# The Real Observer: Constraints on Physical Law Necessitated by the Existence of Embedded Observers

J. B. Thompson

*The first fact is not outside physics: a finite piece of the world can hold a record of the world.*

**Abstract**

We identify the mathematical structure selected by a single invariance principle (Axiom A0) together with four constitutive constraints on any embedded physical observer: a bounded, finite-energy record-keeping system that updates a causally ordered internal memory solely through local interactions at finite speed along a continuous worldline. A0 asserts that physical reality consists solely of what is invariant under physically inert transformations. These assumptions first define an operational quotient: states are equivalence classes of histories under all admissible observer protocols. Part I then reconstructs the nonclassical irreducible sector under explicitly named regularity conditions—finite randomized preparation with operational completion, a countable protocol base, sectorwise continuous accessibility, operational isotropy of independent distinguishability axes, and operational local completeness for independent composites. Under those conditions the surviving structure is complex Hilbert space, the Born rule, wave–particle duality recast as record–coherence complementarity, tensor-product composition, unitary dynamics, and a local U(1) gauge field with massless mediator. Part II gives the laboratory-scale operational consequence of finite control energy: a finite observer/controller cannot refine a protocol without bound. Applying the Margolus–Levitin quantum speed limit to a $\pi$-pulse gives the pulse-count ceiling $n_{\text{max}}(T, E_{\text{op}}) = 2E_{\text{op}}T/(\pi\hbar)$. Coherence-time scalings follow only after adding ideal-pulse filter theory and are treated as model-dependent; thermal timing floors are retained only as conditional clock models. Part III reaches Lorentzian geometry, the local-flatness clause of the equivalence principle, and, with named imports, Einstein's equations. Part IV reaches, using the cyclic separating vacuum as an explicit AQFT structural input, the type II crossed-product algebra on which the Bekenstein bound and generalized entropy become theorems for Rindler and de Sitter horizons; the same structure was reached independently from large-$N$ gravity. A closing extension identifies the Dirac spinor as the form of record-faithful local first-order matter, with the observer's own existence supplying minimal occupation through the fundamental spin-$\frac{1}{2}$ constituents of ordinary matter while leaving the particle census open. The dependency graph records the epistemic status of the major steps, and every import is named at its point of use. This is not a theory of everything or a derivation from nothing: it constrains the form of physical law, not the world's particle multiplicities, couplings, constants, or initial conditions.

## Contents

## 1 Declaration and epistemic gradient

The first fact is not outside physics: a finite piece of the world can hold a record of the world. Every reader of this is such a system, inside the world it describes. This is the operative fact, not a philosophical preliminary: you are made of matter, you arrived here through physical processes, and every piece of information you hold reached you through physical interaction, so you cannot step outside the world to examine it from a neutral vantage, and neither can we. The question this raises is the whole of the work—given that every observer is constitutively *inside* the world, what must the structure of physics look like?—and the answer offered is that quantum mechanics, including the ordinary wave–particle puzzle when read through record formation, Lorentzian geometry, and algebraic quantum field theory are not three unrelated structures but layers of a single structure compatible with observation being finite, local, interaction-based, and situated. The argument is constraint-based, not a derivation from nothing: A0 and O1–O4 first define the operational quotient appropriate to embedded observers; the nonclassical quantum sector is then selected under explicitly named regularity conditions rather than smuggled out of A0 alone. The deepest claim is conditional—a theory of real finite embedded observers must either pass through this sieve of invariant structure or state exactly which observer regularity it rejects. It is not a theory of everything: it does not fix the Standard Model gauge group, particle masses, the Higgs sector, or the cosmological constant, and it derives matter's *form* while leaving multiplicities, masses, and couplings open. Most of its content recovers known physics by a new route; the most concrete near-term laboratory contact is the finite-control-energy ceiling on dynamical-decoupling refinement depth developed in Part II.

The vacuum is therefore acknowledged at the outset as a structural input, not postponed as a technical afterthought. In Part IV it enters as the cyclic separating state that makes the local algebra faithful enough for Tomita–Takesaki modular theory to act. It is not empty nothing; it is the operationally indispensable state whose entanglement structure allows finite observers, horizons, modular flow, and generalized entropy to be discussed in one algebraic language.

## 2 The dependency graph

Before the symbols, a map. The graph below is the argument in one image: the arrows mark dependence, and the labels mark the kind of mathematical or physical step. It is included so the body can focus on the reasoning rather than repeatedly announcing which claims are derived, reconstructed, imported, or conjectural.

**How alternatives exit the sieve**

This is not a claim that alternatives are logically impossible from A0 alone. It is a map of exits. Classical probability satisfies finite observerhood but fails operational isotropy-/strong symmetry because the vertices of the simplex define a preferred operational basis. Superselection theories are handled as disconnected sectors and reconstructed component by component. Real and quaternionic quantum theories preserve much of the single-system structure but fail the composite local-completeness condition used for the tensor product. Non-Euclidean norms fail the redecomposition-isotropy test. Thus the reconstruction is a sieve: a rival theory can escape, but it must identify which observer regularity it rejects.

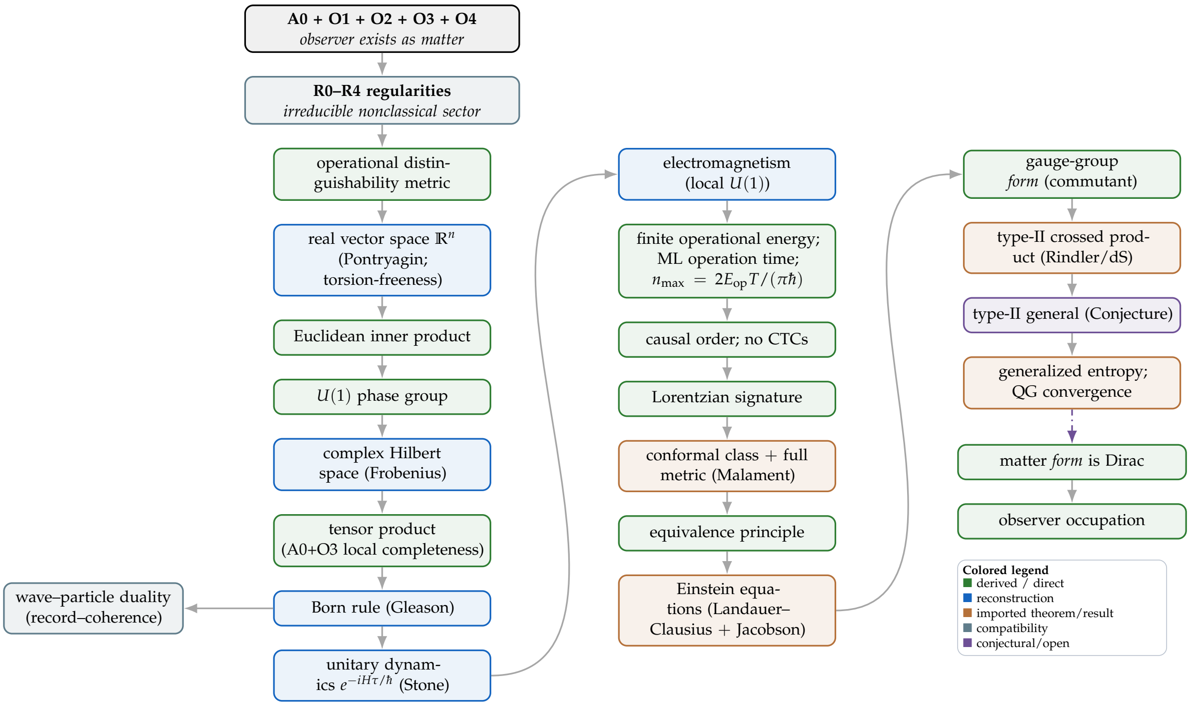


Figure 1: The dependency graph. The QM spine runs down the left; electromagnetism, the finite operational-energy bound, and geometry occupy the centre; the algebraic-QFT results and the matter extension are on the right. A0 and O1–O4 define the observer and the operational quotient; the Part I nonclassical quantum reconstruction additionally uses the explicitly named regularities R0–R4. In Part IV the cyclic separating vacuum is an independent AQFT structural input. The recurring signature of the strongest results is *form, not specifics*: the gauge node fixes the group's form but not its integers; the matter node fixes the field's form while leaving multiplicities, masses, and couplings open.

| **Possible escape** | **Where the sieve catches it** | **Regularity or load-bearing condition** |
|---|---|---|
| Non-Euclidean distinguishability | Redecomposition would become detectable | A0 plus operational isotropy of independent sectors |
| Torsion or cyclic state space | Independent copies cannot return to zero | Pythagorean distinguishability |
| Real-only amplitudes | No continuous physical phase scalar | U(1) phase reconstruction |
| Quaternionic alternatives | Order-sensitive scalars and anomalous composites | Frobenius plus operational local completeness |
| Exceptional Jordan block | Fails independent-composite local completeness | A0+O3 composite closure |
| Super-quantum no-signalling boxes | Exceed Hilbert/Born correlation geometry | inner-product plus Born reconstruction |
| Nonlocal square-root dynamics | Requires unmediated spatial access | pre-metric O3 locality |
| Specific Standard Model census | Not fixed by the form argument | empirical multiplicities/couplings |

Table 1: The sieve at a glance. The graph gives the positive chain; this table records where common alternative structures leave it.

## 3 The one principle and the four constraints

Everything rests on one principle about *what physics is*, plus four facts about *what an observer is*. These are not derived. They are the starting point. The whole framework is the claim that nothing else needs to be assumed.

### 3.1 Axiom A0 — physics is invariant structure

**Definition 3.1** (A0)**.** *Physical reality consists solely of what is invariant under all* physically inert *transformations. A transformation $T$ on histories is physically inert iff for every admissible protocol $\pi$ and history $H$,*

$$P_\pi(o \mid H) = P_\pi(o \mid T(H)) \quad \textit{for every outcome } o.$$

**Symbols and operations**

$P_\pi(o \mid H)$ is the probability of outcome $o$ when you run protocol $\pi$ on history $H$. $T(H)$ is the history after applying the transformation $T$. The equation says: $T$ changes *no* probability of *any* outcome of *any* experiment. “Iff” is “if and only if.”

**In plain English**

If a change to the world makes no possible difference to any measurement anyone could ever do, then it is not a change to the world — it is just a different way of writing the same thing down. Physics is what is left after you throw away all such re-writings.

**Physical justification**

This is the move Leibniz, Mach, and Einstein each made: absolute position, absolute rotation, and absolute simultaneity were all discarded *because* they made no operational difference, and in each case the invariant remainder *was* the physics. A0 generalizes the move and makes it the definition of physical content. Crucially A0 is *ontological, not epistemic*: a non-invariant quantity is not “real but undetectable,” it has no physical content at all. Rejecting A0 means asserting that there exist facts that make no difference to anything, ever, for anyone — and the burden of proof is on whoever asserts that.

*Remark.* The one honest pressure point on A0 is the “future-protocol” objection: absolute simultaneity *looked* inert until specific experiments made it matter. A0 quantifies over all admissible processes, not known ones, and O1–O4 fix what “admissible” means. A0 partitions every claimed distinction into exactly one of three cases: (a) detectable by some O1–O4-admissible process, hence real; (b) detectable only by a process violating O1–O4, hence evidence *against* O1–O4 if such a process is exhibited; (c) detectable by no admissible process, hence inert. There is no fourth case. That trichotomy is what makes A0 falsifiable rather than merely stipulative.

### 3.2 The four constitutive constraints O1–O4

These are not assumptions about the laws of physics. They are what it *means* to be a finite physical system inside the world.

**O1 — Finite energy (two roles).**
$E_{\mathcal{O}} < \infty$, with $E_{\text{op}} \le E_{\mathcal{O}}$ usable for control. One axiom doing two distinct jobs: *(1a)* the recording/control degree of freedom has finite operational energy, which fixes

a minimum duration for committed record-changing operations through Margolus–Levitin (Parts I–II); *(1b)* the observer's total energy is finite, which makes it a bounded stress-energy source contributing to $T_{\mu\nu}$ (Part III). The two roles are flagged at the point of use.

**O2 — Finite, strictly totally-ordered records.**
Finitely many reliably distinguishable records — a *count*, not a Hilbert-space dimension — with at least two records (so recording is nontrivial) and a strict total order on them. The order is a property of the physical *write*, not of recall or later processing, which is why it holds exactly even for biological or distributed systems.

**O3 — Interactive locality.**
All knowledge is acquired only through finite-speed local physical interaction. Locality and finite signal speed $c < \infty$ are not two facts but one: an interaction *is* the finite-speed propagation of influence, and this is constitutive of what observation of invariant structure is, not an imported substrate fact.

**O4 — Proper time, continuous in the limit.**
The observer carries an internal monotone, dynamics-generating proper-time parameter $\tau$. At finite resources, distinguishable record-changing operations have a minimum duration; the continuous, Stone-enabling form is the unbounded-resource refinement limit. Part IV depends most on this constraint.

**O3 is pre-metric**

O3 asserts only a primitive accessibility relation between record-changing events: influence is mediated, finite, and composable. It does not assume a Lorentzian manifold, a metric, null cones, or differentiable spacetime. The causal order and metric geometry reconstructed in Part III are representations of this accessibility relation in the continuum limit, not premises smuggled into the starting point.

**In plain English**

You are made of a finite amount of stuff (O1), you can hold only finitely many ordered records (O2), you learn only by bumping into things at finite speed (O3), and your own clock is implemented by finite physical operations, with continuous time recovered only in the limit of unbounded refinement (O4). That is the entire list. Everything downstream is what the world must look like to *any* system of which all four are true.

*Remark.* The four constraints can be stated at finer grain, but nothing in the mathematics turns on the subdivision. A separate "finite resolution" condition is unnecessary: its upper bound is redundant with O2's finite record count and its lower bound ("at least two states") is part of O2. Local interaction and finite signal speed are a single fact—an interaction *is* finite-speed propagation of influence—carried by O3. O1 plays two roles, operational energy for committed record-changing operations and total energy for the stress-energy source, flagged where each is used. O4's proper time is implemented through finite physical operations and becomes a continuous parameter only in the refinement limit, which makes the unitarity result of Section 10 a limit statement.

**Physical justification**

The decisive feature is that the *same* four constraints do real work in all four domains. O2 alone shows up in the torsion argument, finite protocol refinement, the Second Law, the no-closed-timelike-curve result, the single-timelike-direction condition, and the need for the clock to sit inside the observable algebra. O3 (interactive locality) shows up in cancellativity, operational local completeness, the causal order, Lorentzian signature, and microcausality. This is why quantum mechanics and general relativity turn out compatible: not by coincidence, but because both are answers to the same observer facing the world from the inside.

*Remark.* A0 and O1–O4 are *mutually constitutive*, not circular: A0 says what "same physical content" means; O1–O4 say what counts as an admissible observer. The framework is falsifiable in one clean way — exhibit a system satisfying O1–O4 whose physics does *not* reduce to the structure identified here.

### 3.3 Regularity conditions for the nonclassical sector

A0 and O1–O4 do not by themselves assert that every operational state space is quantum. They define the admissible observer, the admissible protocols, and the quotient by operational indistinguishability. The finite-dimensional quantum reconstruction in Part I applies to an irreducible nonclassical observer sector satisfying the following regularity conditions. They are listed here so that the proof does not hide them in later steps.

**R0 — finite randomized closure and completion.**
If two preparations are admissible, a finite observer can condition on a finite record and choose between them with rational weights. Full convex closure is the operational completion of this finite randomized structure under the distinguishability metric, not an extra claim that A0 alone manufactures arbitrary mixtures.

**R1 — countable operational protocol base.**
The operational topology is generated by finite, countably specifiable protocols and rational tolerances. This is the observer-side content of second-countability: finite records do not make the state space finite, since repeated trials can estimate continuous parameters, but admissible protocols must be finitely describable by the embedded observer.

**R2 — sectorwise continuous accessibility.**
The reconstruction is performed inside one irreducible dynamically accessible component: states in that component are connected by admissible continuous transformations generated along the observer's proper time. Disconnected components are superselection sectors and are reconstructed separately.

**R3 — operational isotropy / strong symmetry.**
In an irreducible nonclassical sector, passive redecompositions of independent distinguishability axes that preserve independent composition and the admissible protocol family are A0-inert, and the resulting inert action is sufficiently transitive on unit distinguishability directions to select the Euclidean norm. Classical simplices fail this condition because their vertices define a preferred operational basis.

**R4 — operational local completeness.**
For independent composites, the physically relevant joint state is exhausted by lo-

cal record-forming protocols on each side together with later comparison of records. This is motivated by A0+O3, but it is stated as a closure condition: no globally hidden composite degree of freedom is admitted unless it is accessible to an admissible protocol.

**R5 — AQFT vacuum input.**
Part IV imports the cyclic separating vacuum/local-algebra structure standard in AQFT. For Rindler, de Sitter, and BTZ settings the crossed-product/type II claims are theorem-level conditional on this input; general-background extensions remain conjectural.

**In plain English**

The paper's claim is therefore not that A0 alone secretly contains quantum mechanics. The claim is that finite embedded observers force an operational quotient, and that once the minimal regularities needed for finite randomized control, countable protocol description, sectorwise accessibility, isotropic redecomposition, and local composite completeness are imposed, the surviving irreducible nonclassical sector is the usual complex Hilbert-space sector.

*Remark* (Operational locality vs geometric spacetime). The first objection a careful reader raises is that the framework appears to *assume* spacetime in O3 (locality, finite speed) and then *derive* spacetime in Part III. The reply should be in the body, not left implicit, because it is load-bearing. O3 asserts only a *primitive accessibility relation*: information transfer between interactions is mediated and finite. It does *not* assume a Lorentzian manifold, light-cones, or a metric. From O3 (mediated finite accessibility) plus O2 (ordered records) the framework reconstructs, in order, the causal partial order, local finiteness, and — given a smooth manifold — the Lorentzian metric. Operational causality is *assumed*; geometric spacetime is *reconstructed*; this is precisely the division of labor causal-set theory and algebraic QFT already use. The point matters twice over: it is also what keeps the first-order matter argument of Part V honest, since the locality O3 contributes there (excluding the nonlocal Foldy–Wouthuysen Hamiltonian) is logically prior to, and separable from, the Lorentz covariance that Part III supplies. The separation is real but should be *argued*, not assumed — which is exactly the second debt flagged in Section 37.

### 3.4 The redecomposition lemma (the seed of everything Euclidean)

**Lemma 3.2** (Redecomposition requires the Euclidean norm)**.** *Any continuous, strictly convex norm on* $\mathbb{R}^n$ *(*$n \geq 2$*) invariant under the orthogonal group* $O(n)$ *is a positive multiple of the Euclidean norm. Consequently, for independent systems in orthogonal subspaces,* $D(x \oplus y)^2 = D(x)^2 + D(y)^2$.

**Symbols and operations**

A *norm* $\|x\|$ measures the size of a vector. $O(n)$ is the group of rotations and reflections of $\mathbb{R}^n$ (the maps preserving right angles and lengths). "$O(n)$-invariant" means $\|Rx\| = \|x\|$ for every such $R$. The Euclidean norm is $\|x\|_2 = \sqrt{\sum_i x_i^2}$. The $\ell^p$ norms are $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$; $p = 2$ is Euclidean.

**In plain English**

If your way of measuring "how distinguishable is this state" is not allowed to care about which orthogonal axes you happened to draw, then it must be the ordinary Pythagorean length — nothing else survives. And the squared distinguishability of two independent things adds, exactly like $a^2 + b^2 = c^2$.

**Physical justification**

Splitting a state space into orthogonal sectors and gluing them back is pure bookkeeping — a redescription. By A0 it must leave total distinguishability untouched. The proof is short: $O(n)$ acts transitively on the unit sphere, so an invariant norm depends only on Euclidean length, $\|x\| = f(\|x\|_2)$; homogeneity forces $f(r) = cr$; and any $\ell^p$ with $p \neq 2$ is detectably *not* rotation-invariant — rotating $(1,0)$ by $45°$ changes its $\ell^p$ norm to $2^{1/p-1/2} \neq 1$. So a non-Euclidean distinguishability measure would let you detect a pure relabeling, which A0 forbids. The Pythagorean addition law for independent sectors is the same fact in disguise, and it returns with decisive force in the torsion argument of Section 5.

**Operational isotropy, not hardware isotropy**

The $O(n)$-invariance used here is not the claim that an observer's memory hardware has no preferred physical basis. A real register may have labels, thresholds, and an internal write order. The claim is narrower: once histories have been quotiented by operational distinguishability, passive redecompositions of *independent distinguishability axes* that preserve the independent-composition structure and the admissible protocol family are A0-inert. Under the regularity condition that independent sectors can be chosen without a privileged state-space axis, this redecomposition group acts as $O(n)$. Thus the Euclidean norm follows from operational isotropy of independent-sector descriptions, not from microscopic rotational symmetry of the observer.

# Part I — Quantum Mechanics

## 4 From histories to physical states

### 4.1 Operational distance and the quotient

We never see histories. We see statistics. So the only honest notion of "how different are two histories" is "how well can any experiment tell them apart."

$$d(H_1, H_2) = \sup_{\pi \text{ admissible}} D_{\mathrm{TV}}\big(X_{H_1}(\pi), X_{H_2}(\pi)\big),$$
$$H_1 \sim H_2 \iff d(H_1, H_2) = 0, \qquad S_{\mathcal{O}} = \mathcal{H}/\sim .$$

**Symbols and operations**

Read left to right: pick a protocol $\pi$; run it on each history to get two outcome distributions; measure their separation with $D_{\mathrm{TV}}$; then take the *supremum* over all admissible $\pi$ — the best any experiment can do. "$\mathcal{H}/\sim$" is the *quotient*: the set $\mathcal{H}$ with every $\sim$-equivalent batch of histories collapsed to a single point.

**In plain English**

The distance between two histories is the score of the best possible experiment at telling them apart. If even the best experiment scores zero, the histories are the same physical state. A "physical state" is therefore not a hidden microscopic truth — it is a whole equivalence class of histories that no experiment can separate.

**Physical justification**

This is the single operational move that turns physics into geometry. By A0, two histories no experiment can separate *are* the same physical situation, so the quotient $S_{\mathcal{O}}$ is forced, not chosen. Because $D_{\mathrm{TV}}$ is a metric (it is non-negative, symmetric, and obeys the triangle inequality) and the supremum of metrics preserves those properties, $d$ is automatically a *pseudometric* on $\mathcal{H}$ (distinct histories may sit at distance 0) and a genuine *metric* on the quotient $S_{\mathcal{O}}$ (distinct states sit at strictly positive distance). The well-definedness on classes is a one-line triangle-inequality check.

*Remark* (The equivalence relation is fixed, not dynamical). $\sim$ is set once and for all by the observer's constitution (O1, O2, O4). A system whose distinguishability rules drift under interaction is not one persistent observer but a sequence of different ones. O2 (stable register) and O4 (continuous proper time) are exactly what guarantee a single persistent observer with a single stable $\sim$.

### 4.2 Convex structure from randomized preparation

The quotient gives us a set of states with a distance on it, but no way to combine them yet. The first combination rule comes from something the observer can already do at finite precision: condition a preparation on a finite record. This directly supplies rational mixing, and its operational completion supplies the usual convex structure.

**Convex closure (finite randomized closure plus completion).** If preparations $P_1, P_2$ are admissible, a finite observer can condition on a finite record and choose "$P_1$ with rational probability $\lambda = m/N$, else $P_2$." Full convex combinations $\lambda s_1 + (1-\lambda)s_2 \in S_{\mathcal{O}}$ are the operational completion of this rational mixing structure under the distinguishability metric.

**In plain English**

If you can prepare two states, you can flip a (possibly biased) coin and prepare one or the other. So the state space has no holes between its points — you can always sit at any weighted average.

*Remark* (Assumption vs consequence). Convex *closure* is an operational assumption. It is worth splitting the claim in two, because half of it is actually a theorem and only the other half is an assumption:

- *Theorem (no assumption needed).* The *action on outcome statistics* is affine: $X_{\lambda s_1 + (1-\lambda)s_2}(\pi) = \lambda X_{s_1}(\pi) + (1-\lambda)X_{s_2}(\pi)$. This is just the law of total probability — the coin flip is independent of the later protocol, so observed frequencies are the weighted average. No extra physics.
- *Assumption (genuinely needed).* That the resulting mixture is *itself an admissible state* of $S_{\mathcal{O}}$ (closure of the set), rather than living outside it.

Stating it this way makes precise what convex closure buys: it does not manufacture the linearity of probabilities (that is free), it only asserts that the state space is closed under the mixing you can already perform. This matters at the completeness question, where the completeness of the Born-rule reconstruction hinges on exactly which structure is assumed versus derived.

### 4.3 Composition, cancellativity, and the Grothendieck group

We now have a state space that is metric (from $d$) and convex (from randomized preparation). Neither of those, by itself, gives us the *algebra* physics runs on — the ability to add states, scale them, and eventually treat them as vectors. That structure has to come from somewhere physical, and the only physical operation available is the one the observer can actually perform: putting two non-interacting systems side by side and treating them as one. The next three results extract a full Abelian group from exactly that operation, in three moves — combine (monoid), undo (cancellativity), complete (Grothendieck) — each forced by a specific observer constraint rather than assumed.

**Proposition 4.1** (Symmetric composition)**.** *For causally disconnected sectors $A, B$: $\mathcal{H}_A \circ \mathcal{H}_B \sim \mathcal{H}_B \circ \mathcal{H}_A$. Restricted to such sectors, $(S_{\mathcal{O}}, \oplus)$ is a commutative, locally compact, Hausdorff topological monoid.*

**Symbols and operations**

A *monoid* is a set with an associative way to combine elements and an identity element (here $\varnothing$), but *no* guarantee you can undo a combination. "Commutative/Abelian" means order does not matter, $a \oplus b = b \oplus a$. "Locally compact + Hausdorff" are mild topological niceness conditions (points are separated; small neighbourhoods are well-behaved), inherited from O2 and the quotient metric.

**In plain English**

Put two non-communicating systems side by side. The order in which you list them is a choice of bookkeeping, not a physical fact, so "$A$ then $B$" and "$B$ then $A$" are the same state. Combining is associative and has a do-nothing element (the empty system). That is exactly a commutative monoid.

**Physical justification**

Independent systems factorize statistically: $X_{H_A \oplus H_B}(\pi_A \otimes \pi_B) = X_{H_A}(\pi_A)\, X_{H_B}(\pi_B)$. This factorization is not a modelling convenience — it is the operational *content* of "causally disconnected." If the joint statistics did *not* factor, then some correlation would survive between sectors that cannot signal, and a protocol on one side would shift outcomes on the other: that is precisely the signalling O3 forbids. So independence $\Rightarrow$ factorization is forced, and multiplication of real numbers is associative and commutative, so composition inherits both properties for free. The identity element is the empty system $\varnothing$: composing with "nothing" multiplies every outcome probability by 1 and changes no statistic. Associativity, commutativity, identity — that is exactly the axiom list for a commutative monoid, and every entry on the list traces to a physical fact rather than a postulate.

A monoid lets us combine but not undo. To reach a group — and ultimately the inverses and scalars a vector space needs — we need to know that composition never destroys information about its parts. That is the next lemma, and it is again O3 that supplies it.

**Lemma 4.2** (Cancellativity). *If $[H_1] \oplus [H_3] = [H_2] \oplus [H_3]$ for any causally independent $H_3$, then $[H_1] = [H_2]$.*

**In plain English**

If gluing the same independent background onto two states makes them look identical, the states were already identical. You cannot hide a real difference behind a shared backdrop.

**Physical justification**

This is O3 (no signalling) doing the work a second time, now in reverse. During causal independence the observer's apparatus cannot reach into $H_3$, so the statistics of any protocol run on $H_1 \oplus H_3$ are determined by $H_1$ alone — the $H_3$ factor contributes the same multiplicative constant on both sides and cancels. Concretely: if $H_1 \oplus H_3$ and $H_2 \oplus H_3$ are indistinguishable for every protocol, then dividing out the common $H_3$ statistics leaves $H_1$ and $H_2$ indistinguishable for every protocol, which is exactly $d(H_1, H_2) = 0$, i.e. $[H_1] = [H_2]$. The physical reading is that a shared, untouchable background cannot generate or mask any real difference between the foregrounds. Cancellativity is the one extra property a commutative monoid must have before it can embed faithfully into a group — without it the group completion would glue together states that are genuinely distinct.

With combination (monoid) and lossless undoing (cancellativity) both secured, the standard algebraic completion applies verbatim, and it is the same one that builds $\mathbb{Z}$ out of $\mathbb{N}$.

**Corollary 4.3** (Grothendieck group)**.** *The Grothendieck group* $G(S_{\mathcal{O}})$ *— formal differences* $[s] - [t]$ *with* $(s, t) \sim (s', t') \iff s \oplus t' = s' \oplus t$ *— is a locally compact Hausdorff Abelian group, and* $S_{\mathcal{O}}$ *embeds in it as the positive cone via* $\iota([s]) = [s] - [\theta]$.

**Symbols and operations**

The *Grothendieck group* is the standard "add subtraction" construction — the same trick that builds the integers $\mathbb{Z}$ from the counting numbers $\mathbb{N}$. A *formal difference* $[s] - [t]$ need not be a preparable state; it is bookkeeping. The *positive cone* is the image of the genuine states inside the group. "$\iota$" (iota) is the embedding map.

**In plain English**

Monoids let you add but not subtract. To get the symmetry physics needs — translations, inverses, eventually a vector space — you formally allow differences of states, the way $3 - 5 = -2$ makes sense even though you cannot have $-2$ apples. The negative elements are scaffolding; no physical claim is made about them. Cancellativity (Lemma 4.2) is exactly what guarantees the real states inject cleanly into this larger group.

*Remark* (Partiality of $\oplus$). $\oplus$ is a *partial* operation: it is defined only for causally *independent* sectors. The monoid, the cancellativity lemma, and the Grothendieck completion all live on this partial structure. The algebra being completed is the algebra of *independent composition*, and "$n\theta$" always means "$n$ mutually independent copies in orthogonal sectors," never "the same system counted $n$ times." This reading is what makes the Pythagorean law of Lemma 3.2 applicable to group multiples, which is the hinge of the next section.

**Causal independence before geometry**

Here "causally independent" is still an O3 notion, not a Lorentzian one. It means that during the protocol there is no finite chain of local interactions by which one sector can update the records of the other, except through later comparison of already-written records. Part III reconstructs the geometric representation of this primitive accessibility relation as causal order and Lorentzian metric structure. The composition algebra therefore uses pre-metric operational independence, not downstream spacetime geometry.

*Remark* (Completeness). A0 forces metric completion: a Cauchy sequence under $d$ has the property that no admissible protocol distinguishes its late elements, so refusing to admit the limit as a state would assert an operationally invisible distinction — which A0 forbids.

## 5 Real vector space structure and torsion-freeness

We now have a complete, cancellative, locally compact, Abelian group with a translation-invariant metric. Three more topological facts turn it into $\mathbb{R}^n$.

**Sectorwise continuous accessibility.** The reconstruction is restricted to a single irreducible dynamically accessible component. Within that component, states considered in the reconstruction are reached from $\theta$ by admissible finite interaction sequences (O3), and each interaction is continuous in the observer's proper time (O4). Disconnected components are superselection sectors and are reconstructed separately.

**Torsion-freeness.** No nonzero element has finite order: there is no $\theta \neq 0$ with $n\theta = 0$.

**Pontryagin.** A complete, torsion-free, locally compact, connected, second-countable Abelian group in the reconstructed component is a finite-dimensional real vector space:

$$S_{\mathcal{O}} \hookrightarrow G(S_{\mathcal{O}}) \cong \mathbb{R}^n, \qquad 2 \le n \le k.$$

**Symbols and operations**

*Torsion* means a nonzero element $\theta$ that returns to the identity after finitely many additions ($n\theta = \theta \oplus \cdots \oplus \theta = 0$) — like the hours on a clock face, where $12 + 12 = 0$. A group is *torsion-free* if this never happens except trivially. *Second-countable* is a smallness condition on the topology (a countable base of open sets). The *Pontryagin structure theorem* classifies all groups meeting these conditions: the only one is $\mathbb{R}^n$.

**In plain English**

Pull together what we have: physical states can be scaled continuously, combined with independent things, cancelled reversibly, and form a connected, well-behaved space with no "clock-face wraparound." The only mathematical object that is all of these at once is ordinary $n$-dimensional real space. So the state space is (sits inside) $\mathbb{R}^n$.

**Physical justification**

Torsion-freeness is the one nontrivial ingredient, and it is genuinely needed: a connected group can still have torsion (the circle $U(1) = \mathbb{R}/\mathbb{Z}$ is connected, locally compact, one-dimensional — and is *not* $\mathbb{R}$). Without ruling out torsion, Pontryagin would permit $U(1)$-like wraparound instead of a vector space. So we must show distinguishability cannot cycle back to zero under repeated independent composition.

**Torsion-freeness.** To exclude torsion one needs a *lower* bound on $d(n\theta, 0)$: a triangle-inequality estimate would only bound it from above, $d(n\theta, 0) \le n\, d(\theta, 0)$, which is consistent with $n\theta = 0$ and proves nothing. The Pythagorean composition law of Lemma 3.2 supplies the lower bound directly. By the partiality of $\oplus$, the group multiple $n\theta$ is $n$ mutually independent copies of $\theta$ in orthogonal sectors, and independent orthogonal composition adds *squared* distinguishability, so

$$\|n\theta\|^2 \;=\; n\,\|\theta\|^2, \qquad \text{i.e.} \qquad d(n\theta, 0) = \sqrt{n}\, d(\theta, 0).$$

If $\theta \neq 0$ then $d(n\theta, 0) = \sqrt{n}\, d(\theta, 0) > 0$ for every $n \ge 1$, so $n\theta \neq 0$: no element has finite order. The bound uses only a law already in hand from A0, no new import. (Independently, a connected, divisible, locally compact Abelian group is torsion-free, which covers group elements not of the independent-copies form.)

*Remark* (Topological fine print). Hausdorff and local compactness come from the operational metric and O2. Second-countability enters as the countable operational protocol-base condition: neighbourhoods are generated by finite lists of finitely describable protocols with rational tolerances. Connectedness enters sectorwise through continuous accessibility inside one irreducible component. These are operational regularity conditions consistent with O1–O4, not deductions from O1–O4 alone.

## 6 The Euclidean norm and the inner product

We have an $n$-dimensional real vector space, but a bare vector space has no notion of length or angle. The observer does have one notion of size already — distinguishability from the zero state — so the question is what *shape* that size-measure has. The next lemma shows it must be the round, Euclidean one, and that single fact is what later forces the inner product, the complex structure, and ultimately the Hilbert space.

**Lemma 6.1** ($O(n)$-invariance of the distinguishability norm)**.** *The norm $\|x\| = d(x,0)$ on $\mathbb{R}^n$ is invariant under every orthogonal transformation $R \in O(n)$: $d(Rx,0) = d(x,0)$.*

**In plain English**

Rotating which axes you call "the basis" is just relabelling which detector clicks count as which outcome. The physical interaction, and the mark it leaves in the register, are unchanged. So the distinguishability of a state cannot depend on that choice of axes.

**Physical justification**

A rotation $R$ carries one orthonormal description basis to another; both describe the same state space provided the rotation is a passive redecomposition of independent distinguishability axes, not a physical scrambling of the observer's memory hardware. An admissible protocol couples to physical states, not to the symbol chosen for a basis element, and the memory register records a classical outcome *index*. Relabelling which index names which redecomposed axis leaves the interaction and the register's post-recording physical state alone. So to every protocol $\pi$ on $Rx$ corresponds a relabelled protocol $\pi'$ on $x$ with identical statistics; the map $\pi \mapsto \pi'$ is a bijection, so the suprema (the distances) agree. If an observable genuinely singles out a physical axis, that is not an inert rotation and is not included in the $O(n)$-invariance claim.

Combining with Lemma 3.2: the norm is $O(n)$-invariant, hence Euclidean, hence $\|x \oplus y\|^2 = \|x\|^2 + \|y\|^2$, and by the parallelogram law (Jordan–von Neumann) it comes from an inner product:

$$\langle x, y\rangle = \tfrac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right).$$

**Symbols and operations**

The *parallelogram law*, $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$, is the exact algebraic fingerprint of a norm that comes from an inner product $\langle \cdot, \cdot \rangle$. The displayed *polarization identity* reconstructs that inner product from lengths alone. An *inner product* is the abstract dot product: it gives lengths ($\|x\|^2 = \langle x, x\rangle$) and angles (orthogonality is $\langle x, y\rangle = 0$).

**In plain English**

Once the size-measure is Pythagorean, the geometry has angles, not just lengths — there is a well-defined dot product. The state space is now a real inner-product space: it has perpendicularity, projections, the works.

## 7 The phase group is U(1)

Free evolution (no interaction, hence no information gained) must preserve the Euclidean norm, so it is a one-parameter subgroup of $O(n)$ of the form $\exp(tA)$ with $A$ skew-symmetric. The real canonical form splits $\mathbb{R}^n$ into two-dimensional rotating planes $\Pi_1, \ldots, \Pi_m$.

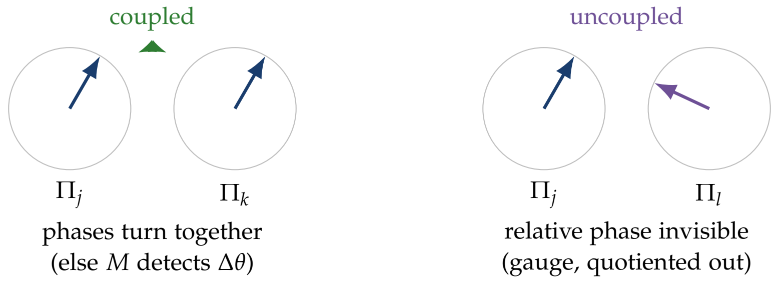


Figure 2: Why one phase survives inside an irreducible observer sector. Free evolution turns each invariant 2-plane $\Pi_j$ at its own rate. Coupled planes must co-rotate because an independent relative rotation $\Delta\theta$ changes a measurable expectation. Permanently uncoupled planes define separate superselection sectors or separate observers; their relative phases are gauge. Within one connected sector a single global phase remains: $U(1)$.

**Lemma 7.1** (Phase group from pairwise structure)**.** *For one irreducible operational sector, the unique physically inert, norm-preserving freedom is the* global *rotation* $\psi \mapsto e^{i\theta}\psi$*. The phase group of such a sector is* $U(1)$*.*

**Symbols and operations**

$A$ *skew-symmetric* means $A^\top = -A$; exponentiating it gives rotations. Each invariant plane $\Pi_j$ is a 2D subspace that the dynamics simply rotates. $U(1) = \{e^{i\theta}\}$ is the circle group of phases, $|e^{i\theta}| = 1$. A "cross-block" $M_{jk}$ of an observable $M$ is the part that links plane $j$ to plane $k$.

**In plain English**

An isolated system has internal motion that changes no probabilities — pure "spinning in place." This spinning decomposes into independent 2D rotations. Now ask: can two of these planes spin at *different* rates without anyone noticing? Only if nothing ever couples them. If any allowed measurement links two planes, spinning them differently shows up as an interference term someone can measure — so A0 forbids it and forces them to spin together. If two planes never couple, their relative phase was never observable and got quotiented away already, or the planes belong to different superselection sectors. For one coherent observer sector, exactly one overall spin rate is left: one global phase. That is the circle $U(1)$.

**Physical justification**

Build a graph on the planes with an edge wherever some admissible observable has a nonzero cross-block. For a coupled pair, an independent relative rotation $\Delta\theta \neq 0$ changes the expectation by the cross term $2\,\mathrm{Re}(e^{i\Delta\theta}\langle\psi_j|M|\psi_k\rangle)$, which depends on $\Delta\theta$ — not inert, so A0 synchronizes them. For an uncoupled pair, every admissible expectation depends only on the moduli $|a|, |b|$, so the relative phase is pure gauge and already gone. Within each connected component synchrony is selected; across permanently disconnected components the relative phase is gauge, or the components are better treated as separate superselection sectors rather than one irreducible observer. Thus the theorem is sectorwise: one coherent observer sector has one physical inert phase parameter, not $U(1)^m$.

**Irreducible observer sector**

The connectivity assumption is not a hidden claim that every mathematical two-plane in a reducible direct sum must couple to every other. It is a definition of the sector being reconstructed: a single observer sector is one coherent physical system whose record-bearing degrees of freedom can, in principle, affect one another through admissible dynamics. If the coupling graph decomposes permanently, the components are separate observers or superselection sectors. The $U(1)$ result applies to each irreducible sector.

## 8 The scalars are $\mathbb{C}$: complex Hilbert space

We now have a real inner-product space carrying a distinguished $U(1)$ of inert phase rotations. A vector space is built over some number field, and so far we have only assumed the reals. But the phase group just derived behaves like multiplication by a square root of $-1$, which the reals cannot supply — so the question of *which* scalars the theory runs on is now forced, and the answer is the source of the $i$ in quantum mechanics.

**Theorem 8.1** (The field is $\mathbb{C}$)**.** *The scalar field over which the state space is built must be a finite-dimensional associative division algebra over $\mathbb{R}$ containing the sectorwise $U(1)$ phase action and compatible with symmetric independent composition. By Frobenius's theorem the only options are $\mathbb{R}, \mathbb{C}, \mathbb{H}$; $\mathbb{R}$ has no continuous scalar phase, and $\mathbb{H}$ is incompatible with operational local completeness for independent composites. The surviving field is $\mathbb{C}$, so the state space is a finite-dimensional complex Hilbert space $\mathcal{H}_\mathcal{O} \cong \mathbb{C}^m$.*

**Symbols and operations**

A *division algebra* is a number system where every nonzero element has a multiplicative inverse (you can always divide). *Associative*: $(ab)c = a(bc)$. $\mathbb{R}$ = reals, $\mathbb{C}$ = complex numbers, $\mathbb{H}$ = quaternions (a 4D number system where $ab \neq ba$). *Frobenius's theorem* says these three are the *only* finite-dimensional associative real division algebras. $J$ is the operator implementing the $U(1)$ generator on each plane; $J^2 = -I$ makes it "multiplication by $i$."

**In plain English**

What kind of numbers multiply the states? They must form a system where you can always divide (otherwise a harmless rescaling could crush a nonzero state to zero — forbidden by A0), where multiplication order does not matter (forced by the symmetric composition of independent systems), and which contains the circle of phases just derived. Run down Frobenius's short list: the reals have no continuous phase, the quaternions multiply in the wrong order, only the complex numbers fit. So the scalars are complex — and that is *where* the $i$ of quantum mechanics comes from. It is not put in by hand; it is the last survivor.

**Physical justification**

Division is selected by the requirement that reversible rescalings of nonzero operational states remain reversible. The real option is too small: it cannot realize the continuous phase action from Section 7 as scalar multiplication. The quaternionic option is too large and too order-sensitive: left and right scalar actions do not commute, and independent subsystem phases become composition-order dependent. In finite composites this produces precisely the anomalous locally invisible degrees of freedom that operational local completeness forbids. Thus $\mathbb{H}$ is excluded not merely because it contains non-commuting symbols, but because its scalar structure conflicts with A0+O3 independent composition. With $J^2 = -I$ in hand, define

$$\langle x, y \rangle_{\mathbb{C}} = \langle x, y \rangle_{\mathbb{R}} + i \, \langle Jx, y \rangle_{\mathbb{R}},$$

a sesquilinear, conjugate-symmetric, positive-definite form: a genuine complex inner product. The reconstruction of the Hilbert space is complete.

**Why quaternions fail**

The quaternions do contain copies of $U(1)$, so they cannot be excluded by the mere existence of a phase. They fail at composition. Independent systems require symmetric, order-insensitive local actions whose composite statistics are exhausted by local record protocols. Quaternionic scalar multiplication makes left/right phase order physically relevant in composites and is well known to generate anomalous composition behavior. Since the next section derives operational local completeness for independent composites, $\mathbb{H}$ is not an admissible scalar field for one embedded-observer theory.

**Where each competitor dies.** Every alternative to standard quantum theory fails on this stretch of the chain, and it is clearest to see the eliminations in sequence: classical probability at Section 7 (no phase group); real Hilbert space here ($J^2 = -I$ has no real solution); quaternionic QM here (non-commutative, by Frobenius and anomalous composition); non-Euclidean/$\ell^p$ state geometry back at Lemma 3.2 (redecomposition becomes detectable); super-quantum no-signalling boxes after the Born rule (they exceed the Hilbert/Born correlation geometry, not O3 alone); and the exceptional Jordan algebra $J_3^8$ at independent-composite local completeness (next section).

## 9 Tensor-product composition

The single-system Hilbert space is now built. Section 4 established that independent systems compose operationally, but it did not yet identify the Hilbert-space operation that represents this composition. Having complex Hilbert spaces in hand, we can finally ask: when two causally independent systems are treated jointly, is the representing space a tensor product, a direct sum, or something stranger?

**Proposition 9.1** (Operational local completeness as a closure principle)**.** *For causally independent sectors $A, B$, the physical state of the independent composite is taken to be exhausted by the statistics of local record-forming protocols on $A$ and $B$, together with ordinary comparison of the resulting records. If two candidate joint states agree on all such admissible statistics, they are identified as the same operational composite state.*

**Theorem 9.2** (Tensor composition)**.** *For independently composable finite systems with reconstructed single-system Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$, bilinear independent preparation, bilinear independent effects, and operational local completeness, the representing composite space is*

$$\mathcal{H}_{AB} \cong \mathcal{H}_A \otimes \mathcal{H}_B.$$

**Symbols and operations**

*Local operations and classical communication (LOCC)* means that each side interacts only with its own system and the resulting classical records are later compared through ordinary causal communication. *Operational local completeness* means that there is no extra joint degree of freedom in an independent composite that is invisible to every admissible local-record protocol. The symbol $\otimes$ denotes the universal bilinear product of Hilbert spaces.

**In plain English**

Two systems that cannot interact cannot be probed by a magic joint measurement from nowhere. Each side can only be touched locally, and whatever is learned must be written into records that can later be compared. If a supposed difference between two joint states makes no difference to any such comparison, then by A0 it is not an additional physical difference. This is the observer-first content behind local completeness: independent composites contain exactly the structure that local record protocols can resolve.

**Physical justification**

O3 identifies the admissible access to causally independent sectors: local interactions on $A$, local interactions on $B$, and finite-speed communication of the resulting records. A0 then motivates quotienting away any proposed distinction between independent composite states that changes none of those statistics. Thus local completeness is not a free theorem of locality alone; it is the operational closure condition for independent composition under A0 and O3, stated explicitly to exclude globally hidden composite degrees of freedom inaccessible to admissible local-record protocols.

The remaining step is representational. Once the factors have already been reconstructed as complex Hilbert spaces, independent preparation is bilinear in the two factors and independent effects are bilinear in the two dual effect spaces. The object representing all such bilinear pairings is the tensor product. Product effects separate the operational states by the preceding proposition, so no additional locally invisible sector may be appended to the independent composite. Direct sums have the wrong bilinear structure; real and quaternionic alternatives have already been eliminated; exceptional Jordan systems fail this local-completeness condition for composites. The result is the ordinary quantum composition rule $\mathcal{H}_A \otimes \mathcal{H}_B$.

This does not say that only product states exist. Entangled states are precisely the non-product states in $\mathcal{H}_A \otimes \mathcal{H}_B$, generated when formerly independent systems interact and are later treated as one composite. The claim is narrower and sharper: there is no extra independent-composite structure beyond the tensor product that an admissible observer could distinguish.

**What is, and is not, proved here**

A0 and O3 motivate operational local completeness for independent composites, but the tensor-product step assumes this closure condition explicitly. The identification with $\mathcal{H}_A \otimes \mathcal{H}_B$ also uses the already-derived complex Hilbert structure of the factors and the universal property of bilinear composition. Thus local completeness is not a free-standing axiom on the level of the cyclic separating vacuum; it is the composite-level A0+O3 closure principle that independent composites carry no globally hidden degrees of freedom inaccessible to all admissible local-record protocols. Super-quantum no-signalling boxes are not eliminated merely by O3, since O3 forbids superluminal signalling rather than super-quantum correlation; they are excluded, if at all, by the Hilbert-space and Born-rule reconstruction that restricts correlations to the quantum convex set.

# 10 The Born rule and unitary dynamics

## 10.1 Born rule via the qubit-gap closure

The observer is *always present* (it is doing the measuring) with at least two distinguishable states (O2), so a real measurement lives on the joint space $\mathcal{H}_{\text{joint}} = \mathcal{H}_O \otimes \mathcal{H}_S$ of dimension $\geq 4 \geq 3$.

- **Move (i) — frame function.** Two orthonormal bases sharing a ray $e_1$ differ by an orthogonal map fixing the projector $P_{e_1}$; by A0 that is inert at the $e_1$ outcome, so $P(e_1)$ depends only on $P_{e_1}$, not on the rest of the basis. That is exactly Gleason's frame-function condition.
- **Move (ii) — Gleason.** Since dim $\geq 3$, Gleason's theorem gives $P(E) = \mathrm{Tr}(\rho_{OS} P_E)$; the system rule follows by partial trace, $P(E_S) = \mathrm{Tr}_S[\rho_S P_{E_S}]$ with $\rho_S = \mathrm{Tr}_O \rho_{OS}$.

**Symbols and operations**

A *ray* is a one-dimensional subspace (a state up to phase). $P_{e_1}$ projects onto it. A *frame function* assigns a number to each ray so that the numbers over any orthonormal basis sum to 1, independent of the rest of the basis. *Gleason's theorem*: on a Hilbert space of dimension $\geq 3$, every frame function has the form $\mathrm{Tr}(\rho P)$ for a unique density operator $\rho$. *Partial trace* $\mathrm{Tr}_O$ "averages out" the observer to leave the system's state $\rho_S$.

**In plain English**

The probability of an outcome should depend only on *that* outcome, not on what other answers were on the menu. That context-independence, plus "probabilities sum to one," is a frame function. Gleason's theorem then says: on any space of dimension three or more, the only consistent way to assign such probabilities is the trace rule — the Born rule. The famous loophole is that Gleason fails for a lone qubit (dimension 2). The framework closes it structurally: because the observer is always there with at least two states of its own, the real arena is always at least $2 \times 2 = 4$-dimensional, comfortably inside Gleason's reach.

**Physical justification**

The Born rule is *not* derived from A0 alone. A0 plus the operational construction establishes the frame-function property; Gleason (imported) supplies the trace form. The OFP-specific contribution is the qubit-gap closure: the observer's own $\geq 2$ states guarantee dim $\geq 4$, so Gleason always applies and the system rule drops out by partial trace.

### 10.2 Unitary dynamics

**Theorem 10.1** (Unitarity, in the refinement limit)**.** *Free evolution is unitary, $U(\tau) = e^{-iH\tau/\hbar}$ with $H$ self-adjoint, in the unbounded-resource limit where the proper-time parameter of O4 is continuous.*

**In plain English**

Evolution with no measurement cannot create or destroy distinguishability, so it preserves the inner product. Wigner's theorem says such a map is either unitary or anti-unitary. Anti-unitary maps reverse time's direction and cannot be reached continuously from "do nothing"; since the observer's clock runs continuously from the identity (O4), only the unitary branch survives. Stone's theorem then writes any such continuous family as $e^{-iH\tau/\hbar}$, and $H$ is the observer's clock generator.

*Remark.* Stone's theorem needs a *genuinely continuous* one-parameter group, and O4 supplies that only in the unbounded-resource refinement limit; at finite resources the proper-time parameter is discrete at $\tau_{\mathcal{O}}$ (O1 via Margolus–Levitin), so the evolution is a discrete-step approximant to the continuous unitary group. Continuous unitarity is therefore exact in the limit and approximate at finite resources — a real correction to any unconditional "derived unitarity" reading, carried wherever Stone is invoked (including Parts II and IV).

**Physical justification**

$U(\mathcal{H})$ is the identity component of the isometry group; the anti-unitaries are the other component. A continuous one-parameter group through the identity (O4) stays in the identity component (Wigner), and Stone's theorem gives the exponential form with self-adjoint generator $H$. This $H$ is the same clock generator that returns in Parts II and IV — the single thread tying the four domains together.

## 11 Wave–particle duality as record–coherence complementarity

The ordinary phrase "wave–particle duality" is misleading if it suggests that a quantum object changes species. In this framework the duality is more precise: it is the tradeoff between phase coherence among alternatives and the creation of finite distinguishable records. The visibility–distinguishability relation used below is standard in two-path complementarity (Jaeger–Shimony–Vaidman; Englert); the contribution here is not the inequality itself but its observer-first reading. A wave is not a second substance. It is the phase-sensitive amplitude structure over alternatives that no admissible record has yet separated. A particle is not a tiny classical bead hidden inside the wave. It is the localized record-event produced when one alternative becomes physically distinguishable.

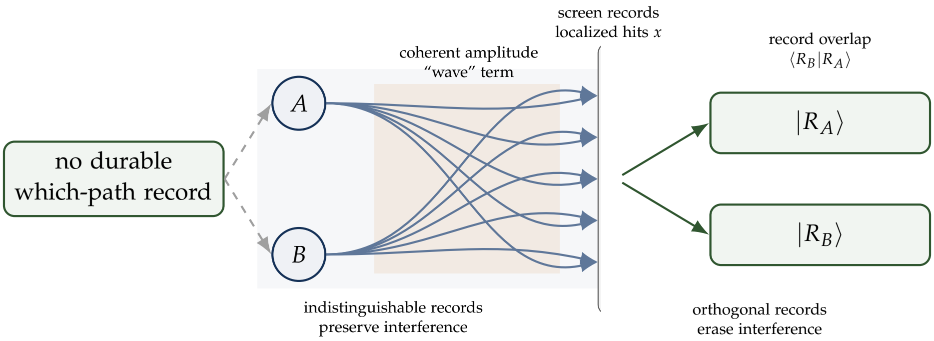


Figure 3: Wave–particle duality in observer-first form. "Wave" is phase coherence among alternatives not yet separated by a durable record. "Particle" is the localized record-event produced when one alternative becomes physically distinguishable. The same record overlap controls both; the figure gives the standard complementarity relation an observer-first reading.

Let two operational alternatives be represented by orthogonal path states $|A\rangle$ and $|B\rangle$. Before a which-path record exists,

$$|\psi\rangle = \alpha|A\rangle + \beta|B\rangle.$$

At a screen point $x$, the amplitude is $\psi(x) = \alpha\psi_A(x) + \beta\psi_B(x)$, so the Born rule gives

$$P(x) = |\alpha\psi_A(x) + \beta\psi_B(x)|^2.$$

The interference term is the part that depends on the relative phase:

$$P(x) = |\alpha\psi_A(x)|^2 + |\beta\psi_B(x)|^2 + 2\,\mathrm{Re}\big(\alpha\beta^*\psi_A(x)\psi_B^*(x)\big).$$

Now let the observer or environment acquire a which-path record:

$$|\Psi\rangle = \alpha|A\rangle|R_A\rangle + \beta|B\rangle|R_B\rangle.$$

Tracing over the record degrees of freedom multiplies the interference term by the record overlap:

$$P(x) = |\alpha\psi_A(x)|^2 + |\beta\psi_B(x)|^2 + 2\,\mathrm{Re}\big(\alpha\beta^*\psi_A(x)\psi_B^*(x)\langle R_B|R_A\rangle\big).$$

If $\langle R_B|R_A\rangle = 1$, no admissible record distinguishes the paths and the interference pattern remains. If $\langle R_B|R_A\rangle = 0$, the records are orthogonal, the path distinction is physical, and the interference term vanishes.

**Symbols and operations**

$|A\rangle, |B\rangle$ are two alternative paths; $|R_A\rangle, |R_B\rangle$ are record states correlated with those paths. The overlap $\langle R_B|R_A\rangle$ measures how distinguishable the records are. Orthogonal records, $\langle R_B|R_A\rangle = 0$, can be perfectly told apart. Identical records, $\langle R_B|R_A\rangle = 1$, contain no which-path information. The screen coordinate $x$ is a localized detector record.

**In plain English**

The double-slit experiment is not telling us that matter is confused about whether it is a wave or a particle. It is telling us that two different record protocols expose two incompatible aspects of one Hilbert-space state. If the paths remain unrecorded, their phases combine and produce interference: wave behavior. If the paths become durable records, the alternatives are separated and the screen registers localized hits without interference: particle behavior. The mystery is not that both pictures appear; the structure says exactly when each appears.

**Physical justification**

This is A0 and O2 in their most familiar laboratory costume. When no admissible record distinguishes two alternatives, treating them as separate physical facts would add structure with no operational content. The alternatives therefore remain one coherent amplitude. When an interaction writes a distinguishable record, the difference becomes physical; coherence is exported into the enlarged observer–environment state and is no longer available as interference in the reduced screen statistics. Collapse is not added here as a new dynamical law. It is the irreversible stabilization of one finite record relative to the observer.

**Scope of the duality claim**

The complementarity formula below is not presented as a new quantum theorem. It is the standard two-path visibility–distinguishability relation read through A0 and finite record formation. The new claim is interpretive and structural: in an observer-first reconstruction, "wave" names coherent amplitude among alternatives not yet separated by a record, while "particle" names the localized finite record produced when such a distinction becomes physical.

The standard two-path visibility–distinguishability tradeoff expresses the same point quantitatively. With balanced pure paths one may write

$$V = |\langle R_B|R_A\rangle|, \qquad D = \sqrt{1 - |\langle R_B|R_A\rangle|^2}, \qquad V^2 + D^2 = 1,$$

and more generally $V^2 + D^2 \leq 1$. Here $V$ is interference visibility and $D$ is which-path distinguishability. Perfect wave visibility means no path record; perfect particle distinguishability means no interference. The complementarity is therefore not a verbal paradox but an algebraic conservation law between coherence and recordability.

## 12 Two cautious corollaries: classical action and electromagnetism

### 12.1 Schrödinger = multi-valued classical least action (Madelung–Bohm)

The substitution $\psi = \sqrt{\rho}\, e^{iS/\hbar}$ rewrites the Schrödinger equation *exactly* as a continuity equation for $\rho$ plus a Hamilton–Jacobi equation for $S$ with the quantum potential

$$Q = -\frac{\hbar^2}{2m}\frac{\nabla^2\sqrt{\rho}}{\sqrt{\rho}}.$$

**Symbols and operations**

$\psi$ is the wavefunction; $\rho = |\psi|^2$ is the probability density; $S$ is the classical action (the phase). $\nabla^2$ is the Laplacian (a second derivative measuring curvature). $Q$ is the extra "quantum" force term. A *continuity equation* expresses conservation of probability; *Hamilton–Jacobi* is classical mechanics written in terms of the action.

**In plain English**

Quantum dynamics is not a different kind of mechanics — it is classical least-action mechanics where the action is allowed to be *multi-valued* and carries one extra potential term. The transformation is exact and is itself an A0-invariant rewriting, so this is a re-description, not new physics. It is why the framework can treat "collapse" as a branch selection from a multi-valued action rather than a mysterious extra process.

### 12.2 Electromagnetism from local phase freedom

**Theorem 12.1** (Gauge structure)**.** *Demanding that the action stay stationary under* local *phase rotations $\psi(x,\tau) \mapsto e^{i\theta(x,\tau)}\psi$ forces a compensating field $A_\mu$ with covariant derivative $D_\mu = \partial_\mu - ieA_\mu$ and kinetic term $-\frac{1}{4}\int F_{\mu\nu}F^{\mu\nu}\, d^4x$: Maxwell's equations. The mediator is massless — a Proca mass term would break the local phase symmetry, violating A0.*

**Symbols and operations**

A *local* phase $\theta(x,\tau)$ varies from point to point, unlike a global constant. $A_\mu$ is the gauge potential; $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ its field strength; $e$ the coupling (charge). A *Proca mass term* $\propto m^2 A_\mu A^\mu$ would give the photon a mass.

**In plain English**

A global phase rotation is inert. But an observer is *local*: it can only fix a phase convention along its own worldline, and forcing one global phase across spacelike-separated regions would need faster-than-light coordination (banned by O3). So the truly inert freedom is *arbitrary local* phase choices. Insisting the physics not notice them forces a new field to "connect" the conventions at neighbouring points — and that connection field is the electromagnetic field. The photon must be massless because a mass term would re-expose the local phase, making an inert transformation observable.

**Physical justification**

Beyond A0 and locality, the theorem also uses *minimality* (an Occam choice, not from A0) and *Lorentz invariance* (supplied globally by Part III). The value of $e$ stays empirical. Identifying this U(1) gauge field with real electromagnetism is empirical; extending to the non-Abelian SU(2)×SU(3) of the Standard Model is *not* derived (a remaining model-selection question).

## 13 Arrow of time, the Second Law, and collapse

The strict causal ordering of the finite register (O2) *is* the observer's internal arrow of time. When the register fills, new records require erasure; by Landauer's principle each erased bit dissipates at least $k_B T \ln 2$:

$$\frac{dS_{\text{env}}}{d\tau} \geq k_B \ln 2 \cdot r > 0,$$

where $r$ is the bit-erasure rate.

**Symbols and operations**

$S_{\text{env}}$ is the entropy of the environment; $\tau$ proper time; $r$ bits erased per unit time; $k_B$ Boltzmann's constant; $T$ temperature. *Landauer's principle* (imported): logically erasing one bit costs at least $k_B T \ln 2$ of dissipated heat.

**In plain English**

You remember the past and not the future because your memory register is ordered and finite. To keep recording, you must eventually erase, and erasing is thermodynamically irreversible — it dumps heat. So the Second Law is not an extra cosmic postulate; it is the running tab of a finite recorder. In the same breath, "wave collapse" is identified with an irreversible *write* to that register: it is irreversible (O2 + Landauer), it selects one branch of the multi-valued action, and it produces environmental entropy. Arrow of time, Born rule, and collapse are three faces of one constraint.

**Physical justification**

The mechanism is concrete. A finite register (O1) with an ordered write-discipline (O2) can only keep recording by overwriting old cells, and Landauer's principle says erasing one bit in contact with a bath at temperature $T$ dissipates at least $k_B T \ln 2$ of heat into that bath. Every recorded observation therefore has a strictly positive minimum entropy cost, paid to the environment and never recovered — which is the Second Law stated at the level of a single observer rather than an ensemble. The *direction* of time is the direction in which the register fills: "past" is what has been written, "future" is what has not, and the asymmetry is the irreversibility of the write, not a boundary condition imposed by hand. What the framework deliberately does *not* settle is the metaphysics underneath the write: whether recording *creates* the outcome (a physical collapse) or merely *reveals* a pre-existing branch is left as an explicit empirical/interpretive question, so that the thermodynamic result does not silently smuggle in a measurement-problem stance.

# Part II — Finite Operational Energy and Protocol Refinement

Part I reconstructed the state-space structure available to an embedded observer. Part II asks a narrower laboratory question: what happens to an ideal protocol when the observer/controller that implements it is treated as a finite-energy physical system? The theorem-level answer is not a universal timing-noise law. It is a minimum duration for distinguishability-changing operations, and therefore a maximum number of protocol refinements in a finite interval.

## 14 Minimum operation time from finite operational energy

Consider a resonant two-level control pulse. In the rotating frame, under the usual rotating-wave approximation, the effective pulse Hamiltonian is

$$H_{\text{eff}}^{\text{pulse}} = \frac{\hbar\Omega_R}{2}\sigma_x.$$

Define the operational control energy as the norm of the Hamiltonian generating the pulse:

$$E_{\text{op}} := \|H_{\text{eff}}^{\text{pulse}}\| = \frac{\hbar\Omega_R}{2}.$$

Equivalently, if the measured Rabi frequency in cycles per second is

$$f_R = \frac{\Omega_R}{2\pi},$$

then

$$E_{\text{op}} = \frac{hf_R}{2}.$$

**Proposition 14.1** (Minimum record-changing operation time)**.** *A committed record-changing operation generated with operational energy $E_{\text{op}}$ cannot be completed faster than the Margolus–Levitin time*

$$\tau_{\text{op}} \geq \tau_{\text{ML}} = \frac{\pi\hbar}{2E_{\text{op}}}.$$

*For the ideal resonant square $\pi$-pulse,*

$$t_\pi = \frac{\pi}{\Omega_R} = \frac{1}{2f_R} = \frac{\pi\hbar}{2E_{\text{op}}}.$$

**Symbols and operations**

$\Omega_R$ is the Rabi angular frequency. $f_R = \Omega_R/(2\pi)$ is the measured Rabi frequency in cycles per second. $t_\pi$ is the duration of an ideal resonant $\pi$-pulse. $E_{\text{op}}$ is the norm of the effective Hamiltonian generating the operation. The Margolus–Levitin theorem is imported: a quantum system with energy $E$ above the relevant ground/reference level requires at least $\pi\hbar/(2E)$ to evolve to an orthogonal state.

**In plain English**

A pulse is not a mathematical dot. It is a physical operation. If the controller has finite operational energy, an operation that writes a distinguishable record or flips a two-level state takes nonzero time. The instantaneous-pulse limit is therefore an infinite-energy idealization.

**Physical justification**

The physical content is deliberately modest. O1 supplies finite operational energy; O2 says a committed record must be reliably distinguishable; Margolus–Levitin supplies the quantum-speed-limit time for reaching an orthogonal state. Therefore a controller cannot execute distinguishability-changing operations arbitrarily quickly at fixed $E_{\text{op}}$. This is a minimum operation time, not by itself a stochastic timing jitter. Treating it as a variance requires an additional noise model and is not part of the theorem-level result.

**Square pulses versus general controls**

The displayed equality $t_\pi = \pi/\Omega_R$ is the clean resonant two-level square-pulse case. For shaped pulses, multilevel systems, detuned drives, leakage-prone hardware, or explicitly time-dependent control Hamiltonians, $E_{\text{op}}$ should be replaced by the appropriate time-averaged or action-integrated quantum-speed-limit resource. The square-pulse expression is the minimal illustrative case used to expose the protocol-refinement ceiling.

## 15 The protocol-refinement ceiling

A protocol consisting of $n$ distinguishability-changing operations in a total duration $T$ must satisfy

$$n\,\tau_{\text{op}} \le T.$$

Using Proposition 14.1,

$$n\frac{\pi\hbar}{2E_{\text{op}}} \le T.$$

Thus the number of physically executable refinements is bounded by

$$\boxed{n \le n_{\max}(T, E_{\text{op}}) = \frac{2E_{\text{op}}T}{\pi\hbar}}.$$

In measured Rabi-frequency units this is

$$\boxed{n_{\max} = 2f_R T}.$$

**In plain English**

The point is not that CPMG or any other control protocol requires new physics. The point is that an infinitely refinable protocol assumes an infinitely capable external controller. Once the controller is treated as a finite physical system, the mathematical limit $n \to \infty$ acquires a quantitative cutoff.

**Physical justification**

The bound is independent of the environmental noise model. It follows before any discussion of dephasing, thermodynamics, or hardware error. Real hardware may have $t_\pi^{\mathrm{real}} > \tau_{\mathrm{ML}}$, in which case the experimentally available pulse count is

$$n_{\max}^{\mathrm{real}} = \frac{T}{t_\pi^{\mathrm{real}}} \le n_{\max}^{\mathrm{ML}}.$$

The Margolus–Levitin expression is the most optimistic quantum ceiling; actual devices usually sit below it.

## 16 Application: CPMG pulse sequences

A CPMG sequence applies $n$ nominal $\pi$-pulses over a sequence duration $T$. Since each pulse takes at least $t_\pi$, the total pulse time obeys

$$n t_\pi \le T.$$

For an ideal resonant pulse, $t_\pi = 1/(2f_R)$, so

$$n \le 2 f_R T = \frac{2E_{\mathrm{op}}T}{\pi\hbar}.$$

This is the direct laboratory form of the finite-refinement result.

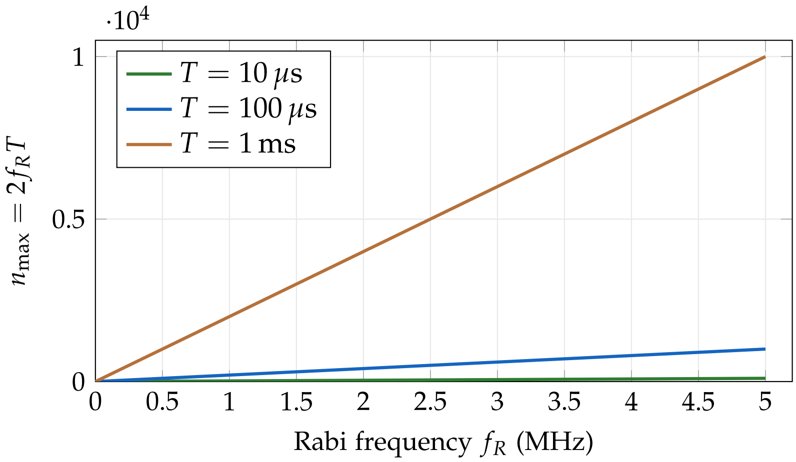


Figure 4: Finite-energy CPMG refinement ceiling. In experimental units the ideal resonant-pulse ceiling is $n_{\max} = 2f_R T$. The graph shows the upper envelope for three sequence durations. Real hardware may saturate below these lines.

**Symbols and operations**

$n_{\max}$ is the maximum number of ideal $\pi$-pulses that can fit into duration $T$ at measured Rabi frequency $f_R$. $n_{\mathrm{sat}}$ denotes the observed pulse number at which additional CPMG pulses cease to improve the measured coherence. The theorem-level statement is the upper envelope $n_{\mathrm{sat}} \le n_{\max}$, not a universal law for the value of $T_2$ at saturation.

**In plain English**

The robust experimental observable is where the improvement stops as pulses are added. The claim is not that the coherence time must obey one universal power law. The clean claim is that pulse-count saturation cannot exceed the finite-energy envelope $2f_R T$.

## 17 Model-dependent coherence consequences

The pulse-count ceiling becomes a coherence-time estimate only after adding a dephasing model. Assume power-law dephasing noise

$$S(\omega) = A|\omega|^{-\beta}, \qquad \beta > 0,$$

and the standard ideal-pulse CPMG asymptotic form

$$\chi(T, n) \sim K_\beta A \frac{T^{\beta+1}}{n^\beta},$$

where $K_\beta$ is a dimensionless filter-function constant. Setting $\chi(T_2, n) = 1$ gives

$$T_2(n) \sim \left(\frac{n^\beta}{K_\beta A}\right)^{1/(\beta+1)}, \qquad T_2(n) \propto n^{\beta/(\beta+1)}.$$

### Fixed-window consequence

Fix a target sequence duration $T_0$ and use a pulse count $n = \alpha n_{\max}(T_0)$ with $\alpha \ll 1$ so that the short-pulse condition remains safe. Then

$$n = \alpha \frac{2E_{\mathrm{op}} T_0}{\pi\hbar},$$

and therefore

$$T_2 \propto E_{\mathrm{op}}^{\beta/(\beta+1)}.$$

This is a fixed-window, ideal-pulse scaling estimate.

### Formal self-consistent envelope

If one formally sets the sequence duration equal to the achieved coherence time and uses the ceiling $n = T/\tau_{\mathrm{ML}}$, then

$$\chi_{\mathrm{env}}(T) \sim K_\beta A \frac{T^{\beta+1}}{(T/\tau_{\mathrm{ML}})^\beta} = K_\beta A \tau_{\mathrm{ML}}^\beta T.$$

Solving $\chi_{\mathrm{env}}(T_2^{\mathrm{env}}) = 1$ gives

$$T_2^{\mathrm{env}} = \frac{1}{K_\beta A}\left(\frac{2E_{\mathrm{op}}}{\pi\hbar}\right)^\beta.$$

Thus the formal envelope scales as

$$T_2^{\mathrm{env}} \propto E_{\mathrm{op}}^\beta.$$

**Epistemic status of the coherence scalings**

The pulse-count ceiling is theorem-level within the specified control model. The two coherence scalings above are not. They assume the ideal-pulse CPMG filter expression. The fixed-window scaling is valid only when $t_\pi \ll T/n$. The formal $E_{\mathrm{op}}^{\beta}$ envelope extrapolates that ideal-pulse expression to the boundary $n \sim n_{\max}$, where finite-pulse filter analysis is required. It is a guide for modeling, not a theorem-level prediction.

*Remark* (Existing $T_2$-versus-$n$ experiments). Existing CPMG measurements that report exponents such as $T_2 \propto n^{\gamma}$ usually characterize the environmental noise spectrum through the standard relation $\gamma = \beta/(\beta+1)$. Such results do not test the finite-energy refinement ceiling unless the drive energy is independently varied and the saturation pulse number is compared with $n_{\max} = 2f_R T$. In particular, an exponent near 0.72 can be explained by an appropriate noise exponent in ordinary filter-function theory and should not be treated as evidence for the observer-floor model.

## 18 Observer-realizable channels and finite refinement

**Definition 18.1** (Observer-realizable channel)**.** *A quantum channel $\mathcal{E}$ is observer-realizable at resources $(M, E_{\mathrm{op}})$ if it has a dilation*

$$\mathcal{E}(\rho) = \mathrm{Tr}_{\mathcal{O}}\left[U(\rho \otimes |0\rangle\langle 0|_{\mathcal{O}})U^{\dagger}\right]$$

*with ancilla dimension $\leq M$, bounded interaction generator $\|H_{\mathrm{int}}\| \leq E_{\mathrm{op}}$, and an operation time compatible with the relevant quantum-speed-limit resource.*

**In plain English**

Not every mathematically allowed quantum operation is one a finite observer can actually perform. An ORC is one that fits inside a real memory, a real interaction-strength budget, and a finite operation time. A pulse sequence is an iterated ORC, so finite $E_{\mathrm{op}}$ bounds not merely the fidelity of a single operation but the number of operations that can be packed into a finite interval.

**Physical justification**

The theorem-adjacent ORC consequences are deliberately limited. Finite memory bounds the ancilla dimension and hence the complexity of realizable dilations. Bounded interaction strength supplies a minimum time for operations that require distinguishable state change. Landauer reset costs remain relevant thermodynamic accounting when memory is erased. By contrast, exact bandwidth bounds, exponential infidelity floors, and detailed gate-quality scalings require additional control-model assumptions. They are not part of the theorem-level pulse-count ceiling.

## 19 Conditional extension: thermal timing floors

In addition to a minimum operation time, one may ask about stochastic timing precision for a finite clock in a thermal environment. This is a different question. It requires specifying a clock Hamiltonian, the thermalization basis, the bath model, and the measurement used to estimate time.

A useful scaling estimate is

$$\delta t_{\text{th}} \gtrsim \frac{c\hbar}{2\sqrt{k_B T E_{\text{op}}}}, \qquad c = O(1),$$

obtained by combining quantum Cramér–Rao timing precision with an energy-fluctuation estimate. For a specified thermalized two-level clock the dependence can instead take forms such as

$$\sigma_{\text{th}} = \frac{\hbar \cosh \eta}{2E_{\text{op}}}, \qquad \eta = \frac{E_{\text{op}}}{k_B T}.$$

**No universal crossover is claimed**

Thermal timing floors are conditional clock models, not theorem-level observer bounds. There is therefore no universal exact crossover coefficient and no single dimensionless number that classifies every real observer. In specified models, crossovers may occur at control energies of order $k_B T$, but the coefficient and even the detailed $\eta$-dependence are model-dependent. The main Part II result does not rely on these thermal estimates.

## 20 Experimental signature: saturation of refinement depth

The primary experimental observable is the saturation pulse number

$$n_{\text{sat}}(f_R, T),$$

not the precise value of $T_2$ at saturation. The finite-energy ceiling gives

$$n_{\text{sat}} \leq n_{\max} = 2f_R T.$$

**Physical justification**

Evidence consistent with the finite-energy ceiling would be saturation that approaches the envelope $2f_R T$ after ordinary hardware limits are controlled. Saturation far below the envelope would usually indicate bandwidth limits, leakage, heating, pulse errors, drive-induced noise, or $T_1$ relaxation, not a violation of the quantum speed limit.

A simple comparison is

$$\text{finite-energy envelope:} \qquad n_{\text{sat}} \lesssim 2f_R T,$$

versus a fixed-bandwidth hardware null such as

$$\text{bandwidth-limited null:} \qquad n_{\text{sat}} \sim T\Delta f_{\text{AWG}},$$

which is independent of $f_R$. More generally, platform-specific hardware mechanisms may also depend on drive amplitude, so the null hypothesis should include leakage, heating, amplifier compression, pulse-shape distortion, and drive-induced dephasing.

**In plain English**

The experiment does not try to falsify Margolus–Levitin itself. It asks whether a concrete dynamical-decoupling platform approaches the finite-energy refinement ceiling once ordinary hardware effects are accounted for. The clean signature is the location of the saturation point in pulse number as the measured Rabi frequency is varied.

**Unit-clean numerical scale**

If $f_R = 1\,\text{MHz}$, then

$$t_\pi = \frac{1}{2f_R} = 0.5\,\mu\text{s}.$$

For $T = 1\,\text{ms}$,

$$n_{\max} = 2f_R T = 2(10^6)(10^{-3}) = 2000.$$

If one imposes a conservative short-pulse margin $t_\pi \le 0.1\,T/n$, then the ideal-pulse CPMG expression should only be trusted up to roughly $n \lesssim 200$. The region near $n \sim n_{\max}$ is precisely where finite-pulse filter analysis is required.

# Part III — Geometry

Parts I–II concern a single observer. Geometry appears when *many* observers must share one world. The invariant structure across an entire family of observers is spacetime.

## 21 A0$^+$, the causal order, and no closed timelike curves

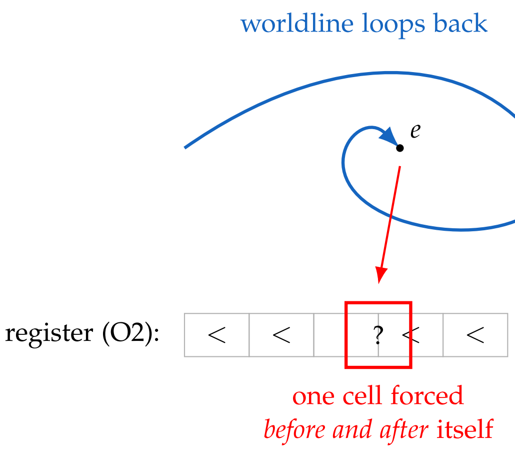


Figure 5: Why there are no closed timelike curves, without a geometric postulate. A worldline returning to its own past would route a record around the loop so that it arrives as an input to its own writing; the strictly ordered memory register (O2) would then have to hold one entry that is simultaneously earlier and later than itself, which it cannot. Acyclicity is a property of memory.

**Definition 21.1** (Admissible family and A0$^+$). *An admissible family $\mathcal{F}$ is a set of observers (each obeying O1–O4) any two of which are linked by a finite chain of physical interactions. A transformation is inert relative to $\mathcal{F}$ iff it changes no outcome for* any *member; physical reality is what is invariant under all such. $|\mathcal{F}| = 1$ recovers A0. For events, $e_1 \preceq e_2$ iff a signal at speed $\le c$ connects them (O3).*

**Proposition 21.2** (No CTCs). *$\preceq$ is a strict partial order: a closed causal chain would force some observer's memory (O2) to contain an entry both before and after itself.*

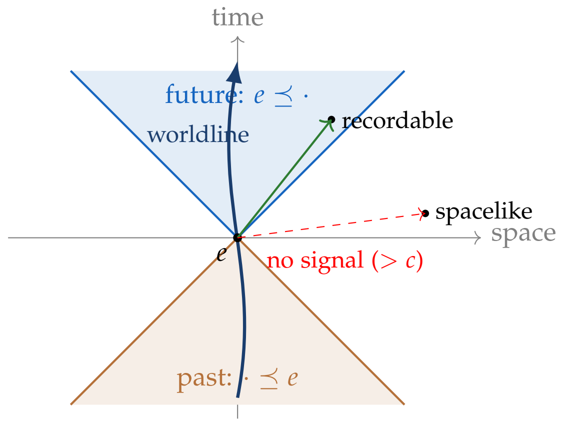


Figure 6: The causal order on the events of a family. A signal at speed $\leq c$ links $e$ to events in its future cone (which the observer can later record) and receives from its past cone; spacelike-separated events cannot be reached, so no information passes between them. This accessibility relation—finite-speed, mediated—is all O3 supplies. The light-cone structure, metric, and Lorentzian signature are reconstructed downstream (Sections 22–23), not assumed here.

**Symbols and operations**

A *partial order* $\preceq$ is a "before/after" relation that is transitive and never loops back on itself. A *closed timelike curve (CTC)* is a worldline that returns to its own past — time travel. $c$ is the finite signal speed (O3).

**In plain English**

Spacetime is what all the observers agree on after throwing away everything none of them can detect. The "before/after" relation between events is just "can a signal get from one to the other." Time loops are impossible for a simple reason: a memory register is strictly ordered (O2), so it can never hold a record that is simultaneously earlier and later than itself. The ban on time travel is a property of *memory*, not an extra axiom about spacetime.

**Physical justification**

Both halves are observer facts dragged into geometry rather than assumptions placed on it. The order itself is just O3 read as a relation: $e_1 \preceq e_2$ holds exactly when a signal at speed $\leq c$ can carry information from the first event to the second, so the "causal structure" of spacetime is nothing but the reachability graph of the no-signalling constraint. The acyclicity — no closed timelike curves — is forced by O2 rather than postulated: a closed causal chain would let a record propagate around the loop and arrive as an input to its own writing, so the register would have to hold one entry that is simultaneously earlier and later than itself, which an ordered finite memory cannot do. The standard chronology-protection assumption of classical GR is here a *theorem* about memory, and it costs no new geometric input.

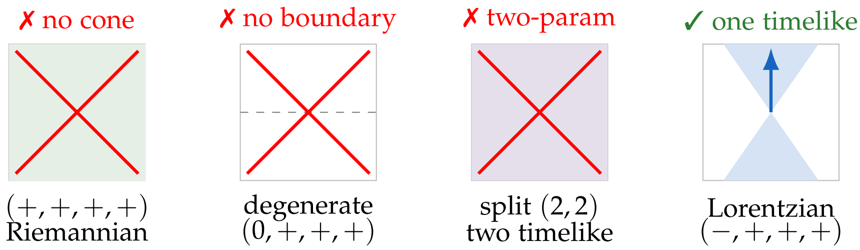


Figure 7: Signature exclusion. Riemannian has no light-cone; degenerate has no sharp causal boundary; split $(2, 2)$ admits a two-parameter family of timelike directions. Only Lorentzian $(-, +, +, +)$ gives exactly one timelike direction—which is forced by the unity of the observer (one system, one ordered register, one proper time).

## 22 Lorentzian signature is forced

**Proposition 22.1** (Lorentzian signature)**.** *Any smooth pseudo-Riemannian substrate consistent with (i) the causal partial order, (ii) finite signal speed, and (iii)* exactly one *continuous timelike direction per observer, must have Lorentzian signature* $(-, +, +, +)$.

**Symbols and operations**

*Signature* counts how many $-$ and $+$ appear when you diagonalize the metric: Riemannian $(+, +, +, +)$ is all-spatial; Lorentzian $(-, +, +, +)$ has one time direction; split $(-, -, +, +)$ has two. A *timelike* direction is one a massive observer can move along; the *light-cone* is the boundary between timelike and spacelike.

**In plain English**

What kind of geometry can host the causal structure? Run the options. All-plus (Riemannian): no light-cones at all, no notion of "faster than light," excluded. Degenerate: no sharp causal boundary, excluded. Two-minus (split): *two* independent time directions at each point, so an observer would have a whole plane of "forward" choices — but an observer has one ordered memory (O2) and one proper time (O4), hence one worldline direction. Only Lorentzian gives exactly one timelike direction. So spacetime has one dimension of time and the rest space — because the observer is one thing with one clock.

**Physical justification**

The "exactly one" clause is the unity of the observer: a single physical system (O1), single ordered register (O2), single proper time (O4). This proposition is also what *retroactively* supplies the global Lorentz invariance that the electromagnetism theorem (Section 12) had to cite as an input — closing that loop.

*Remark* (Time-orientability). "Exactly one continuous timelike direction per observer" is stronger than it looks. A smooth choice of one timelike direction at every point of $\mathcal{F}$'s region is a *global timelike line field*, whose existence is a genuine topological constraint on the manifold (a Lorentzian metric exists on a closed manifold iff its Euler characteristic vanishes, and a *time-orientable* one requires the line field to lift to a vector field). So O2+O4 deliver not merely "signature has one minus" but a time-orientable Lorentzian structure—which is what the arrow-of-time argument of Section 13 needs globally and what the no-CTC result protects. This is a topological selection rule, connected to a remaining topological question on discrete substrates.

## 23 The conformal class and the full metric

By Malament's theorem, any causal-order-preserving bijection between strongly causal Lorentzian manifolds is a smooth conformal isometry. So the causal order fixes the metric up to a local scale, $g_{\mu\nu} \mapsto \Omega^2(x) g_{\mu\nu}$. The observer's own proper-time normalization $g_{\mu\nu}\dot{\gamma}^\mu\dot{\gamma}^\nu = -c^2$ (O4) pins $\Omega$ along each worldline; smoothness extends it everywhere.

**Symbols and operations**

A *conformal* transformation $g \mapsto \Omega^2 g$ rescales lengths point-by-point while preserving all angles and light-cones; $\Omega(x)$ is the local scale factor. *Strong causality* is a mild "no almost-closed causal curves" condition. $\dot{\gamma}^\mu$ is the observer's four-velocity along its worldline $\gamma$.

**In plain English**

Knowing only who-can-signal-whom tells you the entire geometry *except* for an overall stretch factor at each point — the causal structure fixes all the angles and light-cones but not the local ruler. Each observer's own clock supplies the missing ruler: "my proper time ticks at rate $c$" sets the scale along its path, and smoothness spreads it to the whole manifold. Causal order + everyone's wristwatch = the full metric.

**Physical justification**

The geometry is recovered in two physically distinct pieces, and it matters that they come from different observer facts. Malament's theorem (imported) shows that the causal order alone — which events can signal which — already fixes the metric up to an overall scale at each point: the light-cones are determined, so angles and the conformal class are determined, but how much proper time elapses along a given timelike path is not. That remaining scale, the conformal factor, is then nailed down by O4: each observer carries a clock that reads its own proper time, and demanding the metric reproduce those readings fixes the one free function Malament leaves open. The strong-causality hypothesis Malament needs is itself supplied by the no-CTC result of the previous section — a chronology violation would appear as a closed causal curve in the limit. So the full metric is causal structure (from O3) plus proper-time calibration (from O4), with no separately postulated line element.

## 24 The local-flatness clause of the equivalence principle

**Scope of the result**

What is derived here is the local-flatness/Riemann-normal-coordinate clause: a freely falling observer can remove the local gravitational-force term as an inert coordinate artifact. The universality of free fall and the full locality of all non-gravitational interactions require the matter-coupling assumptions used elsewhere in the framework; they are not independently re-derived in this section.

**Proposition 24.1** (Equivalence principle from A0$^+$)**.** *At any worldline point $p$, no admissible protocol confined to a region of size $\ell \ll \mathcal{R}^{-1/2}$ (with $\mathcal{R}$ the local curvature scale) can distinguish curved spacetime from flat.*

**Symbols and operations**

*Riemann normal coordinates* are the coordinates a freely-falling observer naturally uses; at the chosen point they make the metric exactly flat ($g_{\mu\nu}(p) = \eta_{\mu\nu}$) with vanishing first derivatives. $\mathcal{R}$ is the curvature scale; $\ell$ the size of the lab. The *Riemann tensor* is the genuine, coordinate-independent curvature, first appearing at second order.

**In plain English**

At any point you can choose freely-falling coordinates in which spacetime looks flat to first order. Switching to those coordinates is a passive relabelling, so by A0$^+$ it is inert — it changes no measurement. Therefore no small-enough local experiment can feel the difference between curved spacetime and flat: that is the equivalence principle, and here it is a *theorem*, not a postulate. Real gravity — the part you cannot coordinate away — first shows up at second order, as tidal effects: the Riemann tensor.

**Physical justification**

The geometric import is the smooth manifold structure; Riemann normal coordinates always exist, the change to them is passive and inert by A0$^+$, and the metric departs from flat only at order $\ell^2$. The observer-first content is not the existence of Riemann normal coordinates by itself, but the claim that the first-order gravitational terms removed by such a coordinate choice have no invariant local physical content. Curvature and tidal terms remain because they cannot be transformed away.

**Scope of the equivalence-principle claim**

This section derives the local-flatness clause of the equivalence principle: first-order gravitational-force terms removable by a freely falling coordinate choice are inert descriptions, while curvature is physical. It does not independently derive every textbook clause sometimes bundled into the equivalence principle, such as universality of free fall for all matter models or the full locality of non-gravitational physics. Those enter through the matter-coupling and locality assumptions elsewhere in the framework.

## 25 Observer closure: the observer is on the right-hand side

**Definition 25.1** (Observer closure)**.** *Every member of $\mathcal{F}$ is a physical system described by the same ontology it inhabits: observers carry stress-energy (contribute to $T_{\mu\nu}$), consume entropy budget, hold only finite/local state, obey causal structure, contribute to geometry, and their proper-time generators lie* inside *the physical algebra.*

**In plain English**

The observer is not a camera floating outside the scene. It is made of matter, so it gravitates; it burns energy, so it adds entropy; it cannot hold a god's-eye global description, only a local causal patch; and its clock is a physical operator in the world, not a background dial. Six consequences follow, each load-bearing: no global viewpoint;

a personal causal horizon; thermodynamic irreversibility; finite-dimensional accessible algebras; the clock living in the algebra; and — decisively for gravity — the observer appearing in the source term $T_{\mu\nu}$.

**Physical justification**

The crucial one is Consequence 6: because observers carry stress-energy (O1), $T_{\mu\nu}$ *includes* them. Purely thermodynamic derivations of gravity (Jacobson) leave the origin of the source term unexplained; observer closure supplies it. The field equations are not equations about a world watched from outside — the watcher is in them.

## 26 Bekenstein, the Landauer–Clausius bridge, and Einstein's equations

### 26.1 The 1+1D bridge

For a uniformly accelerated observer in 1+1D, the Rindler horizon is the null ray $x = t$. By Bisognano–Wichmann the wedge's modular Hamiltonian is the boost $K = \frac{2\pi}{\hbar} \int dx\, x\, T_{tt}(x)$. Matter $T_{\mu\nu}$ crossing the horizon changes it by

$$\delta K = \frac{2\pi}{\hbar} \int T_{\mu\nu} k^\mu k^\nu \, d\lambda,$$

new information entering the patch. By Bekenstein, $\delta S = \delta K / T_U$, and the Landauer cost of registering then erasing those bits at $T_U$ is

$$\delta Q_{\text{Landauer}} = T_U \, \delta S = \delta K = \frac{2\pi}{\hbar} \int T_{\mu\nu} k^\mu k^\nu \, d\lambda = \delta Q_{\text{Clausius}}.$$

**Symbols and operations**

$k^\mu$ is the null direction along the horizon; $\lambda$ an affine parameter along it; $T_{\mu\nu} k^\mu k^\nu$ the energy flux crossing it. $T_U = \hbar a / (2\pi c k_B)$ is the *Unruh temperature* an accelerated observer feels. *Clausius* $\delta Q = T\, dS$ is the textbook heat–entropy relation. *Bisognano–Wichmann* (imported) identifies the wedge's modular flow with the Lorentz boost.

**In plain English**

An accelerating observer has a horizon; matter falling across it deposits information into the observer's patch. Storing that information has a thermodynamic price — Landauer's $k_B T \ln 2$ per bit, at the local Unruh temperature. The remarkable fact: that bookkeeping price equals, term for term, the heat flow in the ordinary Clausius relation. So Clausius is not assumed; it is the Landauer cost of an observer keeping records at a horizon. Feed that into Jacobson's 1995 argument (Clausius across every local horizon $\Rightarrow$ Einstein's equations) and the field equations follow.

**Physical justification**

The bridge needs $\delta S \leq M \ln 2$: the crossing information must fit in memory — and *that is exactly the semiclassical regime* where Einstein's equations hold. The observer constraint and the domain of validity are the same condition, which is a genuinely elegant point.

**Theorem 26.1** (Einstein's equations, conditional on imports)**.** *Under A0$^+$, O1–O4, the smooth manifold, the equivalence principle, the Bekenstein bound, the Unruh temperature, and Raychaudhuri,*

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

*with $\Lambda$ an undetermined integration constant.*

*Remark* (Scope of the Landauer–Clausius bridge). The bridge is *exact* in 1+1D, where the transverse integral is trivial. In 3+1D the term-by-term Landauer = Clausius identity is *not* re-derived; the lift is carried entirely by Jacobson's construction, which supplies the area/Raychaudhuri machinery the 1+1D toy lacks. What is reconstructed from Landauer is the *Clausius relation* (the equation of state), not the full Einstein equation; Jacobson then does the geometry. Stated precisely: the framework supplies the Clausius relation operationally in all dimensions and verifies the heat-flow identity exactly in 1+1D, while the dimensional lift and the area law are Jacobson's, imported.

### 26.2 The master ledger

Including the observer's memory in the crossed-product algebra, the change in field relative entropy splits into a geometric term and a Landauer term:

$$\boxed{\delta\langle K_\Omega^{\mathrm{QFT}}\rangle = \frac{\kappa}{2\pi}\frac{\delta A}{4G} + \ln 2 \cdot \delta\langle N_c\rangle.}$$

**Symbols and operations**

$K_\Omega^{\mathrm{QFT}} = -\ln\Delta_\Omega$ is the modular Hamiltonian; $\delta\langle K_\Omega\rangle$ the change in relative entropy of the field state (its quantum-information distinguishability from vacuum). $\kappa$ is the horizon surface gravity; $\delta A$ the area change; $N_c$ the number of classical bits the observer records/erases. Each memory bit at the Landauer threshold contributes $-\ln 2$ to the total modular Hamiltonian.

**In plain English**

The ledger is an exact accounting identity, not a new law: every bit of field information that becomes distinguishable to the observer is paid for in two currencies whose sum is fixed — geometric area (the Bekenstein increment) and Landauer erasure cost in the observer's memory. The left side is information; the right side is "area plus memory cost." It does *not* derive Einstein's equations (the area-law step imports them); it is the precise statement of where the information goes once you put the observer inside the algebra.

**Physical justification**

The memory term's coefficient $-\ln 2$ is itself derived: a two-level memory bit in the Gibbs state at $T_U$, *at the Landauer threshold* $\beta E_{\mathrm{bit}} = \ln 2$, contributes exactly $-\ln 2$ per bit. No new input beyond Landauer and the crossed product. Nothing more is claimed here: the ledger is an identity, and promoting it to an operator equation (toward Einstein–Langevin / stochastic gravity) is a future direction.

# Part IV — Algebraic Quantum Field Theory

The final domain. Put the observer of Parts I–III inside relativistic quantum field theory and ask which algebra of observables it can actually access. The answer — the modular crossed product, a type II factor — is where the framework meets the recent gravitational-algebra literature on its own ground.

## 27 Finite-access algebras and the clock in the algebra

A finite-energy observer can excite only modes with $\omega \leq 2\pi\Lambda_{\mathcal{O}}$ in region $\mathcal{O}$; the generated algebra $\mathfrak{A}_{(\Lambda_{\mathcal{O}},\mathcal{O})}$ is a finite-dimensional type I factor. As energy grows, the closure of their union recovers the full Rindler algebra.

**The cyclic separating vacuum is a structural input**

Part IV is conditional on the ambient QFT having the standard cyclic separating vacuum for the local algebras under discussion. This is not derived from A0 or O1–O4. It imports the Reeh–Schlieder/Tomita–Takesaki setting: a vacuum state whose entanglement across regions is strong enough that modular flow is available and faithful. The observer constraints then determine how a finite clock and finite access transform that ambient type III structure into the crossed-product/type II observer algebra. In short: the vacuum is the Part IV AQFT structural input; the crossed product is the observer-specific output. The Rindler, de Sitter, and BTZ claims are theorem-level conditional on this input; general curved-background extensions remain conjectural.

**Proposition 27.1** (Clock generator is in the algebra)**.** *The clock Hamiltonian $H_{\mathcal{O}}$ ($\|H_{\mathcal{O}}\| \leq E_{\mathrm{op}}$) is an element of the physical algebra.*

**Symbols and operations**

A *von Neumann algebra* is a collection of observables closed under the operations physics needs (adjoints, limits). A *factor* is one with trivial center (no nontrivial observable commutes with everything). *Type I* factors are the familiar matrix/operator algebras with well-defined density matrices and entropies; *type III* (the generic local QFT algebra) has *no* trace and no density matrices — entropies are infinite. *Type II* sits between: it has a trace, hence finite *relative* entropies.

**In plain English**

A real observer touches only finitely many field modes (finite energy), so the algebra it can actually manipulate is an ordinary finite matrix algebra — type I — not the pathological type III algebra of the full field. And the observer's clock is not a background parameter painted on from outside: by A0, if the clock operator were *not* among the physical observables, two situations differing only in their time records would be assigned identical statistics, which O2 forbids. So the clock is *in* the algebra.

**Physical justification**

The physical claim is that the abstract type classification of the algebra is decided by observer finiteness, not by the field theory in the abstract. A real observer has finite energy (O1), so it can excite and read out only finitely many field modes; the algebra it can actually manipulate is generated by finitely many bounded operators and is therefore type I, the ordinary case where density matrices, entropies, and traces all exist. The infamous type $\mathrm{III}_1$ algebra of local QFT — on which entropy is not even definable, because every state has infinite entanglement across the boundary — is precisely the idealized object you get by taking the observer constraints to their unphysical limit (unbounded energy, perfect localization). On this reading type III is not the "true" algebra that the observer approximates; it is the operationally *incomplete* one, missing exactly the finiteness that makes physical quantities well-defined. The second claim, that the clock generator $H_{\mathcal{O}}$ belongs to the algebra rather than sitting outside as a c-number parameter, is forced by A0 together with O2: if the clock were external, two histories differing only in their recorded times would carry identical observable statistics, erasing a distinction O2 insists is real.

## 28 Internal symmetry from finite observable accessibility

The finite observable algebra just constructed has more to say before we pass to the modular completion. A finite observer resolves the operators it can measure, but *not* the internal multiplicities those operators commute with — and the unitary freedom on those unresolved multiplicities is exactly what gauge symmetry is. This section makes that precise. The commutant construction at its center is a clean theorem about finite observer algebras; the material that tries to push it toward the Standard Model is honestly conjectural, and the seam between the two is kept visible throughout.

**In plain English**

The deepest reframing here: *gauge symmetry is not fundamental ontology — it is the residual unitary freedom left undetermined by finite observer accessibility.* A finite observer's measurements pin down the observable algebra but say nothing about rotations *inside* the multiplicity spaces that commute with everything measurable. Those rotations are physically inert by A0, and the group of them is the gauge group.

### 28.1 The finite observable algebra, restated

From Section 27, within a single causal patch the operators an observer can actually measure form a finite-dimensional C*-algebra. Every such algebra is, up to isomorphism, a direct sum of matrix blocks (Artin–Wedderburn):

$$\mathfrak{A}_{\text{patch}} \;\cong\; \bigoplus_{i=1}^{k} M_{d_i}(\mathbb{C}),$$

carried on the observer's internal space $\mathcal{H}_{\mathcal{O}} \cong \mathbb{C}^n$ by a representation $\pi$ with $\sum_i m_i d_i = n$, where $m_i$ is the multiplicity of block $i$.

**Symbols and operations**

$M_d(\mathbb{C})$ is the algebra of $d \times d$ complex matrices. A *block* $M_{d_i}$ is one irreducible kind of measurable operator; $d_i$ is how many distinguishable states that kind resolves. The *multiplicity* $m_i$ counts how many identical copies of block $i$ sit inside $\mathcal{H}_{\mathcal{O}}$ — copies the observable operators cannot tell apart, because every measurable operator acts the same way on all of them.

**Physical justification**

Finiteness of the record count caps the number of reliably distinguishable orthogonal states, so the measurable algebra is finite-dimensional; Artin–Wedderburn then forces the block form with no further input. This is operational and patch-local — no Standard Model structure has entered, and none should yet.

### 28.2 Inert internal rotations and the commutant

A unitary $U \in U(n)$ that commutes with every measurable operator changes *no* measurement outcome, so by A0 it is physically inert. The set of such unitaries is the gauge group:

$$G = \{ U \in U(n) \ : \ U\,\pi(a)\,U^\dagger = \pi(a) \ \ \forall a \in \mathfrak{A}_{\text{patch}} \}.$$

**Symbols and operations**

The *commutant* $\mathfrak{A}'$ of an algebra is everything that commutes with all of it. $G$ is the unitary group *of* the commutant. $U(m)$, $SU(m)$ are the unitary and special-unitary groups; $U(1)$ is the circle of phases.

**Theorem 28.1** (Gauge group from the commutant). *With $\mathfrak{A}_{\text{patch}} \cong \bigoplus_i M_{d_i}(\mathbb{C})$ represented on $\mathbb{C}^n$ with multiplicities $m_i$, the commutant is $\mathfrak{A}'_{\text{patch}} \cong \bigoplus_i M_{m_i}(\mathbb{C})$, and the gauge group is its unitary group,*

$$G \cong U(m_1) \times U(m_2) \times \cdots \times U(m_k).$$

*Removing the inert overall phase of each block, the group acting on physical states is*

$$G_{\text{eff}} \cong \big(SU(m_1) \times \cdots \times SU(m_k) \times U(1)^r\big)/\Gamma,$$

*a compact product of simple non-Abelian factors and Abelian phases, with $\Gamma$ a discrete center and $r$ the number of surviving Abelian factors.*

**In plain English**

The structure the observer *cannot* resolve has its own shape, and that shape is a product of unitary groups — one factor for each kind of internal multiplicity. After throwing away the global phases A0 already calls inert, what is left is a compact group built from $SU(m_i)$ pieces and $U(1)$ pieces. This is the standard form of a gauge group, and it has appeared without importing any field theory: it is forced by the block structure of a finite observable algebra.

**Physical justification**

This is the double-commutant theorem applied to a finite direct sum of matrix algebras — mathematically standard and fully rigorous. Its *interpretation* is the load-bearing OFP content: observable algebra = accessible structure; commutant = operationally invisible redundancy; gauge symmetry = invariance under inaccessible internal rotations. This is the first point in the chain where the gauge sector becomes genuinely native to the framework rather than imported, and it fits the observer-finite logic exactly: gauge freedom is the freedom finite accessibility leaves undetermined.

*Remark* (Scope). The theorem fixes the *form* of the gauge group (compact product of $SU(m_i)$ and $U(1)$ factors) from the block structure of whatever finite algebra the observer happens to carry. It does *not* select specific integers $m_i$ — those are inputs read off the observer's algebra, not outputs derived from O1–O4. In this it agrees with the earlier Doplicher–Roberts analysis (which gave compactness from finiteness but provably could not select a particular group), while needing none of the Haag–Kastler / vacuum-sector / split-property machinery that route assumed. The compact-product conclusion rests on elementary finite-dimensional algebra rather than on AQFT structure not yet available at that stage of the argument.

### 28.3 Causal chains and holonomy consistency

The commutant lives inside one patch; the chain machinery links patches. On overlaps of causal patches the transition maps form a (generally non-Abelian) 1-cocycle — the defining data of a principal bundle over the poset of patches — and the holonomy of that bundle around a causal loop is the gauge-field content the chain can carry.

**In plain English**

Chains only constrain *relationally observable* transitions between patches. They do not fix the internal redundancy spaces that commute with all admissible operations. So when two patches overlap, the freedom in gluing their internal multiplicity sectors is exactly the commutant symmetry of Theorem 28.1, and tracking it around a loop gives a holonomy. Gauge structure is the holonomy ambiguity of the causal chain, classified by the commutant.

**Physical justification**

Causal-chain composition induces overlap cocycles; cocycle closure induces the bundle structure; the holonomy ambiguity is classified by the commutant symmetry sectors. This ties the gauge group into the chain construction rather than bolting it on: the same finite-accessibility that produced the patch algebra produces, on overlaps, the structure group of a bundle whose holonomies are the field strengths. What the chain does *not* supply is a reason for any particular holonomy group; that, again, is read from the algebra, not derived.

### 28.4 Thermodynamic accessibility and bounded gauge complexity

A finite observer can resolve only boundedly many holonomy variables, which suggests a ceiling on gauge complexity tied to the observer's information capacity.

*Remark* (A conjectural bound, not a theorem). If an observer can store at most $S_{\max}$ bits, then tracking all independent gauge degrees of freedom plausibly requires

$$\dim G \;\leq\; S_{\max} \ln 2. \qquad (\star)$$

This is *not* derived. It is an operational *hypothesis*, included because it is structurally OFP-consistent (it links gauge distinguishability, holonomy resolution, and observer entropy capacity — exactly the triad the framework couples elsewhere) and because it is *falsifiable*: for a fundamental Planck-scale observer the Bekenstein bound caps $S_{\max}$, giving $\dim G$ of order $10^2$–$10^3$; the observed $\dim G = 8 + 3 + 1 = 12$ sits comfortably inside. A genuinely fundamental gauge group exceeding the Planck-scale ceiling would, under $(\star)$, have to be an effective rather than fundamental description. The bound points in the right structural direction; it has no rigorous derivation, and its status is .

*Remark* (A research direction, not a result). One would like a principle that selects *which* block structure nature realizes — e.g. "the observable algebra of maximal entropy compatible with causal composition." At present this is underdefined: the entropy functional, the poset measure, the causal weighting, the admissible-covering class, and the uniqueness conditions are all unspecified. It is a well-posed *optimization programme* (tractable, perhaps, by discrete Morse theory on the causal poset), not a derivation. Included to mark the direction, not to claim the result.

### 28.5 The Standard Model, very cautiously

It is natural to ask whether the observed $SU(3) \times SU(2) \times U(1)$ drops out. The honest answer is that the framework privileges compact products of simple unitary factors and Abelian centers — and the Standard Model is *one* such pattern — but it does not derive the specific group.

**Physical justification**

The safe and still-substantial statement is: *finite embedded observers naturally induce compact internal symmetry groups arising as commutants of finite observable algebras*, and the observed Standard Model structure is one particularly efficient saturation pattern of the resulting algebraic constraints. The framework sharply constrains the admissible *form* of internal symmetry; it does not derive the specific gauge group, which remains an empirical input.

## 29 The physical algebra is the fixed-point algebra

**Lemma 29.1** (Clock-shift inertness). *The proper-time shift $\tau \mapsto \tau + \alpha$, implemented by $\sigma_\alpha^{\text{clock}}(a) = e^{i\alpha H_\mathcal{O}}\, a\, e^{-i\alpha H_\mathcal{O}}$, is physically inert under A0.*

**Proposition 29.2** (Fixed-point characterization). $\mathfrak{A}_{\text{phys}} = \{a : \sigma_\alpha^{\text{clock}}(a) = a\ \forall \alpha\} = \{a : [H_\mathcal{O}, a] = 0\}$.

**Symbols and operations**

An *automorphism* $\sigma_\alpha$ is a symmetry of the algebra (here, shifting the clock's zero). A *fixed-point algebra / centralizer* is the set of observables left unchanged by it — equivalently, those commuting with $H_\mathcal{O}$. $[H_\mathcal{O}, a] = H_\mathcal{O} a - a H_\mathcal{O}$ is the commutator.

**In plain English**

Choosing when to start your clock is a passive convention — inert by A0, just like a passive coordinate change. So genuine physical observables must not care about that choice: they are exactly the operators that commute with the clock generator. The physical algebra is the "clock-shift-invariant" part of everything.

### 29.1 Finite clock coarse-graining

A finite observer's clock is not an ideal external spectral reference. In any concrete model, finite control energy, finite record resolution, thermal coupling, or instrumental coarse-graining supplies an energy width $\Delta$ for the realizable clock generator. One may represent the accessible generator schematically as

$$\mathbb{H}_\Delta = \int \chi_\Delta(\lambda - H_\mathcal{O})\, \lambda\, d\lambda,$$

with the physical algebra taken relative to the corresponding coarse-grained clock action.

**In plain English**

No real clock is infinitely sharp. The exact "commute with $H_{\mathcal{O}}$" condition is therefore an ideal limit; finite observers work with a physically resolved clock. The source of the width $\Delta$ is model-dependent—thermal, instrumental, or control-theoretic—and is not identified here with a universal Part II timing floor. What matters for the algebraic construction is that the clock is inside the algebra and has finite operational resolution.

**Scope of the type-II claim**

In what follows, theorem-level type-II claims refer only to horizon settings where the needed modular identification is known: Rindler wedges, de Sitter static patches, and BTZ wedges. The general curved-background statement is a conjectural extension of the same observer-clock mechanism.

## 30 The crossed product and type II structure

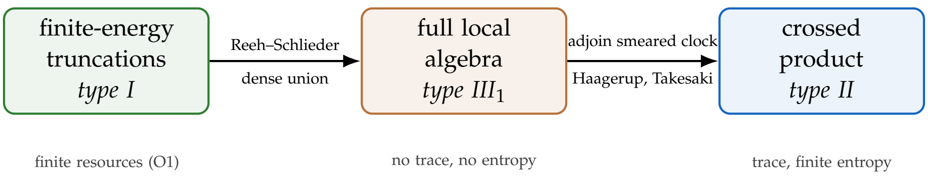


Figure 8: The algebra flow. What a finite observer can access is a type-I matrix algebra; the union of these truncations is dense (Reeh–Schlieder) in the full type-$\text{III}_1$ local algebra, which carries no trace and so no well-defined entropy. Adjoining the observer's smeared clock as a dynamical variable—the crossed product—produces a type-II factor on which a trace, and hence finite vacuum-subtracted entropy, exists. The observer is what regularizes the algebra.

**Theorem 30.1** (OFP local algebra is type II: Rindler and de Sitter)**.** *Under A0$^+$ + O1–O4 + the cyclic separating vacuum, for Rindler wedges in flat space and static patches in de Sitter,*

$$\mathfrak{A}_{\text{OFP}}(\mathcal{O}) \cong \mathfrak{A}(\mathcal{O}) \rtimes_{\sigma^{\text{vac}}} \mathbb{R}$$

*is a hyperfinite type* $\text{II}_\infty$ *factor* ($\text{II}_1$ *in compact regions).*

**Symbols and operations**

A *cyclic separating vacuum* $|\Omega\rangle$ is the standard QFT vacuum (it can generate the whole space and is not annihilated by any algebra element) — the one nontrivial physical import of Part IV. The *crossed product* $\mathfrak{A} \rtimes_\sigma \mathbb{R}$ adjoins the flow $\sigma$ (the clock) as new degrees of freedom. *Hyperfinite* means approximable by finite-dimensional pieces. *Reeh–Schlieder*, *Haagerup*, *Takesaki* are the imported theorems doing the heavy algebra.

**In plain English**

Here is the punchline of Part IV. The full field algebra on a region is type III: it has *no* well-defined entropy, because the vacuum is infinitely entangled across the boundary.

But the observer cannot resolve that infinite entanglement — it has finite energy and a blurred clock. Quotienting by the inert clock-shift, equivalently adjoining the clock as a dynamical variable (the crossed product), *regularizes* the type III algebra into a type II one. Type II has a trace — so entropies, finite at last, exist. The infinity that plagued the field algebra was an artifact of pretending the observer was not there.

**Physical justification**

Two clean steps: (A) finite-energy truncations are type I, and by Reeh–Schlieder their union is dense in the type $\mathrm{III}_1$ Rindler algebra; (B) by Haagerup every type $\mathrm{III}_1$ factor has a type $\mathrm{II}_\infty$ core, reached as the crossed product by modular flow, and Takesaki duality identifies the fixed-point algebra with that crossed product. For Rindler/de Sitter/BTZ, Bisognano–Wichmann identifies the clock-shift with the vacuum modular flow, and the result is a *theorem*. For general curved backgrounds that identification is Conjecture 23.2, reducing to a spectral-convergence statement (Open Problem 6).

*Remark* (Scope of the type-II theorem). Theorem 30.1 is proved *only* for Rindler wedges, de Sitter static patches, and BTZ—which are the physically central cases (every point has a local Rindler horizon by the equivalence principle; de Sitter is the cosmological horizon; BTZ is the stationary black hole). For *general* curved backgrounds it is a conjecture awaiting a proof that the smeared-clock centralizer converges to the modular crossed product. Because the Bekenstein-as-theorem, Unruh-as-KMS, and generalized-entropy results all ride on Theorem 30.1, each inherits the same theorem/conjecture split, and each is tagged inline as *theorem for Rindler/de Sitter, conjectural for general backgrounds* rather than carrying the qualifier silently.

## 31 Microcausality, Bekenstein, Unruh, generalized entropy

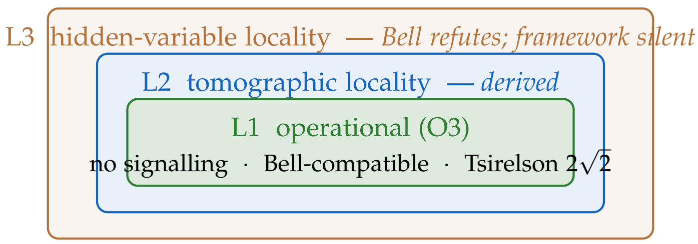


Figure 9: Three senses of locality, nested. The framework lives in the inner two layers: operational locality (O3, no signalling) and tomographic locality (derived, Theorem 9.1). Bell's theorem refutes only the outer layer—hidden-variable locality—about which the framework says nothing. A Bell violation lies between L2 and L3: it breaks the classical-realist picture without breaching no-signalling, so it threatens neither O3 nor the reconstruction.

These four results are short once Theorem 30.1 is in hand. Each is stated with its honest status.

**Microcausality.** Observers on disjoint worldlines cannot signal (O3), so their clock generators commute, $[H_{\mathcal{O}_1}, H_{\mathcal{O}_2}] = 0$; with standard field commutation this gives $[\mathfrak{A}_{\mathrm{OFP}}(\mathcal{O}_1), \mathfrak{A}_{\mathrm{OFP}}(\mathcal{O}_2)] = 0$ for spacelike-separated regions.

**In plain English**

Two observers who cannot talk have independent clocks and independent observables — so everything in their two regions commutes. Spacelike separation means no interference, which is microcausality.

**Bekenstein bound as a theorem (Rindler/de Sitter).** On the type II algebra the vacuum-subtracted entropy $\Delta S(\omega) = S(\omega) - S(\omega_{\text{vac}})$ is finite, and by the KLRSS regulator $\Delta S(\omega) \leq \frac{2\pi R}{\hbar c}\langle E \rangle_\omega$.

**In plain English**

What was an *imported* empirical bound in Part III becomes a *theorem* here: because the type II algebra has a trace, the entropy relative to vacuum is finite, and positivity of relative entropy gives the Bekenstein bound for free. The bound is no longer an extra input — it is a consequence of the same vacuum that standard QFT already assumes.

**Unruh as KMS .** For an accelerated observer the clock is the boost (geometry + Bisognano–Wichmann), the modular flow is the boost (imported), so the vacuum is a KMS (thermal) state at $T_U = \hbar a/(2\pi c k_B)$.

**In plain English**

The framework does not *derive* the Unruh effect — it shows the pieces fit: an accelerated observer's clock *is* the modular flow, and a vacuum is automatically "thermal" with respect to its own modular flow (the KMS condition). The Unruh temperature is that thermality. The matching Landauer dissipation rate $P_{\mathcal{O}} \geq k_B T_U \ln 2 \cdot M/\tau_{\mathcal{O}}$ is the same identification that powered the 1+1D bridge.

**Generalized entropy .** For semiclassical states the von Neumann entropy of the OFP algebra reproduces $S(\omega) = A/(4\ell_P^2) + S_{\text{QFT}}(\omega) + \text{const}$ (CLPW applied to the type II algebra).

**In plain English**

The famous "generalized entropy" — horizon area in Planck units plus the field entropy outside — comes out as the natural entropy on the observer's algebra. The clock degrees of freedom that the gravity literature calls "the observer" are exactly $H_{\mathcal{O}}$ from O4 plus Part II. This is where OFP and the Witten–CLPW program turn out to be describing the same object.

## 32 Independent convergence with the gravitational program

**In plain English**

The strongest piece of corroboration the framework can offer short of a lab result: two completely different starting points land on the same algebra. Witten (2022) and Chandrasekaran–Longo–Penington–Witten reached the type II crossed product from $\mathcal{O}(1/N)$ gravity in large-$N$ holography — they put *gravity* in. OFP reaches the identical structure from six observer constraints — it puts *the observer* in. The two routes

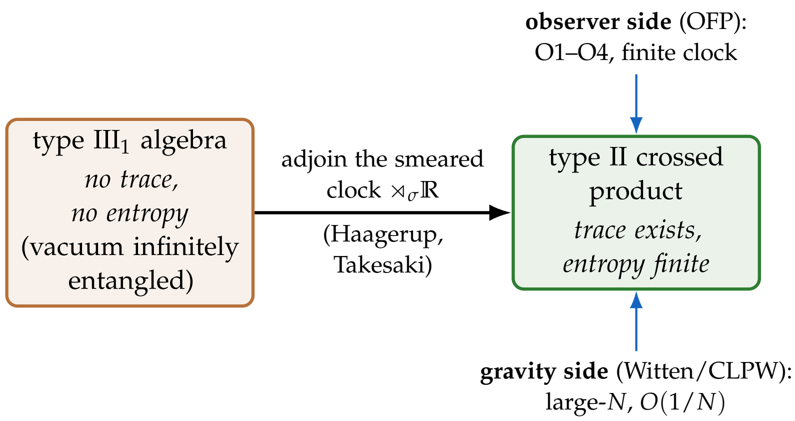


Figure 10: The type-III→type-II regularization, and the convergence. Adjoining the observer's smeared clock as a dynamical variable (the crossed product) turns the trace-free type-III vacuum algebra into a type-II factor on which entropy is finite. Two unrelated routes—OFP from observer constraints, the Witten/CLPW program from large-$N$ gravity—arrive at the same object, each having put in a different ingredient.

start from orthogonal inputs. When two unrelated derivations converge on the same mathematical object, the natural reading is that both are detecting one real constraint: that no finite observer can resolve the infinite entanglement of a type III vacuum, so the operationally honest algebra is always type II.

**Physical justification**

This is convergence, not proof. It does not make Conjecture 23.2 a theorem. But independent arrival at the same structure from gravity (input: large-$N$) and from operational constraints (input: O1–O4) is a non-trivial consistency check, and should be read as structural corroboration rather than confirmation.

**General-background program**

The next technical problem is to show whether smeared-clock centralizers converge to the corresponding modular crossed product under mild curvature and energy conditions. A positive result would extend the construction beyond symmetric horizons; a counterexample would precisely delimit the algebraic scope of the framework.

# Part V — Extension: the form of relativistic matter

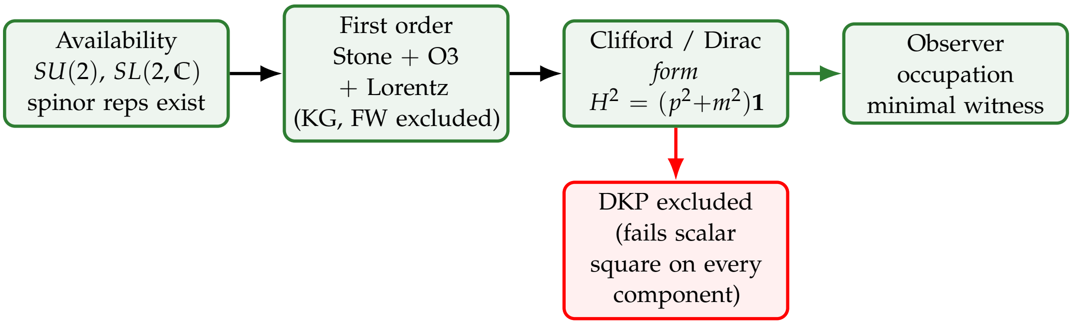


Figure 11: The matter chain. Availability (group theory alone) does not force occupation; the dynamical argument (Stone + locality + Lorentz) selects the first-order form, and the record-faithful scalar-square condition selects Dirac while excluding auxiliary DKP sectors. The observer's own existence as matter supplies the minimal occupied witness of that form through its fundamental spin-$\frac{1}{2}$ constituents (empirically, electrons among them); multiplicities, masses, couplings, and the Standard Model census remain empirical.

This part asks what *form* record-faithful local relativistic matter must take for a finite embedded observer, and identifies the spinorial form as the survivor under the stated conditions. The observer supplies the minimal occupation of that form through the fundamental spin-$\frac{1}{2}$ constituents of ordinary matter (empirically, electrons among them); the detailed particle census remains outside the derivation, in the same way that the gauge-group form is constrained without fixing the specific Standard Model integers. No new axiom is introduced; the central result (Section 34) follows from structure already in hand. Two limitations are noted where they arise: the argument is first-quantized, and the locality premise must remain clearly separated from the Lorentzian geometry derived earlier.

## 33 Availability: the covering group and spinor representations

The Lorentz group acts on the state space by inert transformations (A0), so its action is a symmetry in the sense of Part I. The first question is which representations are *available*.

**Proposition 33.1** (Spinor representations are available)**.** *The proper orthochronous Lorentz group acts on $\mathcal{H}_O$ by a projective unitary representation (Wigner); its connected component cannot carry antiunitaries without reversing the sign of energy (O1), so the representation is unitary up to phase. The phase lifts (Bargmann; $H^2(SO^+(3,1);U(1)) = 0$) to a true unitary representation of the universal cover $SL(2,\mathbb{C})$. The half-integer representations $(\frac{1}{2},0)$ and $(0,\frac{1}{2})$ — the Weyl spinors — are* *therefore available on equal footing with the integer ones. The same argument restricted to spatial rotations gives $SO(3) \to SU(2)$ and is independent of all relativistic structure.*

**Symbols and operations**

The *universal cover* of a group is the simply-connected group mapping onto it; $SL(2, \mathbb{C})$ covers the Lorentz group two-to-one, $SU(2)$ covers $SO(3)$ two-to-one. A *projective* representation is defined only up to a phase $\omega(\cdot, \cdot)$; when the phase can be removed it is a *true* representation. $(\frac{1}{2}, 0)$ labels the left-handed Weyl spinor, $(0, \frac{1}{2})$ the right-handed one.

**In plain English**

States are rays, so symmetries act only up to a phase. For the Lorentz group that phase can always be absorbed by passing to its double cover $SL(2, \mathbb{C})$ — and once you are on the cover, half-integer-spin objects (spinors) are perfectly legitimate representation labels, sitting beside the ordinary integer-spin (tensor) ones. The cheap, robust version uses only rotations: $SO(3)$'s double cover is $SU(2)$, and spin-$\frac{1}{2}$ is born there with no relativity at all.

**Physical justification**

Wigner and Bargmann are imported; the OFP content is merely that Lorentz acts inertly (A0) with positive energy (O1), which selects the unitary branch on the cover. The honest weight of this proposition is small: it shows spinors *can* live in the theory, not that anything *does*. Every purely bosonic theory satisfies it verbatim, because every unitary representation of $SO^+(3,1)$ lifts to $SL(2, \mathbb{C})$, including those that factor back through $SO^+(3,1)$ (integer spin). Availability is not occupation; conflating them is the single most common error in informal "OFP implies spinors" arguments, and Section 36 keeps the two apart.

*Remark* (Availability is not occupation). Proposition 33.1 is an availability statement: it identifies the representations that can carry relativistic local matter, but does not by itself populate them. The force comes from the dynamical argument of the next two sections together with the observer-occupation argument of Section 36. This is the matter-sector twin of Part IV's discipline that the commutant constrains the gauge group's *form* but not its specific integers.

**Two roles of locality**

O3 supplies operational locality: no physical update may require unmediated access to spacelike-separated data. Part III supplies Lorentzian covariance: the geometric form taken by that accessibility relation in the continuum limit. The first-order matter argument uses both, but not circularly: O3 excludes nonlocal generators such as the square-root Hamiltonian, while Lorentzian geometry fixes the relativistic covariance condition.

## 34 The fundamental matter law is first order

This is the one genuinely new result of Part V, and it uses no new axiom: it assembles Stone (Section 10), locality (O3), and Lorentzian structure (Section 22), all already in hand.

**Theorem 34.1** (First-order forcing)**.** *Let a localized matter excitation evolve in an observer's frame. Given (i) Stone evolution $i\,\partial_\tau\psi = H\psi$ with self-adjoint $H$ (Section 10); (ii) locality of the interaction — $H$ a local operator of finite order in spatial derivatives (O3); and (iii) Lorentz covariance (Section 22); the field equation is first order in* all *spacetime derivatives. In particular the second-order Klein–Gordon operator cannot be the fundamental generator; it can appear only as the square of the first-order law.*

**Symbols and operations**

$\partial_\tau$ is the derivative along the observer's proper time. "First order in time" means the equation fixes $\partial_\tau\psi$ from $\psi$ alone, with no $\partial_\tau^2$ term — exactly the Schrödinger/Stone form. A *local* operator is a finite-order differential operator (no integral kernel reaching distant points); the nonlocal alternative is a pseudodifferential operator such as $\sqrt{p^2+m^2}$.

**In plain English**

Part I already proved the observer's evolution is first order in its own time (Stone: $i\,\partial_\tau\psi = H\psi$). So "first-order dynamics" is not something to assume for matter — it is already a theorem about how any observer's state moves. There is exactly one way to wriggle out: keep first-order-in-time but make the energy operator *nonlocal*, $H = \sqrt{p^2+m^2}$ (the Foldy–Wouthuysen Hamiltonian). Locality (O3) forbids that — it reaches across space instantaneously. So the matter law is local *and* first order in time; and because Lorentz covariance puts time and space on the same footing (a boost mixes $\partial_\tau$ with the $\partial_i$, so the highest derivative order is the same in every frame), it must be first order in space too. Klein–Gordon, being second order, is disqualified as the fundamental law — it survives only as the squared, on-shell consistency condition.

**Physical justification**

The three inputs are all prior results: Stone is the unitary-dynamics theorem of Section 10; locality is O3, the same no-signalling constraint that builds the causal order in Section 21; Lorentz covariance is the signature result of Section 22. The genuinely informative move is the explicit exclusion of the *nonlocal* scalar option. "First order in time + positive energy + relativistic dispersion" has three solutions, not one: the nonlocal scalar $\sqrt{p^2+m^2}$ (excluded by O3); the second-order scalar Klein–Gordon (not Stone-form, so not a fundamental generator); and the local first-order multi-component equation, which is what remains. Naming all three and killing two is what turns "first order is natural" into "first order is forced."

*Remark* (First order is a consequence, not a postulate). A tempting shortcut (seen in informal treatments) promotes "the observer's update is Markovian and first order" to a new postulate, sometimes labeled "O5." Reject this on two grounds. First, it collides with the framework's axiom budget — the foundation is A0 + O1–O4 (Section 3), and an inserted "O5" is redundant because first-order-in-time is *already* the content of Stone (Section 10). Second and worse, it is circular: it postulates the very operational fact whose grounding is the point. The honest structure is the reverse — first-order is *derived*, and the only genuinely additional ingredient is locality (O3), which is already an axiom. No new postulate is needed or wanted.

## 35 First order plus dispersion selects the Clifford algebra

**Theorem 35.1** (Clifford forcing and the Dirac equation)**.** *Write the first-order local generator of Theorem 34.1 in Hamiltonian form $i\partial_0\psi = (\alpha^i p_i + \beta m)\psi$. For a record-faithful irreducible matter field, requiring the squared generator to be the relativistic dispersion times the identity, $H^2 = (p^2 + m^2)\,\mathbf{1}$, is equivalent to*

$$\{\alpha^i, \alpha^j\} = 2\delta^{ij}\mathbf{1}, \qquad \{\alpha^i, \beta\} = 0, \qquad \beta^2 = \mathbf{1},$$

*a Clifford algebra. Its minimal faithful representation is four-dimensional over $\mathbb{C}$; the resulting equation is the Dirac equation, whose solution space carries the $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation of $SL(2, \mathbb{C})$. Record-faithful local first-order matter is therefore spinorial in form.*

**Symbols and operations**

$\alpha^i, \beta$ are the matrices in Dirac's Hamiltonian form; $\{A, B\} = AB + BA$ is the anticommutator. $\mathbf{1}$ is the identity on the internal (spinor) space. A *Clifford algebra* $Cl(1,3)$ is generated by elements squaring to $\pm\mathbf{1}$ and anticommuting in pairs; its irreducible representation here is the four-component Dirac spinor. *Record-faithful irreducibility* means that the components of a fundamental observer-occupying field represent genuine propagating distinguishable degrees of freedom, not merely auxiliary algebraic constraint variables.

**In plain English**

Take the local, first-order law and write it as Dirac originally did, an energy operator linear in momentum: $H = \alpha\cdot p + \beta m$. For components that are genuine record-bearing propagating degrees of freedom, demand that squaring it return the correct relativistic energy $p^2 + m^2$ on *every* component. Multiplying out, the cross terms must cancel and each matrix must square to one — which is exactly the Clifford anticommutation relations. The smallest matrices that can do this are $4 \times 4$, and the object they act on is the four-component Dirac spinor. So once the law is local and first order, its internal structure is not a choice: it is a spinor.

**Physical justification**

Elementary once the squaring condition is imposed with the identity on the right-hand side. The load-bearing subtlety is in that "**1**" — this is the central condition, and it is stronger than Lorentz covariance alone. It says that no component of the fundamental observer-occupying field is merely auxiliary or algebraically constrained away.

*Remark* (First order alone does not force the Clifford algebra). The claim "first order + Lorentz-covariant + reproduces Klein–Gordon ⇒ Clifford" is *false* as usually stated, and a referee will catch it. The counterexample is the Duffin–Kemmer–Petiau equation $(i\beta^\mu\partial_\mu - m)\psi = 0$, a first-order, Lorentz-covariant wave equation for spin-0 and spin-1 whose matrices obey the Kemmer algebra

$$\beta^\mu\beta^\nu\beta^\rho + \beta^\rho\beta^\nu\beta^\mu = \beta^\mu\eta^{\nu\rho} + \beta^\rho\eta^{\nu\mu},$$

*not* the Clifford anticommutator; its representations are 5- and 10-dimensional and reproduce Klein–Gordon/Proca. So a first-order covariant equation with the right dispersion can describe integer spin with a non-Clifford algebra. What distinguishes Dirac is the extra observer-first condition stated in the theorem: $H^2 = (p^2 + m^2)\mathbf{1}$ *times the identity on every record-bearing component*. Lorentz covariance supplies the mass shell; it does not by itself forbid auxiliary constrained components. DKP fails the stronger condition (its $\beta$ matrices are singular; some components are auxiliary constraints, and the squared operator is not scalar). The precise statement is therefore: *among local, irreducible, Stone-form, Lorentz-covariant matter laws whose components are all propagating distinguishable modes of one mass, the Dirac equation is unique.* DKP remains a valid first-order description of spin-0/spin-1 fields with auxiliary variables; it is not a counterexample to the record-faithful matter-form claim. The irreducibility/scalar-square condition is essential; bare first-order uniqueness is not the claim.

**Why auxiliary components are excluded here**

Auxiliary components are not unphysical in ordinary field theory; they are often a useful way to write constrained bosonic equations. The narrower claim here concerns fundamental observer-occupying degrees of freedom: a component that never corresponds to an independently distinguishable propagating record state is not part of the minimal matter witness supplied by observer closure. This is the observer-side reason for imposing record-faithful irreducibility before the Clifford conclusion is drawn.

**Field-theoretic lift**

The argument here fixes the local first-order relativistic form of the matter equation. It does not by itself construct the full field-theoretic Fock representation, antiparticle sector, or Standard Model census. Those belong to the operator-algebraic lift, where locality, positivity, modular structure, and statistics must be treated together.

## 36 Occupation: the observer as the existence proof

Availability asks what forms of matter can exist; occupation asks whether any such sector is actually populated. In this framework the minimal answer is supplied by the premise that begins the argument: the observer exists as matter. This is not a fifth axiom. It is the operational fact without which there is no observer to define protocols, records, proper time, or admissible interaction in the first place.

By observer closure (Section 25), the observer is a bounded stress-energy source on the right-hand side of Einstein's equation. By the local $U(1)$ structure (Section 12.2), ordinary material observers furnish at least one sector that couples through the derived charged gauge form. By Stone evolution (Section 10.2) and interactive locality O3, the observer's record-bearing degrees of freedom are local and dynamical in the operational time parameter. The existence of the observer therefore supplies at least one occupied local matter sector of the kind constrained above, while the details of charge assignments and multiplicities remain empirical.

Part V identifies the form of any such record-faithful first-order sector: locality plus relativistic dispersion and scalar squaring select the Clifford algebra, and the irreducible field carrying that structure is the Dirac spinor. Thus occupation is established at the minimal level: some spinorial matter exists, because the observer is made of matter of the relevant record-bearing kind. What remains outside the framework is the detailed inventory of nature: how many spinor families exist, what their masses are, how the non-Abelian charges are distributed, and why the Standard Model has its particular multiplicities.

**In plain English**

The argument does not prove that observers exist; it starts from the fact that they do. Once that fact is admitted, the observer is already a piece of local matter with energy, dynamics, interactions, and records. The previous sections show what form such matter must take. The result is therefore form without particle bookkeeping: Dirac-type matter is occupied, but the full particle census is not derived here.

**Physical justification**

The observer's existence is the same kind of starting recognition as A0 and O1–O4. It is not a hidden empirical add-on introduced at the end. The framework asks what physics must look like given real finite embedded observers, and the matter composing those observers is the minimal witness that the first-order spinorial sector is occupied. This closes only the existence question; it does not derive generations, masses, couplings, or the full Standard Model representation content.

## 37 Fermionic statistics and two limitations

**Proposition 37.1** (Fermi statistics, given occupation)**.** *If the spin-$\frac{1}{2}$ sector is occupied, the spin–statistics theorem — which requires Lorentz invariance, microcausality, and positivity of energy — forces those excitations to obey Fermi statistics. All three hypotheses are already in the framework: Lorentz invariance (Section 22), microcausality (Section 31, from O3), positive energy (O1).*

**In plain English**

The stable-matter argument should run in the consequence direction, not as an anthropic selection principle. Given that a half-integer sector is present, Fermi statistics is not an extra assumption: it is a theorem from ingredients the framework already owns. Stability of bulk matter is then a *consequence*, not a premise.

**Physical justification**

Spin–statistics is the import; its three hypotheses are native (O1, O3, derived Lorentz). The stability of bulk matter is then a consequence of the spinorial sector rather than a premise used to obtain it.

*Remark* (Two limitations)*. First, the argument is first-quantized.* Theorem 34.1 concerns the c-number wave equation of a single relativistic excitation. In full quantum field theory the field is an operator whose Heisenberg evolution is trivially first order in time, so "Klein–Gordon is excluded as the fundamental law" is properly a statement about the first-quantized wave operator, not the operator algebra. This suffices to derive the spinorial *form* (all Part V claims), but the lift to the operator theory — and to the antiparticle/Fock structure — is not automatic and is not done here.

*Second, a possible double-use of locality.* Theorem 34.1 leans on O3-locality to exclude the nonlocal $\sqrt{p^2 + m^2}$, and separately on the derived Lorentz structure of Section 22. If O3's "locality" already presupposed the relativistic causal structure that Section 22 derives, the argument would partly circle. The framework builds the metric *downstream* of O3 (O3 is no-signalling at finite speed; the Lorentzian metric is reconstructed later), so the two are formally separable — but this separation should be argued explicitly in any standalone write-up, not left implicit.

## 38 The matter chain, restated

1. Lorentz acts inertly with positive energy, so spinor representations of $SL(2,\mathbb{C})$ are *available*. *(A0, O1; Wigner, Bargmann)*

2. Stone already makes evolution first order in time; O3 forbids the nonlocal escape $\sqrt{p^2+m^2}$; Lorentz covariance then selects first order in space. *(Section 10, O3, Section 22)*

3. Local + first order + record-faithful scalar squaring $\Rightarrow$ Clifford algebra $\Rightarrow$ the Dirac equation; matter is spinorial *in form*. *(auxiliary DKP sectors excluded)*

4. The observer's own existence supplies the minimal occupied matter witness through fundamental spin-$\frac{1}{2}$ constituents of ordinary matter; the remaining particle census stays empirical. *(observer closure)*

5. Given occupation, Fermi statistics follows. *(spin–statistics; O1, O3, Lorentz)*

The form of any fundamental local relativistic matter field is the Dirac spinor; observer closure supplies the minimal occupied witness of that form through the fundamental spin-$\frac{1}{2}$ constituents of ordinary matter. This remains a *form* result, in the same spirit as the gauge-group-form result of Part IV: it does not derive generations, masses, couplings, or the full particle census.

# Synthesis

## 39 The full chain, restated

Read top to bottom; each line is derived, reconstructed, selected, or imported by the line above plus the named constraint, as indicated by the graph and surrounding text.

1. You exist as a finite physical system inside the world. *(the one recognition)*
2. Physics is what survives all inert re-descriptions. *(A0)*
3. You measure only statistics, so states are distinguishability classes of histories. *(A0, O1–O4)*
4. Finite randomized preparation plus operational completion makes the state space convex; independent composition makes it a cancellative commutative monoid. *(R0, O3)*
5. Cancellativity $\Rightarrow$ a Grothendieck group; completeness, sectorwise connectedness, countable protocol base, and *torsion-freeness* (via Pythagoras) $\Rightarrow$ it is $\mathbb{R}^n$. *(R1–R2; Pontryagin imported)*
6. Operationally inert, sufficiently transitive redecomposition selects the Euclidean norm, hence an inner product. *(A0 + R3 operational isotropy)*
7. Norm-preserving free dynamics leave one sectorwise global phase: the group is $U(1)$. *(A0 + irreducible sector)*
8. A division scalar field containing $U(1)$ and compatible with independent composition leaves $\mathbb{C}$ as survivor: complex Hilbert space. *(Frobenius, imported)*
9. Causal independence + operational local completeness $\Rightarrow$ systems compose by $\otimes$. *(R4 motivated by A0+O3)*
10. Context-independent probabilities are a frame function; on dim $\geq 4$ (observer always present) Gleason gives the Born rule. *(Gleason, imported)*
11. Inner-product-preserving continuous evolution is unitary, $U(\tau) = e^{-iH\tau/\hbar}$. *(Wigner, Stone, imported)*
12. Local phase freedom selects a massless $U(1)$ gauge connection: electromagnetism. *(A0 + minimality + Lorentz)*
13. Finite operational energy sets a minimum duration for distinguishability-changing operations. Therefore a finite observer/controller cannot refine a protocol without bound; for CPMG, $n_{\max} = 2f_R T = 2E_{\text{op}}T/(\pi\hbar)$. *(ML imported; ceiling derived)*
14. Many observers $\Rightarrow$ a strict causal order with no time loops. *(O2, O3)*
15. One clock per observer $\Rightarrow$ Lorentzian signature; causal order + proper time $\Rightarrow$ the full metric; passive flat coordinates identify the local-flatness clause of the equivalence principle. *(Malament, imported)*
16. Landauer cost of horizon record-keeping *is* the Clausius relation; with Jacobson, the metric obeys Einstein's equations (observer included in $T_{\mu\nu}$). *(Bekenstein, Unruh, Jacobson, imported)*

17. The clock lives in the algebra; quotienting clock-shifts and adjoining the physically resolved clock flow gives the modular crossed product—a type II factor—on which Bekenstein and generalized entropy are theorems in the stated horizon settings. *(BW, Reeh–Schlieder, Haagerup, Takesaki, KLRSS, CLPW, imported)*

The deepest structural inputs across the full arc are **A0 itself** and, in Part IV, **the cyclic separating vacuum**. The Part I quantum reconstruction additionally uses the explicitly named regularity conditions R0–R4: finite randomized closure with completion, countable protocol base, sectorwise continuous accessibility, operational isotropy, and operational local completeness. These are not hidden derivations from A0 alone; they are the minimal observer-compatible regularities under which the irreducible nonclassical sector becomes complex Hilbert quantum theory. Everything else is a named theorem used exactly where invoked, or an empirical constant ($G, h, c, \Lambda, e$, the SM gauge groups).

## 40 What the framework establishes

Part II supplies the most concrete laboratory-scale operational contact: a finite-control-energy ceiling on protocol refinement. It is parameter-free at the level of the pulse-count envelope $n_{\max} = 2f_R T = 2E_{\text{op}}T/(\pi\hbar)$, while the associated coherence-time scalings and thermal timing floors are explicitly model-dependent. The local-flatness clause of the equivalence principle (Section 24) is the cleanest geometric surprise: $\mathrm{A0}^+$ identifies first-order gravitational-force terms removable by Riemann normal coordinates as inert, while curvature remains invariant. The type II identification for Rindler and de Sitter (Section 30) is a theorem that the independent gravity-side program (Witten, CLPW) reached from the opposite direction, the strongest external corroboration anywhere in the arc.

The Hilbert-space and Born-rule results lean on Gleason, Frobenius, Wigner, and Stone—all imported. The contribution there is not the machinery but the operational setup that makes it apply: the frame-function property forced by A0, and the qubit-gap closure that guarantees Gleason's dimension hypothesis. That is reconstruction, not derivation from nothing, and is labeled as such.

The open problems, in descending order of how much resolving them would move the picture: the completeness question—whether an exotic theory satisfying O1–O4 could escape the Hilbert-space conclusion; the general-background type II convergence, on which the Bekenstein and entropy results rest outside the cases where it is proved; the finite-control-energy CPMG test, which asks whether a platform approaches the refinement ceiling once ordinary hardware limits are accounted for; and the cosmological constant, an undetermined integration constant rather than a prediction.

This is not a theory of everything. It does not derive the Standard Model gauge group, particle masses, or $\Lambda$; it does not reconstruct Einstein's equations from Landauer alone—the Clausius relation is reconstructed, and Jacobson supplies the dimensional lift and the area law; and it does not derive the full particle census. Observer closure supplies a minimal occupied matter witness, while the detailed multiplicities and couplings remain empirical. It constrains the form of physical law that any finite embedded observer must share, and names at each step exactly which alternatives die and where a result is imported rather than derived. The deepest claim is conditional: any operational theory consistent with real, finite, embedded observers must pass through this structure or exhibit exactly where one of the stated constraints fails.

## A Master symbol dictionary

Every symbol that appears in the chain, defined once. Later sections introduce a few more in purple boxes at the point of first use, but the load-bearing notation is all here.

### Objects the observer deals in

$\mathcal{H}$ The set of all admissible *histories*. A history is one complete physical unfolding — one full causal story of what happened. Not a Hilbert space (that comes later, and gets a subscript).

$H \in \mathcal{H}$
"$H$ is an element of $\mathcal{H}$": $H$ is one particular admissible history.

$\mathcal{O}$ An observer: a physical system that interacts, records, and compares records. Not a disembodied viewpoint — a chunk of matter.

$\mathcal{F}$ An *admissible family*: a set of observers, each obeying O1–O4, any two of which are linked by a finite chain of physical interactions.

$\tau$ The observer's *proper time* — the single continuous parameter its own internal dynamics generate. Its "wristwatch."

$M$ The size of the observer's memory register (a finite number of distinguishable internal states; $\dim \mathcal{H}_{\mathcal{O}} \leq M$).

$E_{\mathcal{O}}$, $E_{\text{op}}$
Total energy of the observer, and the portion of it available for control/measurement ($E_{\text{op}} \leq E_{\mathcal{O}} < \infty$). In Part II, $E_{\text{op}}$ is the norm of the effective pulse Hamiltonian.

$\Omega_R$, $f_R$
Rabi angular frequency and Rabi frequency in cycles per second, with $f_R = \Omega_R/(2\pi)$.

$t_\pi$ Ideal resonant $\pi$-pulse duration, $t_\pi = \pi/\Omega_R = 1/(2f_R)$.

$\tau_{\text{ML}}$ Margolus–Levitin minimum operation time, $\tau_{\text{ML}} = \pi\hbar/(2E_{\text{op}})$.

$n_{\max}$, $n_{\text{sat}}$
Maximum physically executable pulse count $n_{\max} = 2f_R T = 2E_{\text{op}}T/(\pi\hbar)$, and observed pulse number $n_{\text{sat}}$ at which additional CPMG pulses cease to improve coherence.

### Experiments and their statistics

$\pi \in \Pi$
An admissible *protocol*: one experiment, written as a finite adaptive sequence of interactions $\pi = (I_1, f_1, I_2, f_2, \ldots, I_n)$. $I_k$ is an interaction step (a detector firing, a spin probe), $f_k$ is an update rule that chooses the next step from earlier outcomes (feedback).

$\Omega$ The outcome space — the set of things that can be recorded (clicks, spin values, bit strings).

$P(\Omega)$
The set of probability distributions over $\Omega$.

$X_H(\pi)$
The outcome distribution you get by running protocol $\pi$ on history $H$. An element of $P(\Omega)$. This is the *only* thing the observer ever actually sees.

### Distance, equivalence, structure

$D_{\mathrm{TV}}(P, Q)$
: Total-variation distance, $\frac{1}{2}\sum_i |P_i - Q_i|$: how well two distributions can be told apart by the best single guess. 0 means identical, 1 means perfectly distinguishable.

$d(H_1, H_2)$
: *Operational distance* between histories: the best distinguishability any admissible experiment can achieve, $\sup_\pi D_{\mathrm{TV}}(X_{H_1}(\pi), X_{H_2}(\pi))$.

sup
: Supremum — the least upper bound, a "generalized maximum" that exists even when no single $\pi$ attains it.

$\sim$
: Operational equivalence: $H_1 \sim H_2 \iff d(H_1, H_2) = 0$ (no experiment can tell them apart).

$[H]$
: The equivalence class of $H$: all histories indistinguishable from it. This is what the framework calls a physical *state*.

$S_{\mathcal{O}}$
: The operational state space $\mathcal{H}/\sim$: histories grouped into states.

$\oplus$
: Composition of two *causally independent* systems (put two non-communicating things side by side).

$\otimes$
: Tensor / joint composition (run protocols jointly; later, the joint Hilbert space).

$G(S_{\mathcal{O}})$
: The Grothendieck group of $S_{\mathcal{O}}$: the smallest group that lets you "subtract" states, built from formal differences $[s] - [t]$.

$\theta$, 0
: The empty/identity state "$\varnothing$" (composing it changes nothing) and the group identity $0 = [\varnothing] - [\varnothing]$.

### Physics that emerges further down

$\mathcal{H}_{\mathcal{O}} \cong \mathbb{C}^m$
: The observer's state space *after* the reconstruction: a finite-dimensional complex Hilbert space.

$\rho$, $P_E$
: A density operator (a state, written as an operator) and a projector onto a measurement outcome $E$; the Born rule reads $P(E) = \mathrm{Tr}(\rho P_E)$.

$U(\tau) = e^{-iH\tau/\hbar}$
: Unitary time evolution generated by the self-adjoint Hamiltonian $H$ (the observer's clock generator).

$A_\mu$, $F_{\mu\nu}$, $D_\mu$
: The gauge potential, its field strength $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and the covariant derivative $D_\mu = \partial_\mu - ieA_\mu$; these are the electromagnetic field.

$g_{\mu\nu}$, $G_{\mu\nu}$, $T_{\mu\nu}$, $\Lambda$
: The spacetime metric, the Einstein tensor, the stress-energy of matter (observers included), and the cosmological constant.

$\preceq$
: Causal order: $e_1 \preceq e_2$ means a signal at speed $\leq c$ can run from $e_1$ to $e_2$.

$\mathfrak{A}(\mathcal{O})$, $\mathfrak{A}_{\mathrm{OFP}}$
: The local field algebra on region $\mathcal{O}$, and the observer's physical algebra built from it plus the clock.

$\sigma_t$, $\Delta_\Omega$, $K_\Omega$
: Modular flow, modular operator, and modular Hamiltonian $K_\Omega = -\ln \Delta_\Omega$ of Tomita–Takesaki theory.

$\rtimes_\sigma \mathbb{R}$
: Crossed product: the algebra you get by adjoining the flow $\sigma$ (here, the clock) to $\mathfrak{A}$ as new dynamical degrees of freedom.