thermality, stated in correlators |
| $f(r)$ | Ch. 7 | Hislop–Longo weight $(R^2 - r^2)/2R$; the diamond's local modular temperature $1/2\pi f(r)$ |
| $\widehat{M} = M \rtimes_\sigma \mathbb{R}$ | Ch. 7 | crossed product; adjoins the observer's clock, type III→II, restores a trace |
| *The correspondence and the three axioms (Ch. 8, 9)* | | |
| the correspondence | Ch. 8 | QM and GR are two restrictions of one state $\omega$ on one algebra; geometry is the commutant of matter [DERIVED] (semiclassical) |
| Axiom I | (9.2) | generalized entropy: $S_{gen}$ is the ledger [ASSUMED] (area law is the geometric input) |

*continued on next page*

| Symbol | First defined | Meaning |
|---|---|---|
| Axiom II | (9.2) | entanglement equilibrium $\delta S_{\rm gen}(\mathcal{D}) = 0$ [DERIVED] (yields Einstein) |
| Axiom III | Ch. 9 | modular time $K = 2\pi H_{\rm phys} + O(G^2)$ [CONJECTURED] (proven in expectation values / solvable models) |
| $H_{\rm phys}$ | Ch. 7, 9 | physical Hamiltonian; the generator the modular clock is claimed to equal |
| *Einstein, constraint, and rate (Ch. 10–12)* | | |
| $\delta Q = T\,\delta S$ | Ch. 10 | Clausius relation; with Unruh $T$, area $S$, and Raychaudhuri, forces $G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G\,T_{\mu\nu}/c^4$ [DERIVED] |
| $\hat{H}_{\rm tot}\,\|\Psi\rangle = 0$ | Ch. 11 | Wheeler–DeWitt constraint; the total Hamiltonian annihilates the physical state |
| $F(t)$ | Ch. 11 | Feynman–Vernon influence functional / coherent-state decoherence functional; its exponent yields $E_{\rm G}$ |
| $\Gamma_{\rm sat}$ | (12.5) | saturation exponent ($\approx$ 0.154 at $1\,\mu$g/1 mm); a static overlap exponent, not a rate |
| $\Gamma_{\rm dec} = C\,E_{\rm G}/\hbar$ | Ch. 12 | decoherence rate [CONJECTURED]; inherits Axiom III's status |
| $C \in [2/\pi, 1]$ | Ch. 12 | rate coefficient, natural $C = 1$ (Diósi–Penrose); floor $2/\pi$ (Margolus–Levitin) [DERIVED] (bounded, not pinned) |
| $G^1$ vs. $G^2$ | Ch. 12 | manifestation (constraint-enforced, physical) vs. generation (perturbative); QGC predicts $G^1$; gap $\approx 3 \times 10^{34}$ |
| *The frontier (Ch. 13)* | | |
| R1 | Ch. 13 | the one open construction: crossed product for a conformal-Killing modular flow with interacting matter (a rigor gap, not a size gap) |
| M7, M8 | Ch. 13 | the two controlled-model values R1's residue reduces to; M8 kernel $h(\omega) = \omega/[2\sinh(\omega/2)]$, $h(0) = 1$ |
| $O(G^2)$ residue | Ch. 13 | the physically negligible ($\sim 10^{-64}$) remainder; a size, not a proof |
| *Cosmology (Ch. 14)* | | |
| $H_0$ | Ch. 14 | present Hubble rate, $2.268 \times 10^{-18}\,{\rm s}^{-1}$ |
| $T_{\rm dS}$ | Ch. 14 | de Sitter horizon temperature $\hbar H/2\pi k_{\rm B}$; the horizon clock |

*continued on next page*

| Symbol | First defined | Meaning |
|---|---|---|
| $a_0 = cH_0/2\pi$ | Ch. 14 | MOND acceleration scale $\approx 1.08 \times 10^{-10}\,\mathrm{m/s^2}$ [DERIVED] (parameter-free, $\sim 10\%$ from data) |
| $\rho_{\mathrm{DE}} = \alpha c^2 H^2/G$ | Ch. 14 | dark-energy density, $w = -1$; form [DERIVED], $\alpha \approx 0.082$ [MOTIVATED] (fitted to $\Omega_\Lambda$) |
| $g(x) = 1 - \frac{2}{\pi}\,\mathrm{Si}(x)$ | Ch. 14 | de Sitter form factor, $g(0) = 1$, $g(1) \approx 0.398$; cosmic rate $\Gamma_{\mathrm{total}} = g(1)H_0/4\pi \approx 7 \times 10^{-20}$ Hz [DERIVED] |
| $2/g(1)$ | Ch. 14 | "the lock": the cancellation $a_0/(c\,\Gamma_{\mathrm{total}}) = 2/g(1)$ in which cosmology's input drops out |
| *Experiment (Ch. 15)* | | |
| $N_{5\sigma}$ | Ch. 15 | runs for a five-sigma visibility deficit, $25/(1-\mathcal{V})^2$; $\approx 37$ at the envelope |
| the envelope | Ch. 15 | the 1.5 fg silica-sphere operating point that makes the test feasible |
| $\theta_m$ (magic angle) | Ch. 15 | the BMV/QGEM geometry ($\cos^2\theta_m = 1/3$) where the entanglement witness has a true zero; QGC's fingerprint against collapse models |
| kill-switches | Ch. 15 | the five results (coherence to a $G^2$ timescale, $G^2$ scaling, exponential onset, collapse-model heating, $a_0(z) \neq cH(z)/2\pi$) that would end the theory |

## Terms of art

**Bekenstein–Hawking entropy** the entropy $A/4\ell_{\mathrm{P}}^2$ (nats) of a horizon of area $A$; the 1/4 is forced by the first law once $T_{\mathrm{H}}$ is fixed (Ch. 5).

**Bisognano–Wichmann** the theorem that the modular Hamiltonian of the Rindler wedge is exactly $2\pi$ times the boost generator; the exact, flat-space case of Axiom III (Ch. 7) [PROVEN].

**Bogoliubov transformation** a linear mixing of a field's creation/annihilation operators; whenever the coefficient $\beta \neq 0$, two observers disagree about the vacuum and the particle count (§4.3).

**Coherent state** the field state that best imitates a classical wave, $|\alpha\rangle = D(\alpha)\,|0\rangle$; minimum-uncertainty, eigenstate of $a$, and the form the Newtonian field of a mass takes (§4.2).

**Correspondence (Quantum-Geometric)** the thesis that QM and GR are two restrictions of one quantum state on one algebra of observables, geometry being the commutant of matter (Ch. 8).

**Crossed product** the construction $M \rtimes_\sigma \mathbb{R}$ that adjoins the observer's clock to a type III region, lowering it to type II and restoring a trace and a finite renormalized $S_{\mathrm{gen}}$ (Ch. 7).

**Decoherence** the decay of off-diagonal density-matrix elements through entanglement with unobserved degrees of freedom; visibility law (2.12). Not a collapse; unitarity holds globally.

**Diósi–Penrose** the heuristic 1980s–90s proposal of $\hbar d/GM^2$ as a superposed geometry's lifetime; QGC recovers the *scale* but not the *collapse*, keeping quantum mechanics exactly linear (Ch. 1, 15).

**Einselection / pointer states** the selection, by the form of the system–environment coupling, of the basis whose superpositions decohere (§2.4.3).

**Entanglement entropy** $S(\rho_A)$ of a reduced state of a pure joint state; $\ln 2$ for a Bell pair.

**Entanglement equilibrium** Axiom II: the physical state extremizes the generalized entropy of every small causal diamond, $\delta S_{\text{gen}} = 0$; jointly with Axiom I it yields Einstein's equations (Ch. 9, 10).

**Gaussian onset** the universal quadratic-in-$t$ start of visibility decay, (2.15); basis of the $\Gamma(2t)/\Gamma(t)$ discriminator (4 for QGC, 2 for collapse models) (Ch. 12, 15).

**Generalized entropy** Axiom I's ledger $S_{\text{gen}} = A/4\ell_{\text{P}}^2 + S_{\text{ext}}$, one monotonic quantity for geometry and matter together; only its renormalized and relative forms are physical (Ch. 5, 9).

**Horizon** an observer-relative causal boundary walling off a hidden region; Rindler (flat space), Schwarzschild ($r_s$), and de Sitter ($c/H$) each carry a surface gravity, a temperature, and an entropy (Ch. 3, 5).

**Influence functional** the Feynman–Vernon object $F(t)$ encoding an environment's effect on a system; for a gravitational bath its exponent is the self-energy $E_{\text{G}}$ (Ch. 11).

**KMS condition** periodicity of correlators in imaginary time; thermal equilibrium stated in the only language that survives when $\rho$ dissolves (Ch. 4, 7).

**The lock** the cancellation $a_0/(c\,\Gamma_{\text{total}}) = 2/g(1)$, in which every cosmological input drops out — a consistency check on the cosmological sector (Ch. 14).

**Mean-field (semiclassical) gravity** sourcing geometry by $\langle \hat{T}_{\mu\nu} \rangle$, (1.3); fails on superpositions (§1.3).

**Modular flow** the automorphism $\sigma_t$ a faithful state induces on its algebra; the state's own thermal time, generated by $K = -\ln\Delta$ (Ch. 7).

**MOND** the modified-dynamics phenomenology of galactic rotation curves, governed by the scale $a_0$; QGC derives $a_0 = cH_0/2\pi$ with no free parameters (Ch. 14).

**Purification** the representation of any mixed state as the reduced state of a pure entangled state (§2.3); formalised as GNS in Ch. 6.

**Raychaudhuri equation** the focusing equation for a null congruence; ordinary matter always focuses light, the engine that turns entanglement into Einstein's equations (§3.9; Ch. 10).

**Type III$_1$** the von Neumann factor of a spacetime region: no trace, no density matrix, infinite entanglement across the boundary at every scale (Ch. 6).

**Unruh effect** the thermality of the Minkowski vacuum to a uniformly accelerated observer, $T_{\text{U}} = \hbar a/2\pi c k_{\text{B}}$; the physically transparent window onto Bisognano–Wichmann (§4.4).

**Wheeler–DeWitt constraint** $\hat{H}_{\text{tot}} |\Psi\rangle = 0$: gravity's total Hamiltonian annihilates the physical state, making every physical quantity relational and forcing the $G^1$ manifestation (Ch. 11).

**Status tags** [PROVEN] / [DERIVED] / [CONJECTURED] / [MOTIVATED] / [ASSUMED] — the honesty system, defined in the Preface; cool = established, warm = open. See Appendix E.

# B
# Constants and Planck Units

*The three input constants, and every derived scale used in the volume, recomputed from them. Nothing here is a fitted or measured "extra" number: given $G$, $\hbar$, $c$ (and, where thermodynamics enters, $k_B$), everything below follows by dimensional combination. These are the computations checked by `TOOLS/verify-paper-numerics.py`.*

## B.1 The input constants

| Symbol | Value (SI) | Name |
|---|---|---|
| $G$ | $6.67430 \times 10^{-11}\ \mathrm{m^3\,kg^{-1}\,s^{-2}}$ | Newton's constant |
| $\hbar$ | $1.054571817 \times 10^{-34}\ \mathrm{J\,s}$ | reduced Planck constant |
| $c$ | $299792458\ \mathrm{m\,s^{-1}}$ (exact) | speed of light |
| $k_B$ | $1.380649 \times 10^{-23}\ \mathrm{J\,K^{-1}}$ (exact) | Boltzmann constant |

Only $G$ carries experimental uncertainty at a level that matters here (about $2 \times 10^{-5}$ relative); $c$, $\hbar$, and $k_B$ are fixed exactly by the 2019 SI redefinition. Every derived quantity is recomputed from these — never pasted as a truncated literal, because a truncated literal breaks the machine-precision identity tests in the verification suite.

## B.2 Planck units

Formed as the unique dimensional combinations (Chapter 1, Worked Example 1.2):

| Quantity | Formula | Value |
|---|---|---|
| Planck mass | $m_P = \sqrt{\hbar c/G}$ | $2.176 \times 10^{-8}$ kg |
| Planck length | $\ell_P = \sqrt{\hbar G/c^3}$ | $1.616 \times 10^{-35}$ m |
| Planck time | $t_P = \ell_P/c$ | $5.391 \times 10^{-44}$ s |
| Planck energy | $E_P = m_P c^2$ | $1.956 \times 10^9$ J $= 1.22 \times 10^{19}$ GeV |
| Planck area | $\ell_P^2 = \hbar G/c^3$ | $2.612 \times 10^{-70}\ \mathrm{m^2}$ |

The Planck area $\ell_P^2$ is the natural unit of horizon entropy (Chapter 5): $S_{\rm gen} = A/4\ell_P^2 + S_{\rm ext}$.

## B.3 Derived physical scales used in the volume

| Quantity | Formula | Benchmark value |
|---|---|---|
| Gravitational self-energy | $E_G = GM^2/d$ | $6.7 \times 10^{-26}$ J |
| Decoherence time | $\tau_{dec} = \hbar d/GM^2 = \hbar/E_G$ | 1.6 ns |
| Unruh coefficient | $\hbar/(2\pi c k_B)$ | $4.05 \times 10^{-21}$ K/(m s$^{-2}$) |
| Hawking temp. (solar) | $\hbar c^3/(8\pi G M_\odot k_B)$ | $6.2 \times 10^{-8}$ K |
| MOND scale | $a_0 = cH_0/2\pi$ | $1.08 \times 10^{-10}$ m s$^{-2}$ |
| Cosmic decoherence rate | $g(1)H_0/4\pi$ | $\sim 7 \times 10^{-20}$ Hz |

The laboratory benchmark is $M = 1$ $\mu$g $= 10^{-9}$ kg, $d = 1$ mm $= 10^{-3}$ m. The cosmological benchmark uses $H_0 \approx 2.2 \times 10^{-18}$ s$^{-1}$ ($\approx 70$ km s$^{-1}$ Mpc$^{-1}$); $H_0$ is the one astronomical input, not derivable from $G, \hbar, c$.

## B.4 The rate coefficient window

The gravitational decoherence rate is $\Gamma_{dec} = C\, E_G/\hbar$ with the coefficient bounded (Chapter 12) to

$$C \in \left[\tfrac{2}{\pi},\, 1\right] = [0.637,\, 1.000], \qquad \text{natural value } C = 1 \text{ (Markovian dephasing / Diósi–Penrose).}$$

The floor $2/\pi$ is the Margolus–Levitin quantum-speed-limit bound. The window is a rigorous consequence of the framework, but $C$ is *bounded*, not uniquely pinned; and the whole rate carries the status of the modular identity $K = 2\pi H_{phys}$: [CONJECTURED].

## B.5 Useful astronomical and atomic values

| Quantity | Value |
|---|---|
| Solar mass $M_\odot$ | $1.989 \times 10^{30}$ kg |
| Earth mass | $5.972 \times 10^{24}$ kg |
| Electron mass | $9.109 \times 10^{-31}$ kg |
| Proton mass | $1.673 \times 10^{-27}$ kg |
| Age of the universe | $4.35 \times 10^{17}$ s (13.8 Gyr) |
| $\ln 2$ (nats per bit) | 0.693 |
| $2/\pi$ | 0.6366 |
| Si(1) (sine integral) | 0.9461 |

# C
# Running the Verification Harness

*Every displayed number in this volume is recomputed from $G, \hbar, c$ (and $k_B$) by a script in this repository and compared to the value printed on the page. This appendix shows how to run those scripts. The rule of the book: if a number here ever disagrees with the harness, the harness wins.*

## C.1 What the harness is

The file `TOOLS/verify-paper-numerics.py` recomputes every headline number in the QGC papers *and* in these lecture notes from the physical constants and the stated formula, then checks it against the value written down (PASS/FAIL, non-zero exit on any failure). The lecture-notes checks live in the block labelled *section N* ("notes"); each RUN IT footer in the text names the chapter tag you will see in that block, e.g. `ch5 Hawking temperature`.

The design principle is the discipline the whole research programme runs on: no number enters the text unless code can reproduce it from first principles. A derived constant is *recomputed* from $G, \hbar, c$ inside the script — never hard-coded as a truncated literal, because a truncated $\ell_P$ would fail the machine-precision identity tests (for example $S_{BH} = A/4\ell_P^2 = 4\pi GM^2/\hbar c$, checked as an exact identity, not a coincidence of rounded digits).

## C.2 Running it

From the repository root, with a Python that has `numpy`/`scipy` available (the `conda base` environment):

```
python3 TOOLS/verify-paper-numerics.py
```

A green run ends with a line of the form "`N checks, 0 FAIL (ALL GREEN)`" and exits zero. Each row prints the claim, the value written in the text, the value recomputed from constants, their ratio, and the verdict. To see only the lecture-notes rows:

```
python3 TOOLS/verify-paper-numerics.py | grep notes
```

## C.3 The wider suite

The numerics harness is one member of the internal-consistency suite:

```
./TOOLS/run-tests.sh
```

runs the axiom tests, the dedicated derivation checks (modular-operator identification, rate-coefficient, cosmic-decoherence, graviton modular-reduction, the falsification envelope, the MOND scale, ...), and the paper-numerics verifier together. A zero exit certifies that both the internal-consistency tests and every headline number are green. The individual scripts named in the Run it footers — for instance `TOOLS/verify_falsification_5sigma.py` (Chapter 15), `VERIFIED/tests/mond_scale_verification.py` and `test_cosmic_decoherence.py` (Chapter 14), `test_rate_coefficient.py` (Chapter 12) — can be run on their own the same way.

## C.4 Reproducing a single number by hand

Nothing in the harness is a black box. Every Run it footer points at a formula stated in the text; to reproduce, say, the benchmark decoherence time, open a Python prompt and type the formula with the constants of §B:

```
G, hbar = 6.67430e-11, 1.054571817e-34
M, d = 1e-9, 1e-3
tau = hbar*d/(G*M**2)   # 1.58e-9 s
```

That three-line computation is exactly what the harness automates, once per number, across the whole volume.

# D
# Master FAQ

*The reader who works straight through this book accumulates the same few objections again and again — about honesty, about the measurement problem, about $G^1$, about the rate, about rivals, about how it could be wrong. Each is answered where it arises; this appendix gathers the recurring ones in one place, gives the short answer in the book's own honest voice, and points to the full treatment. Nothing here is a new claim. The status tags are exactly those of the ledger (Appendix E): no silent upgrades.*

## Is this real physics or numerology?

**Q: What is actually *new* here, if the theory changes neither quantum mechanics nor general relativity and adds no new field?** A single quantitative, falsifiable prediction, and a unification. Semiclassical gravity predicts *no* gravitational decoherence; perturbative quantized gravity predicts an $O(G^2)$ effect suppressed by $\sim 10^{34}$; QGC predicts an $O(G^1)$ effect at the 1.6 ns benchmark. Three mutually exclusive numbers for one laboratory observable is a fork, not a relabelling (Chapter 8, "if QGC changes neither theory…"). The unification is that one axiom set yields both the laboratory decoherence scale (Diósi–Penrose, 1980s) and the cosmic acceleration scale (Milgrom, 1983), with quantum mechanics kept exactly linear (Chapter 15, "if QGC predicts nothing new…").

**Q: Isn't "QGC predicts nothing new" a fatal admission?** It is the book's most load-bearing limitation, and it is stated as such. QGC contributes *unification and a coefficient*, not a new force or particle. Unification is epistemically weaker than a novel prediction, and Chapter 15 presents it that way rather than dressing it up. What makes it more than numerology is that the two scales fall out of the *same* three axioms, and that the laboratory number comes with sharp experimental discriminators against the collapse models that share its timescale.

**Q: How honest are the status labels — couldn't "derived" be hiding a fit?** The labels are the discipline of the whole project. The energy scale $E_{\mathrm{G}} = GM^2/d$ is [PROVEN]; the operator equation $K = 2\pi H_{\mathrm{phys}}$ and hence the *rate* are [CONJECTURED]; the dark-energy coefficient $\alpha \approx 0.082$ is [MOTIVATED] (openly a fit to $\Omega_\Lambda$); the Born rule is [ASSUMED]. The internal grade is B+. Appendix E is one page precisely so that no claim can be quietly promoted; every chapter's reader-questions return to these tags rather than around them.

## The measurement problem and the Born rule

**Q: Is the conflict this book starts from just the measurement problem in disguise?** No — adjacent but distinct. The measurement problem asks why experiments have single outcomes at all, and exists with or without gravity. The conflict here is sharper: two spectacularly confirmed dynamical equations make contradictory demands on one observable, the gravitational field of a superposition. You could solve the measurement problem tomorrow and this conflict would remain (Chapter 1, "is the conflict…not just the measurement problem"). Conversely, QGC takes no interpretational position.

**Q: Does QGC solve the measurement problem or derive the Born rule?** It does neither, and says so. The Born rule enters as [ASSUMED], exactly as in ordinary quantum mechanics (Chapter 2). QGC explains why superpositions of *distinct geometries* decohere — coherence relocates into matter–geometry correlations, globally unitary, locally mixed (Chapter 8, "if the global state stays pure…") — but it does not tell you why one branch is experienced. Decoherence selects a basis; it does not select an outcome.

## Why $G^1$ and not $G^2$?

**Q: Real graviton emission is a textbook $O(G^2)$ effect. Isn't the $G^1$ claim simply wrong?** Both scalings are correct; they compute different things. $O(G^2)$ is the rate of entanglement *generation* from a bare, unconstrained initial state — the perturbative-QFT answer. $O(G^1)$ is the rate of entanglement *manifestation* from the physical state that already satisfies the constraint. Nature sits in the dressed, constraint-satisfying state, so $G^1$ is the physical rate (Chapter 12, Q 12.4 and the Chapter FAQ). The two differ by $\sim 3 \times 10^{34}$ at the benchmark.

**Q: Why doesn't the same argument decohere every charged particle through its Coulomb field?** Because electromagnetism has only a *spatial* constraint (Gauss's law), which generates internal $U(1)$ gauge rotations and leaves time external — modular flow $\neq$ physical flow. Gravity has a *Hamiltonian* constraint (Wheeler–DeWitt, $\hat{H}_{\text{tot}} |\Psi\rangle = 0$), which makes time emergent and relational and converts the constraint-entanglement into decay (Chapter 12, Chapter FAQ; Chapter 11 builds the Gauss-law→Wheeler–DeWitt ladder). The distinction is the whole reason the mechanism is gravitational.

**Q: Is the Wheeler–DeWitt constraint the $G^1$ argument leans on even established?** The constraint *structure* is textbook Dirac since the 1960s; what is open is the full non-perturbative quantization. The book uses only the linearized, weak-field face — the quantized Poisson equation $\nabla^2\hat{\Phi} = 4\pi G\hat{\rho}$ — which is under complete control (Chapter 11, Chapter FAQ, "is the Wheeler–DeWitt equation…established"; and Chapter 9, "the QED contrast turns on…").

## Is the rate proven?

**Q: In one sentence — is the theory proven?** No: the central operator equation $K = 2\pi H_{\text{phys}}$, and therefore the decoherence *rate*, are [CONJECTURED] — [PROVEN] in every expectation value and solvable model, open as a full operator equation for the interacting gravitational diamond, and reduced to one construction (R1) whose remaining content is two controlled-model values (M7, M8) and a physically negligible $O(G^2)$ residue. The energy scale is [PROVEN]; the prediction is falsifiable; the grade is B+ (Chapter 13, Q 13.1 and Chapter FAQ).

**Q: If $K = 2\pi H_{\text{phys}}$ holds in every expectation value, in what sense is it not just true?** Because decoherence lives in *off-diagonal* matrix elements, not first moments. An expectation-value identity is far weaker than an operator identity; two operators can agree on every $\langle\psi|\cdot|\psi\rangle$ and differ on $\langle\psi|\cdot|\phi\rangle$, which is exactly where coherence is lost (Chapter 9, "you say $K = 2\pi H_{\text{phys}}$ is proven in expectation values"). The open question is whether the full operators coincide in the interacting theory.

**Q: The $O(G^2)$ residue is $10^{-64}$. Why not just declare victory?** Because $10^{-64}$ is a *size*, and what is missing is a *proof*. The number is computed in a controlled Gaussian model where everything diagonalises; the interacting field theory is where the closed-form diagonalisation fails, and the controlled value, however tiny and however robust, is not a theorem for the interacting case. R1 is a rigor gap, not a size gap (Chapter 13, Chapter FAQ; and Q 13.3 on why the IR cancellation is content, not a rigged tautology).

**Q: Is $C = 1$ the prediction, or is the prediction a range?** Rigorously, a range: $C \in [2/\pi, 1]$, floored by Margolus–Levitin and Mandelstam–Tamm. For the visibility decay an interferometer actually reports, the natural coefficient is the ceiling $C = 1$ (Markovian dephasing = Diósi–Penrose). It is bounded, not uniquely pinned; the residual freedom is the clock-model normalization (Chapter 12, Q 12.3 and Chapter FAQ).

## How is this different from Diósi–Penrose, collapse models, string theory, LQG?

**Q: Isn't the timescale $\hbar d/GM^2$ just Diósi–Penrose with extra steps?** The *number* is the Diósi–Penrose scale, and the book credits it. What QGC claims as its own is the derivation and its consequences: the scale emerges from an axiomatized correspondence rather than being postulated; quantum mechanics stays exactly linear (no stochastic collapse terms); and the same axioms yield the cosmological results with no further input (Chapter 1, "which existing approach…"). Crucially, Diósi–Penrose is a *collapse* model — it modifies quantum mechanics and injects energy — whereas QGC is unitary (Chapter 15, "isn't this just Diósi–Penrose…").

**Q: If an experiment measures $\hbar d/GM^2$, haven't QGC and Diósi–Penrose *both* been confirmed?** Only the shared scale. Three discriminators separate them: (i) decay shape — Gaussian onset, $\Gamma(2t)/\Gamma(t) = 4$ for QGC, versus exponential, ratio 2, for collapse; (ii) the coefficient window $C \in [2/\pi, 1]$ versus a single collapse value; (iii) energy conservation — QGC is unitary, collapse models heat spontaneously (Chapter 15, "you keep contrasting QGC with Diósi–Penrose"). These are the kill-switches of §15.6.

**Q: How does this relate to string theory and loop quantum gravity?** QGC is not a theory of the Planck scale and makes no claim there. It takes no position on the microscopic origin of the axioms; it addresses only the *semiclassical seam* where a superposition sources distinct geometries (Chapter 9, "three axioms…isn't that suspiciously few"). It uses established operator-algebra results (Bisognano–Wichmann, and the crossed product of Chandrasekaran–Longo–Penington–Witten) as a base (Chapter 13, "if CLPW already proved the crossed product…"), and inherits the holographic principle only as far as it is established (Chapter 5, "is the holographic principle really established").

## How could it be wrong?

**Q: Five kill-switches sounds like a lot of ways to be wrong. Isn't that a weakness?** It is the opposite — it is what makes the theory science. Each kill-switch is a concrete, near-term measurement whose result would end the theory: coherence surviving to a $G^2$ timescale; a $G^2$ rate scaling; exponential rather than Gaussian onset; collapse-model heating; or $a_0(z) \neq cH(z)/2\pi$ (Chapter 15, §15.6). A theory that could fail in five independent ways and has not yet is more constrained, not less.

**Q: How can you separate a gravitational effect from ordinary environmental decoherence?** By shape, not size. Environmental and gravitational decoherence obey different laws in mass, separation, and time profile; the strategy is to push environmental sources (residual gas below $\sim 7 \times 10^{-15}$ mbar, cryogenic walls) until $t_{\text{env}} \gg t_{\text{grav}}$, so the first coherence loss the fringes feel is the uncoolable floor (Chapter 15, "how can you possibly separate. . . "). That floor depends only on $M$, $d$, and constants of nature (Chapter 1, "doesn't heavier $\Rightarrow$ faster. . . ").

**Q: What if the experiment sees nothing — the fringes survive?** Then a kill-switch has fired. "Nothing" at the envelope — coherence maintained where QGC predicts a deficit — is itself a falsification of the $O(G^1)$ mechanism, pushing the effect to the $G^2$ timescale and refuting the central operator equation (Chapter 15, "what if the experiment sees nothing"). The rate being [CONJECTURED] is not a hedge: the conjecture is exactly what the experiment tests (same chapter, "the ledger lists the rate as CONJECTURED").

## Cosmology: is MOND just a fit? Is dark energy predicted?

**Q: Isn't MOND just a fit — you tuned one number $a_0$ to rotation curves?** MOND *was* a fit. QGC *computes* $a_0 = cH_0/2\pi$ from $c$ and the Hubble rate — quantities measured with no reference to galaxies — and lands within $\sim 10\%$ of the value rotation curves require. A number you can derive from unrelated inputs has become a prediction, and the $\sim 10\%$ miss is reported honestly, not buried (Chapter 14, "isn't MOND just a fit"). The chapter takes no position on clusters or the CMB, where MOND struggles (Chapter FAQ, "MOND without dark matter").

**Q: Why is $\alpha$ (dark energy) not predicted, when $a_0$ was?** Different kinds of quantity. $a_0$ is set by the horizon temperature alone. The dark-energy coefficient $\alpha$ in $\rho_{\text{DE}} = \alpha c^2 H^2/G$ is a density *normalization* that depends on apportioning horizon entropy against matter; the equilibrium condition constrains it but does not uniquely pin it at the present epoch. So the functional form $\rho_{\text{DE}} \propto c^2 H^2/G$ ($w = -1$) is [DERIVED] and $\alpha \approx 0.082$ is [MOTIVATED] — fitted to $\Omega_\Lambda$ (Chapter 14, "why is $\alpha$ not predicted").

**Q: A naive estimate gave cosmic decoherence at $10^{103}$ Hz. Doesn't that destroy the universe?** That number was a Minkowski-formula artifact — the wrong propagator. The de Sitter recomputation gives the form factor $g(x) = 1 - \frac{2}{\pi}\,\text{Si}(x)$ and a cosmic rate $\Gamma_{\text{total}} = g(1)H_0/4\pi \approx 7 \times 10^{-20}$ Hz: finite, sub-causal, utterly unobservable (Chapter 14, "was the naive $10^{103}$ Hz just the wrong propagator"). Its importance is conceptual: it shows the framework predicts no cosmic catastrophe, and the "lock" $a_0/(c\,\Gamma_{\text{total}}) = 2/g(1)$ checks the sector's internal consistency.

# E The Status Ledger

*One table for the whole theory: every major claim, the chapter that establishes it, and its honest epistemic status. Cool tags ([PROVEN], [DERIVED]) mark established ground; warm tags ([CONJECTURED], [MOTIVATED], [ASSUMED]) mark open ground. This ledger is the single most scientifically load-bearing page of the book: it is what keeps the theory honest. It matches the canonical ledger in `THEORY/axioms/CANONICAL-MINIMAL-AXIOMS.md` and the audited statuses of the June-2026 full-theory review.*

## E.1 The claims and their status

| Claim | Status | Where / caveat |
|---|---|---|
| *Foundations* | | |
| The correspondence: QM and GR are two restrictions of one state on one algebra | [DERIVED] | Ch. 8; semiclassical; strict duality aspirational |
| Axiom I — generalized entropy $S_{\mathrm{gen}} = A/4\ell_{\mathrm{P}}^2 + S_{\mathrm{ext}}$ | [ASSUMED] | Ch. 5,9; area law is the geometric input; split is scheme-dependent |
| Axiom II — entanglement equilibrium $\delta S_{\mathrm{gen}} = 0$ | [DERIVED] | Ch. 9; from the modular structure |
| Axiom III — modular time $K = 2\pi H_{\mathrm{phys}} + O(G^2)$ | [CONJECTURED] | Ch. 9,13; proven in expectation values / solvable models |
| Born rule $p_n = \|c_n\|^2$ | [ASSUMED] | Ch. 2; as in all of quantum mechanics |
| *What the framework reproduces* | | |
| Einstein's field equations from the axioms | [DERIVED] | Ch. 10; given the area law as input |
| Bekenstein–Hawking entropy, Unruh, Hawking temperature | [DERIVED] | Ch. 5; standard results recovered |
| Bisognano–Wichmann $K = 2\pi K_{\mathrm{boost}}$ (flat space) | [PROVEN] | Ch. 7; the exact case of Axiom III |
| *The laboratory prediction* | | |

*continued on next page*

| Claim | Status | Where / caveat |
|---|---|---|
| Energy scale $E_G = GM^2/d$ | [PROVEN] | Ch. 11; classical gravity, rigorous |
| $G^1$ scaling (manifestation, not generation) | [DERIVED] | Ch. 12; linearized, controlled approximation |
| Decoherence rate $\Gamma_{dec} = C\, E_G/\hbar$ | [CONJECTURED] | Ch. 12; inherits Axiom III's status |
| Coefficient $C \in [2/\pi, 1]$, natural $C = 1$ | [DERIVED] | Ch. 12; bounded, not uniquely pinned |
| R1 closure (operator equation for the dynamical diamond) | [CONJECTURED] | Ch. 13; residue = two controlled-model values (M7, M8) |
| *The cosmological sector* | | |
| MOND scale $a_0 = cH_0/2\pi$ | [DERIVED] | Ch. 14; parameter-free, $\sim 10\%$ from data; $z \sim 1$ point 24% high |
| Dark-energy structure $\rho_{DE} \propto c^2H^2/G$, $w = -1$ | [DERIVED] | Ch. 14; functional form only |
| Dark-energy coefficient $\alpha \approx 0.082$ | [MOTIVATED] | Ch. 14; fitted to $\Omega_\Lambda$, not predicted |
| Cosmic decoherence $\Gamma = g(1)H_0/4\pi$ | [DERIVED] | Ch. 14; controlled approximation |
| GUP / minimum length / modified dispersion | [MOTIVATED] | phenomenological; unobservably small |

## E.2 The one-line summary

The energy scale $E_G = GM^2/d$ is proven; the identification of the modular clock with physical time, $K = 2\pi H_{phys}$, and hence the absolute decoherence *rate*, is conjectured — proven in expectation values and every solvable model, open as a full operator equation, with the residue sharply isolated in one construction (R1, Chapter 13). The programme's internal grade is **B+**. What separates B+ from a theorem is mathematics (R1); what separates it from a fact is experiment (Chapter 15). Both are identified, both are in progress, and the theory names the results — coherence surviving a $G^2$ timescale, a $G^2$ rate scaling, exponential rather than Gaussian-onset decay, collapse-model heating — that would end it.