Synthetics: Statistical Lawhood Under Finite Description

Gerard McCaul
(Dated: May 15, 2026)

Synthetics is the study of finite representations of lawful systems: how laws are produced, transformed, and partially lost under changes of description. This paper develops an operational framework for Synthetics by showing how a representation C acting on an antecedent generator G_μ sorts its structure into a *synthetic decomposition* $S_C(G_\mu) = (q_s, K_s, \Delta_C)$: the null law absorbed into the coarse background, the visible generator pricing retained complexity, and the closure residue carrying structure with no closed coarse image.

Four requirements on scalar law-content—relabelling invariance, atomwise locality, neutral refinement, and order independence—force the comparison between descriptions to the KL form in a finite positive-support setting; adding visible constraints forces the variational selector to the free-energy family. The free-energy functional $\mathcal{F}_s[p] = \langle K_s \rangle_p + M D_{\mathrm{KL}}(p\|q_s)$ is the invariant scalar grammar of the synthetic decomposition in the order-independent regime; thermodynamics, rate-distortion theory, the information bottleneck, the free-energy principle, and minimum description length are instances under different choices of visible generator K_s.

Order independence is the scalar statistical form of partition-dynamics commutation: where it holds, KL is uniquely forced; where it fails globally, noncommutativity is forced. When dynamics and description-change fail to commute, the closure residue $\Delta_C = CG_\mu - G_M C$ composes coherently across scale towers; its structured modes—renormalised couplings, memory kernels, noise, emergent order parameters, gauge freedom, holonomy, and spectral noncommutativity—are a complete classification of effective physics as structured closure failure.

Taking the infinitesimal scale limit yields a residue density ω_s whose off-diagonal component drives spectral-frame rotation. The dimensionless ratio χ_s of off-diagonal residue to spectral gap measures distance from classical closure; when χ_s flows to a non-trivial fixed point, the canonical commutator $[Q, P] = i\hbar$ is the algebraic signature of partition-dynamics obstruction stabilised at a scalar, central, universal value.

I. FINITE LAWHOOD AND RECURSIVE CLOSURE

Theories idealise. Any theory selects a finite set of variables, declares the rest invisible, and writes its laws in the compressed representation that remains. Formally this is a coarse-graining; every object of theory—particle trajectories, density matrices, continuum fields, block-spin configurations—is a compressed image of phenomena richer than itself.

Every mature quantitative theory built this way ends up with the same variational architecture: an expected cost plus a divergence from a reference law. Statistical mechanics takes this form with energy as cost [1, 2]; rate-distortion theory with distortion [3]; variational inference with log-likelihood [4]; the information bottleneck with negative relevance [5]; minimum description length with code length [6]. The coincidence has been noted, and partially explained from different consistency axioms [1, 7, 8], but not derived from the one property all these theories share: the act of declaring structure invisible. This paper derives it from there.

This convergence has a name. *Synthetics* is the study of finite representations of lawful systems: how laws are produced, transformed, and partially lost under changes of description. The projection-operator formalism [9, 10], the renormalisation group [11, 12], effective field theory [13, 14], maximum entropy methods [1, 7], and quantum reconstruction programmes each engage this problem; this paper develops one operational framework inside it. The name is literal: a finite description *synthesises* its law from antecedent structure, sorting that structure into what closes exactly, what survives only approximately, and what has no closed coarse image.

Which kind of idealisation a theory performs depends on the relation between its register and what it describes (Fig. 1). Let Ω be the configuration space of the world, $R \subset \Omega$ the physical register carrying the theory, and $\pi : \Omega \to X$ the symbol map by which the register represents the world. Three regimes are structurally distinct: a *disjoint* theory has a register lying entirely outside its target; a *coupled* theory partially overlaps its target; a *reflexive* theory lies inside it, so that $R \subset \Omega$ and the target is Ω itself. Fundamental physics, cosmological theories, computational theories, and biological self-models all occupy the reflexive category [15, 16].

A finite reflexive theory compresses: $|X| < |\Omega|$. Exact reflexive self-description would require the compressed image to contain its own representational action without residue—a self-similar fixed-point condition that generically fails. A consistent finite self-description must close under its own application; wherever closure fails, the failure is structured residue whose composition laws determine whether the theory can remain predictive. What those laws must be is the question this paper answers.

The structural question is:

> *How can a finite idealisation stand in lawful relation to phenomena richer than itself?*

Let X_μ denote the antecedent state space and X_M the

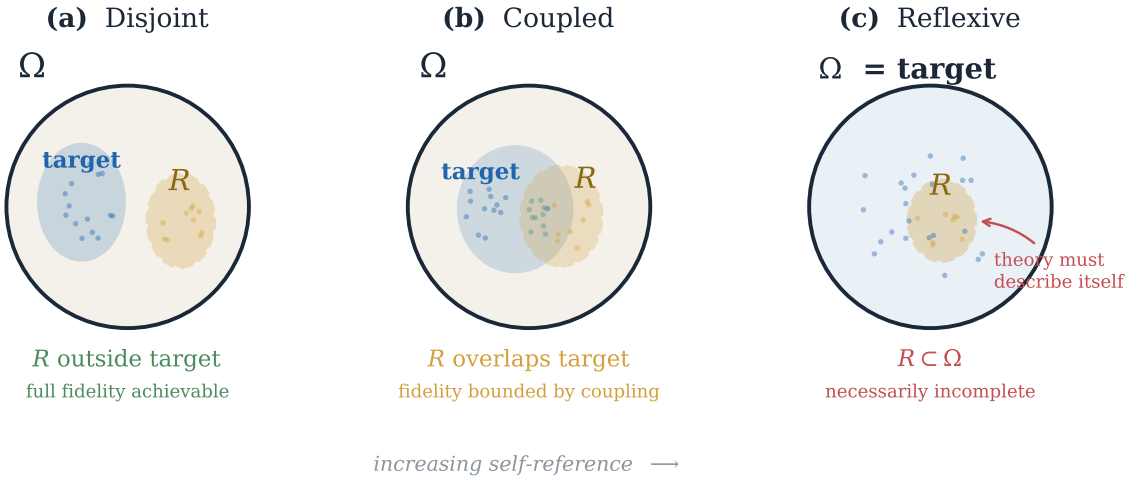

FIG. 1. **Three representational regimes.** *Disjoint:* the register R lies entirely outside its target; the theory and the modelled system share no state space. *Coupled:* R partially overlaps its target. *Reflexive:* $R \subset \Omega$ and the target is Ω itself. Fundamental physics, computation, and biological self-models are reflexive. A finite reflexive theory can close exactly only at a self-similar fixed point; generic finite compression leaves structured residue whose composition laws are the subject of this paper.

coarser state space of the theory. A representation is a map

$$C : X_\mu \to X_M. \tag{1}$$

The fibre $C^{-1}(m)$ collects all antecedent states the theory declares equivalent. Throughout, *antecedent* means representationally prior to the coarse description—the structure on which C acts. The term carries no requirement of a physical scale hierarchy; the antecedent generator G_μ may govern a field theory, a microscopic model, an agent's internal model, or any dynamics richer than what C retains.

Whether the coarse description can generate its own dynamics—without constant reference to the antecedent level—is the *closure question*. Renormalisation-group decimation [11, 12], projection-operator elimination [9, 10, 17], and effective-field-theory truncation [13, 14] all perform irreversible changes of description, and their residue structure is the subject of this paper. If G_μ is the antecedent generator and G_M the coarse generator, exact closure requires

$$CG_\mu = G_M C : \tag{2}$$

evolving at the antecedent level then coarse-graining gives the same result as coarse-graining first then evolving with the coarse law. When this fails, the *closure residue*

$$\Delta_C = CG_\mu - G_M C \tag{3}$$

is the determinate part of antecedent dynamics not internally generated by the coarse law but still relevant for predictions. Because G_μ evolves states *within* a description while C changes the description itself, Δ_C is a higher-order obstruction—the failure of partition-change and dynamics to compose independently of their order—rather than an operator commutator on a fixed state space. How it transforms under iterated coarse-graining, and at what scale it stabilises at a non-trivial fixed point, are the subjects of Sections III and VI.

II. SYMMETRY, CONSTRAINT, AND THE FORCED LAW-FORM

A statistical law is a generator: it drives the system toward a stationary state and determines how fast. The antecedent generator G_μ carries structure beyond its stationary state: symmetries, conserved quantities, an eigenspectrum. Coarse-graining C preserves some of that structure exactly, demotes some to approximate, and erases the rest (Fig. 3). Each fate has a name, and the names together force the law-form.

Structure that C inherits exactly from the antecedent level is absorbed into the coarse background as the *null law*: the minimally committed distribution compatible with the current partition, generated recursively by the coarse-graining itself. Let q_{s-ds} be the null law at the antecedent scale, and let $C_* q_{s-ds}$ denote its pushforward to X_M — the measure defined by $(C_* q)(m) = \sum_{x:\, Cx=m} q(x)$. Projecting onto the admissible class $\mathcal{N}_s$ of distributions on X_M consistent with the current partition gives the null law at scale s:

$$q_s = \arg\min_{q \in \mathcal{N}_s} D_{\mathrm{KL}}(C_* q_{s-ds} \| q)\,. \tag{4}$$

The null law carries exactly the structure preserved whole by the coarse-graining: antecedent dynamical symmetries that close exactly under C are absorbed without

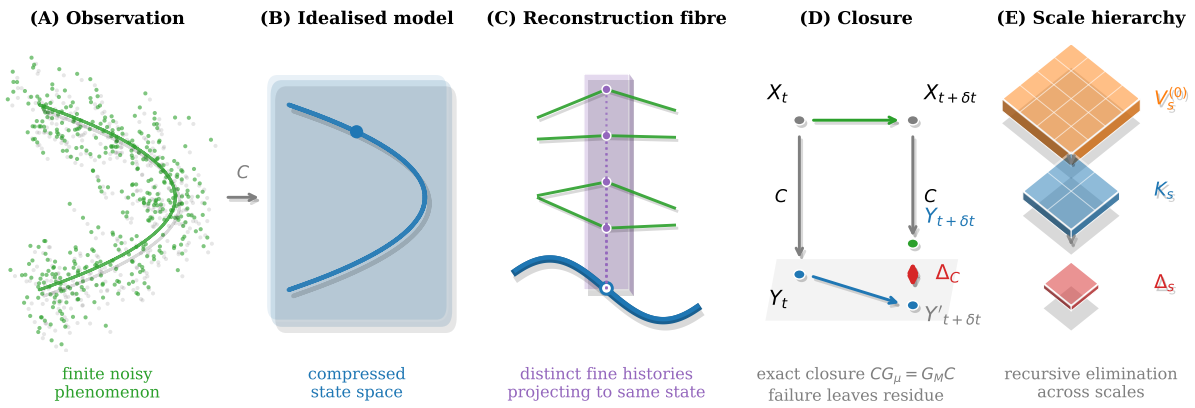

FIG. 2. **Finite law-form under change of description.** *(A)* Finite, noisy phenomenon: data lying near an idealised curve in the antecedent state space. *(B)* Coarse-graining C compresses it to a smaller state space X_M; the compressed curve is the idealised model. *(C)* Reconstruction fibre: each coarse state admits many antecedent histories; the fibre $C^{-1}(m)$ collects all of them. *(D)* Closure residue: evolving at the antecedent scale and coarse-graining yields $Y_{t+\delta t}$ (green); coarse-graining first and evolving coarsely yields $Y'_{t+\delta t}$ (grey); the gap $\Delta_C = CG_\mu - G_M C$ (red) is the determinate closure failure. *(E)* Scale triage: antecedent structure distributes into null potential $V_s^{(0)}$ (amber), visible complexity K_s (blue), and closure residue Δ_s (red). See Sections II–III.

remainder into the coarse background; everything else has been discarded. Define the *inherited null potential*

$$V_s^{(0)} = -M \log q_s \quad \text{(up to additive constant)}, \tag{5}$$

where $M > 0$ is a positive scale parameter whose role is clarified by Result 2 below. The null potential is accordingly a coarse-level encoding of perfectly compressible antecedent structure: the generator component whose partition-dynamics commutator vanishes exactly appears at the coarser scale as a static potential, requiring no explicit tracking. Derivation of the null-law recursion in Appendix B.

Demoted symmetries carry a cost: the coarse description must impose as an explicit constraint what the antecedent law satisfied automatically. Those demoted constraints are the visible observables A_a; their conjugate coefficients g_a are the costs of making those distinctions matter. The cost function of retained complexity is

$$K_s = \sum_a g_a A_a, \tag{6}$$

the image in the coarse description of antecedent symmetry-constrained structure that the coarse-graining has demoted from exact to approximate. Together, the null potential and the visible cost determine the stationary state:

$$\gamma_s \propto \exp\left[-\frac{K_s + V_s^{(0)}}{M}\right]. \tag{7}$$

Appendix J derives this projection explicitly for integrable generators, where action-angle coordinates make the absorbed/demoted split geometric.

The third class of structure—broken symmetries—has no coarse image. The coarse description must still compare representations consistently: a comparison that gives different answers depending on the order in which irrelevant structure is discarded measures nothing objective.

Four requirements express what any consistent scalar comparison $\mathcal{I}(p\|q)$ between distributions on a finite state space must respect.

Relabelling invariance. The comparison cannot depend on arbitrary names of states. Bijections of the state space leave $\mathcal{I}$ unchanged.

Atomwise locality. Disjoint alternatives contribute separately: $\mathcal{I}(p\|q) = \sum_i \ell(p_i, q_i)$ for a common local density ℓ.

Neutral refinement. Splitting an alternative into subalternatives with the same likelihood ratio p_i/q_i and no new visible distinction leaves $\mathcal{I}$ unchanged.

Order independence. Eliminating hidden distinctions in one step or in several must give the same total accounting:

$$\begin{aligned}\mathcal{I}_X(p_X\|q_X) &= \mathcal{I}_Y(p_Y\|q_Y) \\ &+\mathbb{E}_{p_Y}\mathcal{I}_{X_y}\left(p_{X|Y}(\cdot|y)\|q_{X|Y}(\cdot|y)\right).\end{aligned} \tag{8}$$

Order independence is the substantive requirement. Any scalar failing (8) assigns different statistical content to a representation depending on the order in which hidden structure is discarded—incompatible with that structure being genuinely irrelevant. It is the scalar statistical form of partition-dynamics commutation: it requires that changing description and measuring law-content compose in the same order. In the finite positive-support setting this condition holds and the KL form follows. The regime where order independence fails globally—where

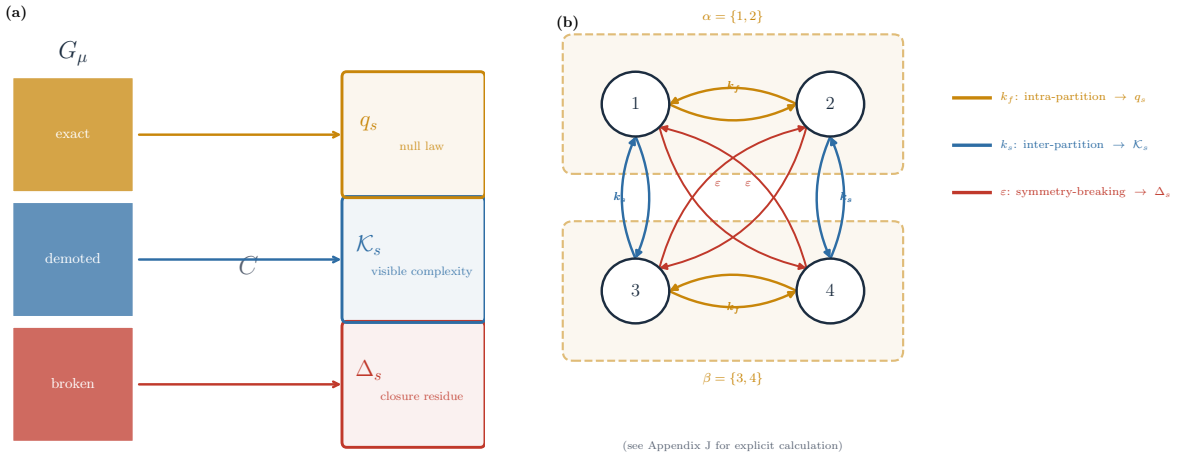

FIG. 3. **Coarse-graining sorts the symmetry content of a generator.** *(a)* The antecedent generator G_μ contains three classes of structure, distinguished by how the coarse-graining C acts on them. Exact symmetries are absorbed into the null law q_s (amber); demoted symmetries become visible constraints entering $\mathcal{K}_s$ (blue); broken structure has no coarse image and contributes to the closure residue Δ_s (red). *(b)* A four-state Markov chain with coarse partition $\alpha = \{1,2\}$, $\beta = \{3,4\}$ instantiates all three fates explicitly. Intra-partition transitions at rate k_f (amber) equilibrate instantly inside each partition and are absorbed into the null law q_s. Inter-partition transitions at rate k_s (blue) project cleanly through C and enter $\mathcal{K}_s$ as explicit constraints. Diagonal transitions at rate ε (red) break the partition symmetry and cannot be expressed in any coarse functional, contributing to Δ_s. Full derivation in Appendix J.

partition-dynamics non-commutation leaves a persistent spectral residue—is the subject of Section VI.

Result 1. *For finite positive-support distributions, relabelling invariance, atomwise locality, neutral refinement, and order independence together force*

$$\mathcal{I}(p\|q) = C \sum_i p_i \log \frac{p_i}{q_i}. \tag{9}$$

Proof in Appendix A. Relative entropy emerges here as the unique scalar bookkeeping compatible with consistent multi-scale description. Shore and Johnson [7] derived the same uniqueness from different consistency axioms on inference; the present derivation traces it to the more primitive requirement that the order of elimination be irrelevant. Every other continuous comparison violates (8) by an $O(1)$ amount on the simplest finite examples (Fig. 4).

With the comparison uniquely determined, imposing visible constraints $\langle A_a \rangle_p = a_a$ on top of the order-independent law-content yields the forced variational family.

Result 2. *For finite states with compatible support and $M > 0$, order-independent law-content combined with visible constraints gives*

$$\mathcal{F}_s[p] = \langle K_s \rangle_p + M D_{\mathrm{KL}}(p\|q_s), \tag{10}$$

where $K_s = \sum_a g_a A_a$ is the cost function of retained visible complexity and M is the positive scale parameter of (5).

The free-energy functional trades two costs: the expected cost of retained complexity, and the cost of departing from the null law. Proof and contraction-closure property in Appendix A.

The free-energy functional (10) is the invariant scalar grammar of the synthetic decomposition in the order-independent regime. Thermodynamics, rate-distortion theory [3], the information bottleneck [5], the free-energy principle [18], and minimum description length [6] are all instances under different choices of retained observable A_a: thermodynamics takes $A = H$ (energy) with $M = k_B T$; rate-distortion takes $A = d(x, \hat{x})$ (distortion) with the rate as null content; the information bottleneck takes $A = -I(T;Y)$ (relevance gap); Friston's free-energy principle takes $A = -\log p(\mu|\mathbf{s})$ (prediction error) on a Markov blanket; Rissanen's MDL takes $A = \ell_\theta$ (code length); variational Bayes is recovered when q_s is the prior and K_s is the negative log-likelihood. These are identifications within a single forced structure, not analogies (Fig. 5; Appendix A). The scale parameter M sets the units of law-content and is fixed by the choice of representation; it is not determined by the framework.

The representation C sorts antecedent generator structure into three branches:

$$\text{antecedent law} \longrightarrow \begin{cases} V_s^{(0)} = -M \log q_s, & \text{null potential,} \\ K_s, & \text{visible complexity,} \\ \Delta_s, & \text{closure residue.} \end{cases} \tag{11}$$

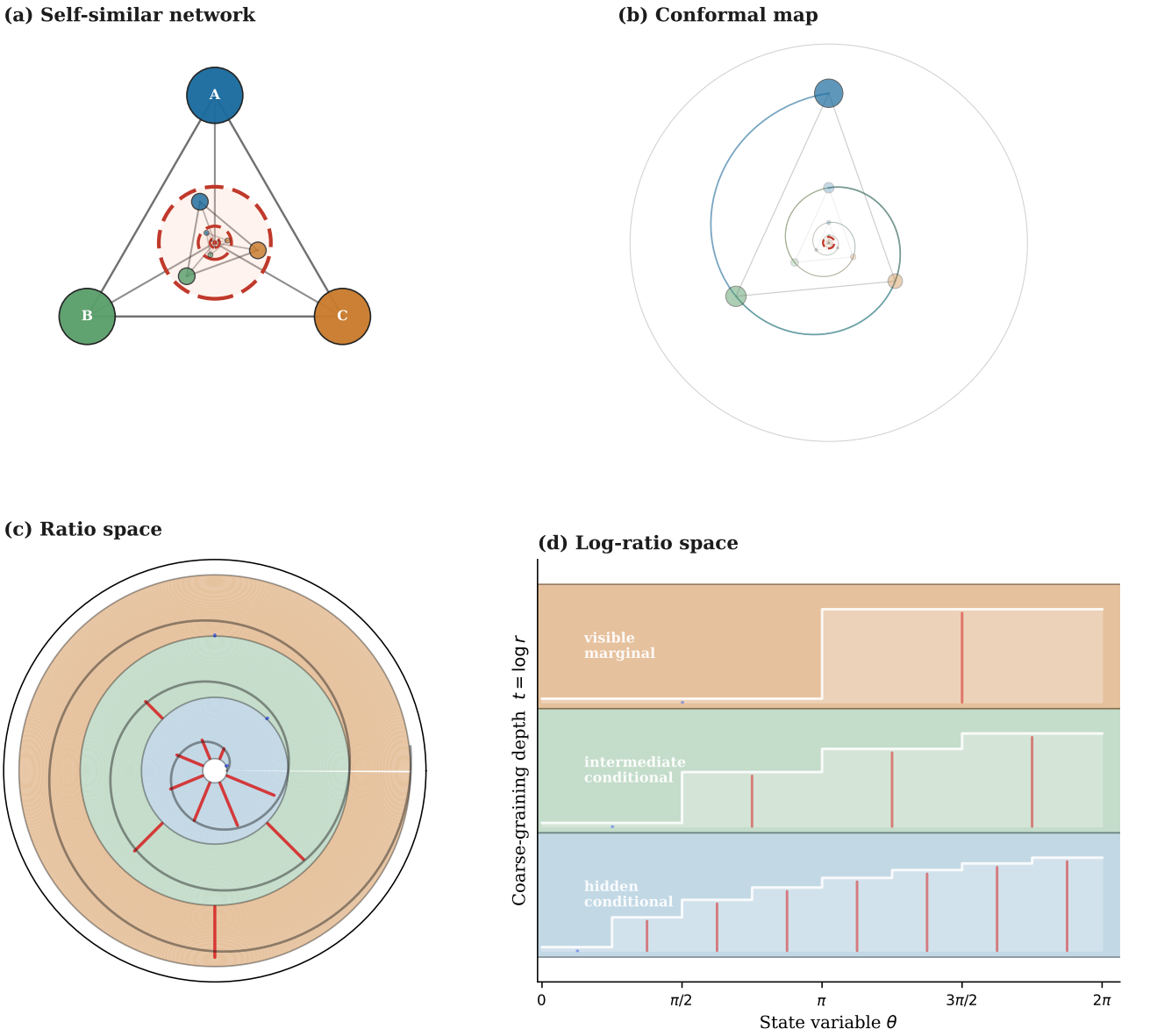

FIG. 4. **The logarithm is forced by order independence.** *(a)* Self-similar network: the wheel graph embeds its own hub at smaller scale, so ratio self-similarity is built into the structure. *(b)* Conformal log-spiral: the map $w = \log z$ converts multiplicative nesting into additive periodicity; each revolution of the spiral carries the same ratio structure. *(c)* Ratio space: probability ratios p_i/q_i for the rank ensemble $P_I(n) \propto n^{-I}$ displayed in polar coordinates, one ring per coarse-graining level. Red bars indicate $p_i/q_i > 1$; blue bars $p_i/q_i < 1$. *(d)* Log-ratio space: $\log(p_i/q_i)$ as a function of state variable θ, one band per level (fine, intermediate, visible). Logarithms add across levels: $\log r = \log r_Y + \log r_{X|Y}$, the chain rule (8). The additivity condition $\phi(ab) = \phi(a) + \phi(b)$ forces $\phi(a) = C \log a$, selecting the KL divergence uniquely. Derivation in Appendix A.

We call

$$S_C(G_\mu) = \big(q_s,\ K_s,\ \Delta_C\big) \tag{12}$$

the *synthetic decomposition* of the antecedent generator under C. This bookkeeping is the spine of the paper.

The coarse generator G_s is the drift toward γ_s; $\mathcal{F}_s$ is the Lyapunov functional that witnesses the contraction. The triage answers the closure question—whether G_M can be written in coarse variables—and the variational question—what functional form the law must take—simultaneously: the generator structure that closes gives null potential and visible complexity; the structure that cannot close gives the closure residue.

III. CLOSURE RESIDUE AND EFFECTIVE PHYSICS

Section II identified three branches of the triage (11); this section develops the third.

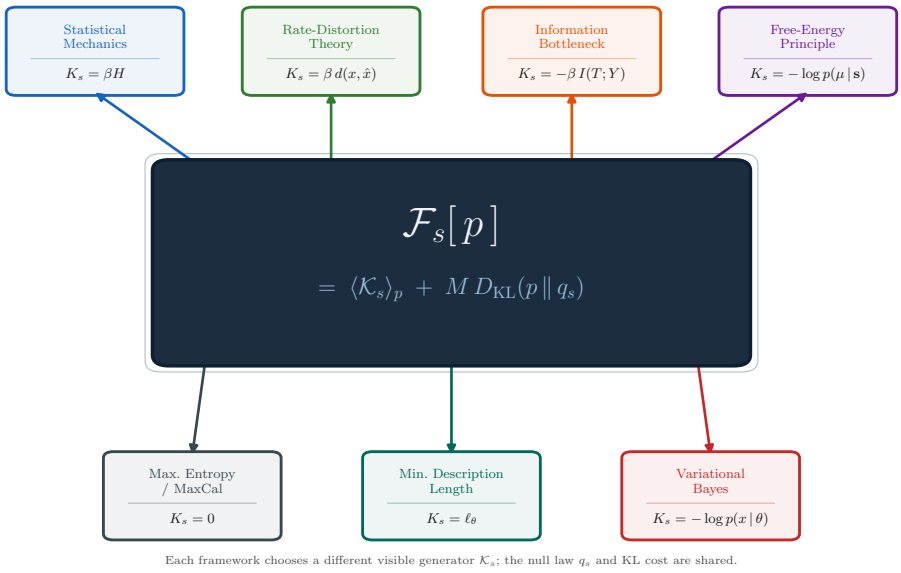

FIG. 5. **The forced law-form unifies seven variational frameworks.** The free-energy functional $\mathcal{F}_s[p] = \langle K_s \rangle_p + MD_{\mathrm{KL}}(p\|q_s)$ (centre) is shared by all seven frameworks. Each spoke chooses a different visible generator K_s; the null law q_s and the KL cost are common to all. Thermodynamics takes $K_s = \beta H$; rate-distortion theory [3] takes distortion; the information bottleneck [5] takes relevance gap; the free-energy principle [18] takes prediction error on a Markov blanket; Rissanen's MDL [6] takes code length; variational Bayes takes negative log-likelihood; MaxEnt [1] takes no visible constraint ($K_s = 0$) and returns directly to the null law. The correspondences are identifications within a single forced structure. Every framework in the figure is the case where order independence holds: partition and accounting commute, and the KL form is uniquely forced.

When an antecedent law is pushed forward through a coarse-graining, the result may not be representable exactly in the coarse law-form. The mismatch is the closure residue

$$\Delta_C = CG_\mu - G_M C, \tag{13}$$

already introduced in Section I. Whether it accumulates arbitrarily or composes with structure determines whether it represents physics or noise.

Result 3. *For a scale tower*

$$X_0 \xrightarrow{C_{01}} X_1 \xrightarrow{C_{12}} X_2, \tag{14}$$

with one-step residues $\Delta_{C01} = C_{01}G_0 - G_1C_{01}$ *and* $\Delta_{C12} = C_{12}G_1 - G_2C_{12}$*, the two-step closure residue satisfies*

$$\Delta_{C02} = C_{12}\,\Delta_{C01} + \Delta_{C12}\,C_{01}. \tag{15}$$

Proof in Appendix C. The cross term $C_{12}\,\Delta_{C01}$ is the key feature: residue at one scale is *filtered* by the next coarse-graining before entering the two-step mismatch. A mode that C_{12} kills contributes nothing to the accumulated residue; a mode it preserves carries its residue forward. This is why residue acquires the structure of effective physics rather than the statistics of noise: it has passed through the same filter as the surviving degrees of freedom. The structure of (15)—partition-dynamics obstruction at one level filtered by the next coarse-graining—has the character of a 1-cocycle in algebraic deformation theory [19]: the obstruction is itself obstruction-free one level up, and what propagates is the projection of the mismatch through the subsequent partition, not a free accumulation of errors. Effective physics is structured closure residue.

Appendix C derives a canonical classification of Δ_C (Proposition 6). The canonical modes are organised by three structural conditions: the spectral structure of the retained mismatch $A = CG_\mu C^+ - G_M$ (where C^+ is a right inverse of C satisfying $CC^+ = I_M$), the exact Nakajima-Zwanzig splitting of the hidden coupling, and the geometry of the coarse-graining fibration. These conditions yield the following residue modes; whether the list is exhaustive in all generality is not claimed. R1 (scalar retained): absorbed by coupling-constant shift—renormalisation-group flow [11, 20]. R2 (diagonal retained): alters transition rates—friction and kinetic corrections. R3 (off-diagonal retained): new coarse couplings—emergent order parameters [21, 22]. H1 (memory-kernel hidden): non-Markovian coarse dynamics—the Nakajima-Zwanzig kernel [9, 10, 17, 23]. H2 (incoherent hidden): fluctuating forces [24, 25]. G1 (fibre ambiguity): gauge freedom [26, 27]. G2 (incompatible projectors): spectral noncommutativity [28, 29].

R1–H2 are local corrections to rates, potentials, or cou-

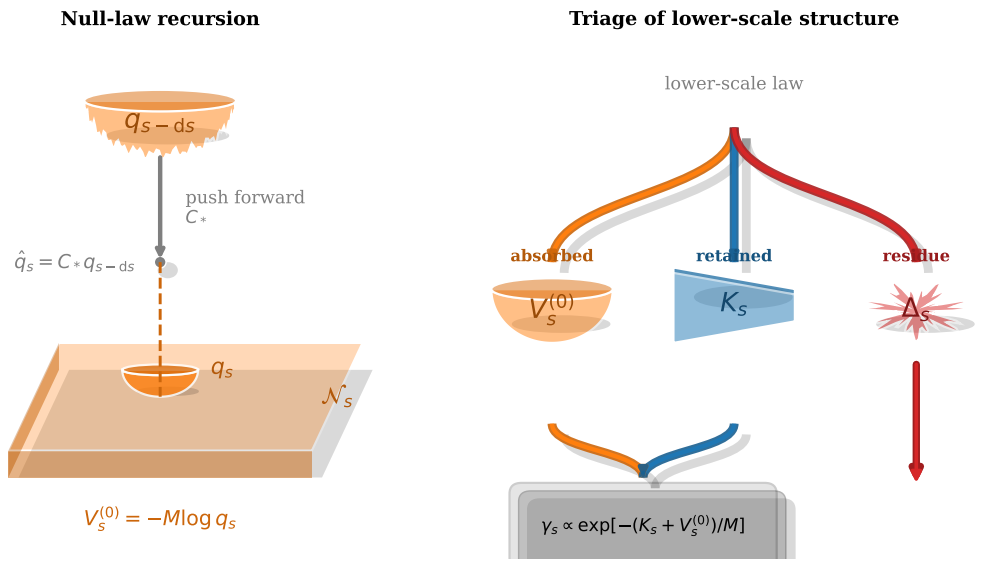

FIG. 6. **Null law recursion and residue triage.** *Left:* The antecedent null law q_{s-ds} is pushed forward through C to give $\hat{q}_s = C_* q_{s-ds}$, then projected onto the admissible class $\mathcal{N}_s$ to give q_s (amber, Eq. 4). The inherited null potential $V_s^{(0)} = -M \log q_s$ (amber) is the antecedent law structure absorbed into the current-scale background. The visible deformation K_s (blue) is resolved structure remaining after absorption; their sum $K_s + V_s^{(0)}$ is the total effective generator. *Right:* The triage of antecedent law (Eq. 11). Every part of the antecedent law is either absorbed (amber), retained (blue), or left as closure residue (red). Recursive null-law derivation in Appendix B.

plings. G1 and G2 are global obstructions—measuring the topology of the space of descriptions: whether the fibre can be globally trivialised and whether outcomes can be globally assigned across measurement contexts [28]. Full derivation and the two-mode open-system example are in Appendices C and D.

IV. CONSTRUCTIVE COARSE-GRAINING AND GENERATOR-CLOSING BASES

The preceding section classifies the closure residue Δ_C of a given coarse-graining. The same criterion can *select* the coarse-graining. This section turns the theory constructive: we derive a dynamics-aware objective for choosing the representation C, show why fast-mode compression alternatives fail, and demonstrate that the residue of a suboptimal representation points to what was forgotten.

An *antecedent generator* G_μ governs $\dot{p} = G_\mu p$ over n_μ antecedent states; a coarse-graining $C : \mathbb{R}^{n_\mu} \to \mathbb{R}^{n_M}$ ($n_M < n_\mu$) aggregates them into coarse variables. In the Markov-chain setting C is a binary aggregation matrix ($C_{\alpha i} \in \{0, 1\}$, exactly one 1 per column). Exact closure requires $CG_\mu = G_M C$—"evolve then coarsen" and "coarsen then evolve" must agree—so the closure residue $\Delta_C = CG_\mu - G_M C$ measures the failure and, more usefully, can guide the next revision.

The closure criterion

Given stationary distribution π_μ ($G_\mu \pi_\mu = 0$, $\pi_\mu > 0$) and $W = \mathrm{diag}(\pi_\mu)$, the stationary-weighted Frobenius norm is $\|A\|_{W,F}^2 = \sum_{ij} \pi_j A_{ij}^2$. The *best-fit coarse generator* for a given C minimises the weighted squared closure residue; its analytic solution is

$$G_M^*(C) = \left(CG_\mu W C^\top\right)\left(CWC^\top\right)^{-1}. \tag{16}$$

The *closure loss* is the residual after fitting the best possible coarse generator:

$$J(C) = \left\| CG_\mu - G_M^*(C)\, C \right\|_{W,F}. \tag{17}$$

Minimising $J(C)$ over candidate coarse-grainings yields a *generator-closing basis*—the representation that best compresses the dynamics, not merely the data.

Result 4. *Exact closure $J(C) = 0$ holds if and only if the row space of C is invariant under right multiplication by G_μ. Equivalently, in the Koopman dual, observables $\{\phi_i\}$ close exactly when*

$$G_\mu^\dagger \,\mathrm{span}\{\phi_i\} \subseteq \mathrm{span}\{\phi_i\}. \tag{18}$$

The proof is direct: $CG_\mu = G_M C$ for some G_M iff every row of CG_μ lies in the row space of C, i.e. the row space is G_μ-invariant under right action.

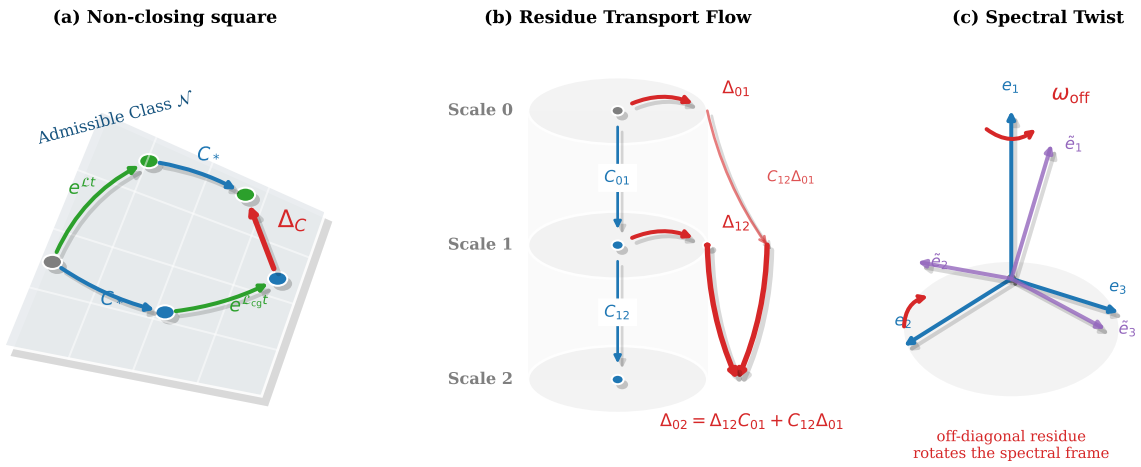

FIG. 7. **Closure residue: non-closing square, composition, and spectral limit.** *Left:* Two paths from (X_μ, t) to the coarse level at $t+\delta t$: evolve at the antecedent level then coarsen (green/blue), versus coarsen first then evolve coarsely (orange). The gap is $\Delta_C = CG_\mu - G_M C$ (red). Exact closure is the commuting square; closure residue is the failure. *Centre:* Three-scale composition identity $\Delta_{C02} = C_{12}\Delta_{C01} + \Delta_{C12}C_{01}$ (Result 3). The cross term shows residue is transported by the next coarse-graining, not accumulated as additive error. Numerical verification: residual 3.93×10^{-17}. *Right:* In the infinitesimal limit the finite residue Δ_C yields the residue density ω_s (Eq. 25, Section VI). Off-diagonal ω_s drives spectral-frame rotation; diagonal ω_s renormalises generator/potential data. Full derivation in Appendix F.

Why fast-mode compression fails

Clustering antecedent states by the *fastest* eigenvectors of G_μ captures within-scale relaxation rather than the slow variables governing long-time behaviour, maximising short-time descriptive power at the cost of long-time predictive failure—precisely what $J(C)$ penalises. Slow-eigenvector clustering is better motivated but fails when hidden couplings contaminate the slow spectrum.

Residue anatomy

When $J(C) > 0$, the *weighted column profile*

$$\sigma_j = \sqrt{\pi_j \sum_i \Delta^2_{C,ij}} \tag{19}$$

assigns each antecedent state its contribution to the residue, weighted by stationary probability. A state with anomalously large σ_j is a *hidden coupling state* whose transitions the current representation cannot predict. Promoting it into the coarse description converts hidden structure into visible generator content; a flat, structureless σ_j profile signals an adequate representation.

Demonstration

We test the procedure on a designed 12-state continuous-time Markov chain (CTMC) with three hub-and-spoke blocks, each with identical internal structure, connected in a chain by slow hub-to-hub rates ($r = 0.4$) (Fig. 8). A hidden shortcut between spoke states $1 \in A$ and $9 \in C$ (rate $r_h = 2.0$) raises the effective A–C coupling to $\approx$0.64, exceeding the inter-block rate. The shortcut is hidden because it connects non-hub states: the hub-to-hub topology suggests $\{A, B\} \mid \{C\}$ or $\{A\} \mid \{B, C\}$, missing the true dynamical similarity between A and C. Full specification and additional figures are in Appendix K.

Among all binary block partitions, $J(C)$ selects $\{A, C\} \mid \{B\}$ (loss 0.26) over the topologically natural alternatives $\{A, B\} \mid \{C\}$ (0.93) and $\{A\} \mid \{B, C\}$ (0.90)—a factor of 3.6× lower. Fast-mode compression yields loss 30.9, two orders of magnitude worse. Mean L^1 prediction error at $t = 5$ is 6×10^{-4} for $\{A, C\} \mid \{B\}$ versus 0.044 for the topology-guided partitions: a factor of 72× (Fig. 9a), confirming that $J(C)$ tracks predictive accuracy.

The column profile σ_j of the wrong partition $\{A, B\} \mid \{C\}$ spikes on states 1 and 9 by a factor of $\sim$6 above the next contributor (Fig. 9b)—precisely the shortcut endpoints, identified without prior knowledge of the shortcut. The analyst can then reason: states $1 \in A$ and $9 \in C$ carry a direct coupling, so A and C belong together. Recovering $\{A, C\} \mid \{B\}$ from the residue demonstrates that the profile is an instruction to revise the partition, not an irreducible noise floor.

The transport identity $\Delta_{C02} = C_{12}\Delta_{C01} + \Delta_{C12}C_{01}$ (Result 3) holds to 2.8×10^{-15} across the three-scale tower $12 \to 3 \to 2$, verifying exact algebraic consistency. Connections to Koopman-based methods, Mori–Zwanzig projection, and spectral clustering are discussed in Appendix K.

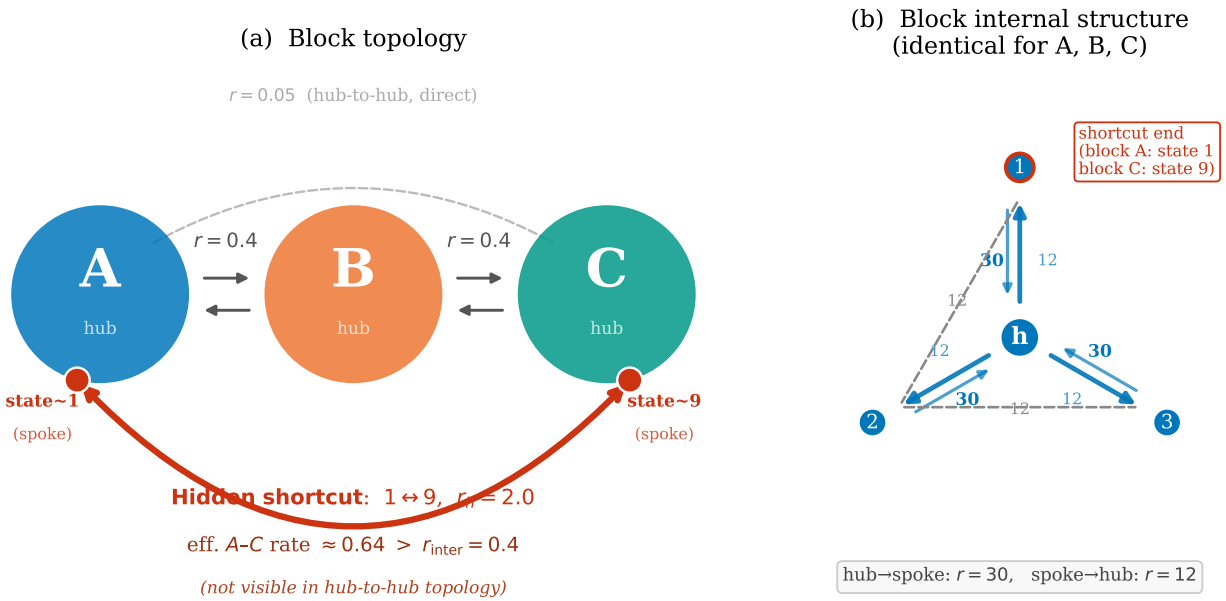

FIG. 8. **Markov chain geometry. (a)** Block topology: three blocks connected by slow hub-to-hub rates ($r = 0.4$, grey). A hidden shortcut (red arc, $r_h = 2.0$) connects spoke states $1 \in A$ and $9 \in C$, raising the effective A–C coupling to $\approx$0.64. The shortcut is hidden because it runs through non-hub states and is invisible to hub-level topology analysis. **(b)** Representative block structure (all three blocks are identical): one hub node connected to three spokes in a star-and-chain arrangement, with fast intra-block rates ($r = 30$ hub→spoke, $r = 12$ spoke→hub and spoke-spoke). Spoke 1 (red outline) is the shortcut endpoint.

V. EQUILIBRIUM, EMERGENCE, AND CONSERVED STRUCTURE

Equilibrium is the state in which both branches of the triage (11) close without residue. Symmetry and conservation determine the algebra of what the null-potential branch inherits across scale transitions.

Equilibrium

A state is stationary when $G_s\gamma_s = 0$; a driven steady state satisfies this condition while maintaining itself through hidden fluxes that would be visible at a finer description [30, 31].

Equilibrium requires in addition that the stationary law closes:

$$G_s\gamma_s = 0 \qquad \text{and} \qquad \Delta_C\,\gamma_s = 0. \tag{20}$$

Equilibrium is the closed sector of finite description: the part that reproduces itself without leaving closure residue under the relevant scale change. Nonequilibrium is persistent closure residue.

The triage identifies four structurally distinct regimes of partition-dynamics commutation. Equilibrium is the commuting case: partition and dynamics agree on the stationary law, and the order of operations leaves no trace. Effective physics is structured non-commutation: residue is nonzero, coherent, and classifiable by the modes of Section III, appearing as renormalisation-group flow, dissipation, memory kernels, emergent order parameters, gauge freedom, and decoherence. Quantum mechanics is self-similar non-commutation: partition-dynamics obstruction stabilises at a non-trivial spectral fixed point rather than flowing to zero (Section VI). At that fixed point, the canonical commutator $[Q, P] = i\hbar$ is the algebraic signature of the stabilised obstruction.

Entropy measures the hidden structure that can be absorbed into the null law q_s without leaving closure residue: when the entropy of the hidden conditional is the only contribution to the change in null potential, the system closes exactly. Unabsorbable structure—memory, correlated residues, coherent modes—leaves $\Delta_C \neq 0$. Details in Appendix E.

Symmetry and emergence

To coarse-grain is to choose a symmetry structure: to decide which distinctions are gauge-like from the coarser point of view and which remain physical.

A fine symmetry g_μ transports to a coarse symmetry g_M only if

$$Cg_\mu = g_M C. \tag{21}$$

Three cases arise: the symmetry transports and remains a symmetry; the transformation satisfies $Cg_\mu = C$ and becomes a gauge redundancy at the coarse level; or no

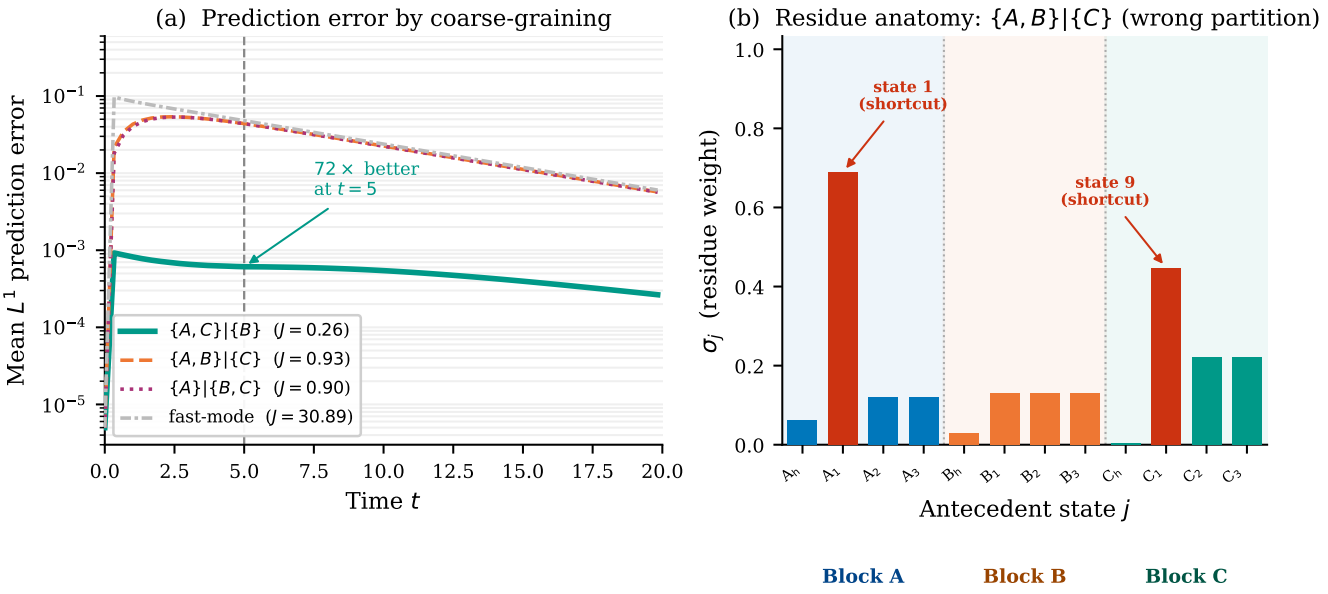

FIG. 9. **Closure results. (a)** Mean L^1 prediction error vs time (log scale) for key partitions; legend gives closure loss $J(C)$ for each. $\{A, C\} \mid \{B\}$ (teal) achieves loss 0.26 and is 72× more accurate at $t = 5$ than the topology-guided alternatives. Fast-mode compression ($J = 30.9$, grey) is two orders of magnitude worse. **(b)** Weighted residue column profile σ_j for the wrong partition $\{A, B\} \mid \{C\}$. States 1 and 9 (red bars) spike by ∼6× above all other contributors, identifying the hidden-shortcut endpoints without prior knowledge of the shortcut; the profile instructs the revision to $\{A, C\} \mid \{B\}$.

g_M closes the diagram. The third case is *emergence*: the residue of the failed symmetry closure appears as new structure at the coarser scale. An order parameter is the coarse-scale observable whose existence depends on that failed lift; it has no antecedent precursor because the commuting square does not close. Emergence is the specific algebraic failure of the commuting square, measured by the closure residue of the symmetry [21, 22].

The dual closure residue $\Delta_C{}^\dagger = G_\mu^\dagger C^* - C^* G_M^\dagger$ is the obstruction to lifting coarse conservation laws to the antecedent level; a macro conservation law is inherited when $\Delta_C{}^\dagger I_M = 0$ and emergent or approximate otherwise. Full derivations and the connection to integrable systems are in Appendix E.

VI. INFINITESIMAL RESIDUE AND QUANTUM-LIKE CLOSURE

The spectral residue mode G2 of Section III—the failure of incompatible retained projectors to admit joint probability assignment—is the regime where order independence fails globally. This section takes the infinitesimal scale limit of the closure residue and asks what fixed-point behaviour the resulting residue density ω_s can exhibit.

Let the coarse-graining between adjacent scales be

$$C_{s\to s+\delta s} = I + \delta s\, L_s + O(\delta s^2), \tag{22}$$

and let the generator vary as

$$G_{s+\delta s} = G_s + \delta s\, \dot{G}_s + O(\delta s^2). \tag{23}$$

A direct expansion of the closure residue gives

$$\Delta_C{}^{s\to s+\delta s} = \delta s\big([L_s, G_s] - \dot{G}_s\big) + O(\delta s^2). \tag{24}$$

The raw residue vanishes with δs, but its ratio to δs remains finite. Define the *residue density*

$$\omega_s = [L_s, G_s] - \dot{G}_s. \tag{25}$$

The fixed-point object is ω_s, not the raw finite residue.

For a non-degenerate spectrum $G_s = \sum_i \lambda_i \Pi_i$, the off-diagonal residue drives spectral-frame rotation:

$$\Pi_j (D_s \Pi_i) \Pi_i = -\frac{\Pi_j\, \omega_s\, \Pi_i}{\lambda_i - \lambda_j}, \tag{26}$$

where $\mathrm{ad}(L_s)$ denotes the commutator action $\mathrm{ad}(L_s)[\,\cdot\,] = [L_s, \cdot]$, and $D_s = \partial_s - \mathrm{ad}(L_s)$ is the covariant scale-derivative. Diagonal residue $(\omega_s)_{ii}$ renormalises eigenvalues and can be absorbed into K_s or $V_s^{(0)}$. Off-diagonal residue $(\omega_s)_{ij}$, $i \neq j$, rotates spectral projectors and cannot be absorbed into any scalar potential.

For the two-level case $G_s = \frac{\Delta}{2}\sigma_z$, $\omega_s = \kappa\,\sigma_x$, one obtains

$$\lim_{\delta s\to 0} \frac{[\Pi_+(s),\, \Pi_+(s+\delta s)]}{\delta s} = -i\,\frac{\kappa}{\Delta}\,\sigma_y. \tag{27}$$

Derivation in Appendix F. The commutator density is finite even as the raw commutator vanishes.

Define the dimensionless spectral residue ratio

$$\chi_s = \frac{\text{off-diagonal residue density}}{\text{spectral gap}} = \frac{\kappa}{\Delta} \quad \text{(two-level case)}. \tag{28}$$

Two regimes follow from the scale evolution of χ_s:

$$\chi_s \to 0: \quad \text{classical closure}. \tag{29}$$

Spectral frames become asymptotically compatible; a single joint distribution assigns probabilities to all contexts simultaneously.

$$\chi_s \to \chi^* \neq 0: \quad \text{self-similar spectral residue}. \tag{30}$$

No scale transition can be made invisible; the spectral-frame rotation rate is finite and scale-stable.

Any description stabilising at a non-zero spectral fixed point $\chi^* \neq 0$ must have the algebraic structure that quantum theory has. At this fixed point the canonical form of the Markovian generator is the GKSL equation [32, 33] (Appendix G): the Hamiltonian commutator is the coherent generator and the Lindblad terms are the residual incoherent modes. No-go theorems [29, 34, 35] are obstructions to making all reconstruction choices jointly flat—they are the statement that spectral residue cannot be globally gauged away, which is the content of partition-dynamics non-commutation at this fixed point. Decoherence [36, 37] is the decay of off-diagonal closure residue toward the classical sector as $\chi_s \to 0$. In the stable quantum regime, the operational residue scale is structurally identified with the universal constant:

$$\hbar_{\text{op}}(s, \rho, C, R) = \hbar\, I. \tag{31}$$

The identification places quantum mechanics at this fixed point; the universality of $\hbar$ reflects the rigidity of the fixed-point structure.

The canonical commutator $[Q, P] = i\hbar$ is the algebraic signature of this fixed point: the Heisenberg algebra—the unique algebra with a central, universal, scalar generator—is the internal spectral readout of partition-dynamics obstruction after stabilisation. Dirac's identification of Poisson brackets with commutators [38] and the deformation-quantisation programme [39, 40] pursue the same structural identification from the classical side; the partition-dynamics fixed-point framing locates the structural reason.

The connection to the KL axioms of Section II is direct: order independence—the condition that partition and accounting commute across refinements—is precisely the requirement that fails at $\chi_s \to \chi^* \neq 0$. Bell's theorem [34] and the Kochen-Specker theorem [29, 35] state that this failure cannot be made globally invisible by any relabelling of contexts: spectral residue cannot be gauged away. Extended treatment of self-similar residue dynamics, generalised quantum-like regimes, and the scope of this identification are in Appendix G.

VII. NEURAL SCALING LAWS AS CLOSURE-RESIDUE DECAY

The closure-residue branch of the triage has a direct empirical target: modern neural network training, where the statistical law, the representation scale, and the residue are all measurable.

Consider a data distribution p_{data} and a parametrised model distribution p_θ. Neural network training loss curves follow empirical power laws [41, 42] whose exponents are stable across model families and modalities. The standard training loss is the cross-entropy,

$$\mathcal{L}(\theta) = -\mathbb{E}_{x\sim p_{\text{data}}} \log p_\theta(x). \tag{32}$$

This decomposes into the irreducible entropy $H(p_{\text{data}})$ of the data source and a KL divergence,

$$\mathcal{L}(\theta) - H(p_{\text{data}}) = D_{\text{KL}}(p_{\text{data}} \| p_\theta). \tag{33}$$

The irreducible entropy is the null statistical content of the data source, fixed independently of the model. The excess loss is the scalar representational residue: the part of the data law not generated by the model family.

Scaling model size, data, and compute changes the admissible representation. Let $s = \log N$, where N is a representation-scale variable such as parameter count. If the representation flow is locally self-similar, residue mode k evolves as $a_k(s) \sim e^{-\lambda_k s}$, giving $a_k(N) \sim N^{-\lambda_k}$. The slowest-decaying mode dominates,

$$\mathcal{L}(N) - \mathcal{L}_\infty \sim AN^{-\alpha}, \tag{34}$$

where $\mathcal{L}_\infty = \lim_{N\to\infty} \mathcal{L}(N)$ is the irreducible loss, $A > 0$ is a normalisation amplitude, and $\alpha > 0$ is the dominant residue eigenvalue (scaling exponent). Power-law scaling is the signature of self-similar residue decay, not an empirical coincidence [43, 44].

There are at least two leading residues: $\Delta_{\text{cap}}(N)$, the capacity residue controlled by model size; and $\Delta_{\text{samp}}(D)$, the sampling residue controlled by dataset size. Therefore,

$$\mathcal{L}(N, D) - \mathcal{L}_\infty \approx AN^{-\alpha} + BD^{-\beta} \tag{35}$$

plus interaction terms, where $B > 0$ and $\beta > 0$ are the analogous amplitude and exponent for the data-size residue. Under a compute constraint $C \sim ND$, compute-optimal scaling balances the marginal reduction of each residue: neither mode should dominate.

The smooth power law is the self-similar decay of residue modes with no residue-class transitions. Deviations from smoothness mark residue changing class: induction heads [45] are closure events (history-dependent residue becomes an internal transition operator); superposition [46] is coherent unresolved residue compressed into too few visible dimensions. These discrete structural transitions—documented as "emergent abilities" [47]—are predicted to correspond to knees or exponent changes in the residual $R(N, D)$.

Defining a smooth scaling baseline and extracting the residual,

$$R(N, D) = \mathcal{L}_{\text{obs}}(N, D) - \mathcal{L}_{\text{fit}}(N, D), \tag{36}$$

where $\mathcal{L}_{\text{obs}}$ is the empirically measured loss and $\mathcal{L}_{\text{fit}}$ is a smooth power-law baseline fitted to the self-similar scaling regime, the framework predicts that circuit formation, feature splitting, induction-head emergence, and superposition reorganisations should correlate with structured features of $R(N, D)$—extrema, knees, or exponent changes—rather than occurring randomly along the scaling curve. Neural networks are not special physical systems; they are unusually well-instrumented finite theories in which the dynamics of representation is directly measurable. (Appendix H gives the full treatment.)

VIII. DISCUSSION

Synthetics names the common problem that these constructions share: how lawful structure is produced, transformed, and partially lost when a description is changed. The projection-operator formalism, the renormalisation group, effective field theory, information geometry, quantum reconstruction programmes, and contextuality cohomology each engage this problem; they differ in entry point and vocabulary, not in the structural question. The synthetic decomposition $S_C(G_\mu) = (q_s, K_s, \Delta_C)$ is one operational framework inside that discipline: a representation absorbs some antecedent structure as null law, retains some as visible cost, and leaves the rest as closure residue whose composition identity (15) ensures it is transportable rather than arbitrary.

Every physical variable is a coarse image of phenomena richer than itself; every equation governs one level in a hierarchy that extends above and below. This is the normal condition for physics, and the consistency conditions it imposes on law-forms are operative in every equation in use. Their answer is the forced triage (11): what closes exactly becomes the null law; what remains visible is priced by K_s; what cannot be absorbed or retained is the closure residue. The KL functional is the invariant content across every change of partition; the free-energy functional is its expression once visible constraints are added.

The modes of that residue are the modes of effective physics (Proposition 6, Appendix C). Renormalisation-group flow is absorbable scalar residue [11, 20, 48]: the running of couplings is the scale-by-scale absorption of antecedent structure into the null potential. Memory kernels are nonlocal residue [9, 10, 17]: the past of the hidden degrees of freedom is not gone, only unrepresented. Emergent order parameters are coherent unresolved residue [21, 22]: structure that survives coarse-graining with internal phase relations intact. Gauge freedom is reconstruction residue [26, 27]: the ambiguity in lifting coarse states back through the fibre. Entropy is the scalar bookkeeping of structure that closes exactly [1, 49]. These are the algebra of what closure failure can look like, exhausted by the triage condition and the coherent composition law (15).

Four levels of partition-dynamics commutation organise the framework's range. Equilibrium is exact commutation: partition and dynamics agree on the stationary law, residue is zero, and the order of operations leaves no trace. Effective physics is structured non-commutation: residue is nonzero, coherent, and classifiable by the modes of Section III. Quantum mechanics is self-similar non-commutation: the partition-dynamics obstruction stabilises at a non-trivial spectral fixed point rather than flowing to zero. The canonical commutator $[Q, P] = i\hbar$ is the algebraic signature of that fixed point—the form taken by the residue when it has become universal, scalar, and central. One structural demand produces two regimes: where order independence holds, the KL form is forced; where it fails globally, noncommutativity is forced. These are two limiting behaviours of the same partition-dynamics consistency condition, not separate axiom systems.

The reflexive character of the framework carries the sharpest consequence. A theory of Ω is a finite register $R \subset \Omega$, subject to the same triage as everything it describes. When the consistency conditions close the loop on the observer—when the act of distinguishing outcomes is itself a coarse-graining subject to the same calculus—self-consistent closure requires residue. The observer has a finite partition; its distinguishable elements are exactly that, and no finer; no finite description absorbs the consequences of applying its own coarse-graining to itself without remainder. Quantum mechanics is the case in which this loop has a stable, rigid closure (Appendix G). Its universality and rigidity are evidence that the loss is structured. The Hilbert-space formalism is the algebraic structure of self-consistent description at a non-trivial spectral fixed point—a structural identification rather than a derivation from coarse-graining alone. The Born rule, complex amplitudes, and the specific algebra of observables are the contents of that fixed point, addressed by the full theory. The relational [50] and Bayesian [51] traditions in quantum foundations pursue the same structural claim from different starting assumptions; the present framework supplies a derivation of the structural claim itself.

Thermodynamics, rate-distortion theory [3], the information bottleneck [5], Friston's free-energy principle [16, 18], and Rissanen's minimum description length [6] are all instances of Result 2 under different choices of visible generator K_s (Fig. 5). They share a null law, a KL cost, and a forced law-form because they are all variational descriptions of finite representations under order independence. That apparently unrelated frameworks are structurally identical under different choices of retained observable is itself a headline result: the apparent multiplicity of variational principles in physics, biology, and information theory collapses to a single forced structure. Watanabe's real log-canonical threshold [43, 44]

is the dominant residue-mode eigenvalue in this framework; neural scaling curves are self-similar residue decay (Section VII).

A finite reflexive description cannot eliminate all residue under self-application—this follows directly from the triage. The remaining open problems are the next residues to be classified: uniqueness theorems need extension to continuous settings requiring additional regularity [52, 53]; the non-convex null-law case—symmetry breaking, phase coexistence [54]—requires a branch-selection principle; the spectral fixed-point characterisation leaves open what determines the algebra of distinguishable events at that fixed point [28, 32, 33, 55]; and the prediction that residue-class transitions correlate with structured features of the loss curve [41, 42, 47] is a falsifiable experimental claim awaiting controlled perturbation studies.

Appendix A: KL and free-energy derivations

This appendix proves Results 1 and 2 and establishes contraction closure of the free-energy family.

1. Local ratio form

Proposition 1 (Local ratio form)**.** *For finite positive distributions p, q satisfying relabelling invariance and atomwise locality, the scalar law-content has the form*

$$\mathcal{I}(p\|q) = \sum_i p_i \, \varphi\left(\frac{p_i}{q_i}\right) \tag{A1}$$

for a continuous function φ on $\mathbb{R}_+$.

Proof. Atomwise locality gives $\mathcal{I}(p\|q) = \sum_i \ell(p_i, q_i)$ with a common density ℓ. For fixed ratio $r = a/b$ define $g_r(t) = \ell(rt, t)$. Neutral refinement: splitting a cell of q-mass $t_1 + t_2$ into two neutral subcells with q-masses t_1, t_2 and the same ratio r gives

$$g_r(t_1 + t_2) = g_r(t_1) + g_r(t_2).$$

Continuity forces $g_r(t) = t\, g_r(1)$, so $\ell(a, b) = a\, \varphi(a/b)$ with $\varphi(r) = g_r(1)/r$. □

2. KL uniqueness under staged elimination

Proposition 2 (KL uniqueness)**.** *For finite positive distributions satisfying the local-ratio form of Proposition 1, the staged-elimination chain rule* (8) *forces*

$$\mathcal{I}(p\|q) = C \sum_i p_i \log \frac{p_i}{q_i}. \tag{A2}$$

Proof. Refine coarse state i into conditional alternatives j: set $p_{ij} = p_i r_{j|i}$, $q_{ij} = q_i s_{j|i}$, $a_i = p_i/q_i$, $b_{j|i} = r_{j|i}/s_{j|i}$. The direct fine content is $\sum_{ij} p_i r_{j|i} \, \varphi(a_i b_{j|i})$. The chain rule requires this to equal

$$\sum_i p_i \, \varphi(a_i) + \sum_i p_i \sum_j r_{j|i} \, \varphi(b_{j|i})$$

for all positive refinements.

Step 1: $\varphi(1) = 0$. Apply neutral refinement: set $b_{j|i} = 1$ for all j, i (equal likelihood ratios in every conditional cell). Then $a_i b_{j|i} = a_i$, so both sides reduce to $\sum_i p_i \varphi(a_i)$, which forces $\sum_i p_i \sum_j r_{j|i} \, \varphi(1) = 0$ for every admissible p, r. Choosing any nondegenerate p, r gives $\varphi(1) = 0$.

Step 2: Isolate arbitrary a, b. Fix two arbitrary positive reals a, b. Choose a coarse description with a single nontrivial state $i = 1$ with ratio $a_1 = a$. Refine state 1 into two sub-states $j = 1, 2$ with conditional weights $r_{1|1} = \epsilon$, $r_{2|1} = 1 - \epsilon$ for small $\epsilon > 0$, and conditional ratios $b_{1|1} = b$, $b_{2|1} = 1$. All other coarse states are refined neutrally.

Step 3: Apply chain rule and solve. The chain rule equation for this refinement reads, after cancelling neutral terms,

$$\epsilon \, \varphi(ab) + (1 - \epsilon) \, \varphi(a) = \varphi(a) + \epsilon \, \varphi(b) + (1 - \epsilon) \, \varphi(1).$$

Using $\varphi(1) = 0$ and cancelling $(1 - \epsilon)\varphi(a)$ from both sides:

$$\epsilon \, \varphi(ab) = \epsilon \, \varphi(a) + \epsilon \, \varphi(b).$$

Dividing by $\epsilon > 0$ gives

$$\varphi(ab) = \varphi(a) + \varphi(b) \quad \text{for all } a, b \in \mathbb{R}_+.$$

Since a and b were arbitrary, this holds on all of $\mathbb{R}_+$. Continuity forces $\varphi(a) = C \log a$. Positivity forces $C \geq 0$; nontriviality fixes $C > 0$. □

The results align with the Csiszár characterisation of f-divergences [52] and the functorial characterisations of [53, 56]. The setting here is strictly finite and positive-support; extensions to measurable spaces require domain and regularity conditions beyond the scope of this paper.

3. Free-energy law-form

Proposition 3 (Free-energy law-form)**.** *For finite states with compatible support and $M > 0$, the unique staged-elimination law-content combined with visible constraints $\langle A_i \rangle_p = a_i$ gives the variational family*

$$\mathcal{F}_s[p] = \langle K \rangle_p + M D_{\mathrm{KL}}(p\|q) \tag{A3}$$

with generator $K = \sum_i g_i A_i$ and minimiser $p^*(x) \propto q(x) \exp[-K(x)/M]$.

Proof. With KL law-content and Lagrange multipliers g_i and α, the Euler equation gives $M(\log p - \log q) + K + \alpha = 0$, hence $p^* \propto q \exp[-K/M]$. Substituting back gives minimum free energy $\mathcal{F}^* = -M \log Z$ with $Z = \sum_x q(x) e^{-K(x)/M}$. □

4. Contraction closure

Proposition 4 (Contraction closure)**.** *For a finite coarse map $\pi : X \to Y$ with decomposition $q_X(x) = q_Y(y)\, q_{X|Y}(x|y)$, the fibre contraction of the free-energy functional is again a free-energy functional:*

$$\inf_{\pi_* p_X = p_Y} \mathcal{F}_X[p_X] = \sum_y p_Y(y) \, E_{\mathrm{eff}}(y) + M D_{\mathrm{KL}}(p_Y\|q_Y), \tag{A4}$$

where

$$E_{\mathrm{eff}}(y) = -M \log \sum_{x \in X_y} q_{X|Y}(x|y) \, e^{-E_X(x)/M}. \tag{A5}$$

Proof. Apply the KL chain rule to decouple the conditional minimisation over fibres. At fixed p_Y, minimising over $p(\cdot|y)$ for each y gives the Gibbs conditional

$$p^*(x|y) = \frac{q_{X|Y}(x|y)\, e^{-E_X(x)/M}}{\sum_{x' \in X_y} q_{X|Y}(x'|y)\, e^{-E_X(x')/M}}.$$

Substituting gives the effective energy $E_{\text{eff}}(y)$ as stated, and the remaining p_Y-minimisation is exactly a free-energy problem with generator E_{eff}. □

Contraction closure is the precise content of vertical covariance of law-form: eliminating hidden refinements does not change the *kind* of law, only the generator and its coordinates. Ordinary renormalisation-group coupling flow is the induced flow of K within this invariant free-energy family (Appendix D).

5. Rate-distortion and information bottleneck as special cases

The free-energy family (10) subsumes two foundational information-theoretic optimisation frameworks.

a. Rate-distortion. Shannon's rate-distortion problem [3] asks for the minimum mutual information $I(X;\hat{X})$ achievable over encoders $p(\hat{x}|x)$ subject to a distortion constraint $\mathbb{E}[d(X,\hat{X})] \le D$. The Lagrangian over encoders is

$$\mathcal{L}_{\text{RD}}[p(\hat{x}|x)] = I(X;\hat{X}) + \beta\, \mathbb{E}[d(X,\hat{X})], \tag{A6}$$

which expands as $\sum_x p(x)\big[D_{\text{KL}}(p(\hat{x}|x)\|p(\hat{x})) + \beta\langle d(x,\hat{x})\rangle_{p(\hat{x}|x)}\big]$. This is exactly $\mathcal{F}_s[p(\hat{x}|x)]$ with $q_s = p(\hat{x})$ (the encoder marginal serving as the null law) and $K = \beta d(x,\hat{x})$ as the visible cost function. Rate-distortion theory is the free-energy law-form with the distortion measure as visible generator and the marginal as null law.

b. Information bottleneck. The information bottleneck [5] compresses X to a representation T while preserving information about a relevant variable Y. Its Lagrangian is

$$\mathcal{L}_{\text{IB}}[p(t|x)] = I(T;X) - \beta\, I(T;Y). \tag{A7}$$

The first term $I(T;X) = \sum_x p(x) D_{\text{KL}}(p(t|x)\|p(t))$ is again $M\, D_{\text{KL}}(p\|q_s)$ with $q_s = p(t)$ as null law and $M = 1$. The second term $-\beta I(T;Y)$ imposes a visible constraint: retain information about Y. Writing the mutual information as an expectation of log-likelihood ratios, $-\beta I(T;Y) = \beta \sum_{t,y} p(t,y)[\log p(y) - \log p(y|t)]$, the constraint enters as a visible generator $K_{\text{IB}}(t) = -\beta \sum_y p(y|t) \log p(y|t) + \text{const}$: the visible cost of the representation is its reduction of uncertainty about Y. The IB is the free-energy law-form with information relevance as visible generator and the encoder marginal as null law.

Both frameworks inherit the chain-rule requirement that forced KL in the first place; neither is a modelling choice grafted onto the free-energy functional. They differ only in the choice of visible generator K_s. The null law in both cases is the marginal produced by the encoder, which is precisely the pushforward of the source law through the coarse-graining—the recursively generated null law of Definition 1.

Appendix B: Null laws and kinetic/potential split

1. Recursive null-law projection

The null law at each scale is defined recursively by the coarse-graining operation itself. Let $\mathcal{N}_s$ denote the admissible class at scale s: the set of distributions compatible with the partition and its defining constraints, and no others.

Definition 1 (Null law)**.** *The null law q_s at scale s is*

$$q_s = \arg\min_{q \in \mathcal{N}_s} D_{\text{KL}}(C_* q_{s-ds} \| q)\,, \tag{B1}$$

where $C_* q_{s-ds}$ is the push-forward of the antecedent null law through the coarse-graining C.

At a self-similar fixed point $q^* = \mathcal{P}(C_* q^*)$, where $\mathcal{P}$ denotes projection onto $\mathcal{N}_s$, the null law is self-consistently the baseline produced by the scale's own coarse-graining. The null law at the finest scale (the ground-level description) is the uniform distribution over the primitive alternatives, which is the only distribution in $\mathcal{N}$ when no constraints are available.

The null law is not a prior in the Bayesian sense. It is not chosen before evidence or placed externally by an agent. It is the output of the same projection operation that defines the scale.

2. Inherited null potential

Write the null law in exponential form:

$$q_s(x) = \frac{1}{Z_s^{(0)}} \exp\Big[-V_s^{(0)}(x)/M\Big]\,. \tag{B2}$$

The *inherited null potential* $V_s^{(0)} = -M \log q_s$ encodes the accumulated effect of all antecedent laws that have been successfully absorbed into the current-scale background through repeated projection.

At each new scale:

1. The antecedent null law q_{s-ds} is pushed forward to $\hat{q}_s = C_* q_{s-ds}$.
2. $\hat{q}_s$ is projected onto $\mathcal{N}_s$ to give q_s.
3. $V_s^{(0)} = -M \log q_s$ is the null potential at the new scale.

Any antecedent structure that cannot be represented within $\mathcal{N}_s$ after this projection contributes to the closure residue Δ_s, not to the null potential.

The compact identity is:

$$\Delta V_s^{(0)} = -M \log Z_{\text{conditional}} = \langle K_{\text{hidden}} \rangle + M H_{\text{conditional}}, \tag{B3}$$

where $H_{\text{conditional}}$ is the Shannon entropy of the hidden conditional. Entropy is the scalar bookkeeping of structure that *can* be absorbed into the null potential without closure residue.

3. Path formulation and the kinetic/potential split

The kinetic-potential split follows from applying the free-energy construction to path-space rather than instantaneous state-space.

For a path $x(t)$ over the interval $[0, T]$, the effective action selecting physical trajectories decomposes as

$$\mathcal{A}_s[\text{path}] = T_s[\text{path}] + V_s[\text{path}], \tag{B4}$$

where:

- $T_s[\text{path}]$ is the transport cost of the trajectory given the admissible null dynamics: the kinetic contribution from allowed transitions of the represented variables.
- $V_s[\text{path}]$ is the occupancy cost relative to the null law: the integrated KL weight of the trajectory against the scale-s background.

Both contributions arise from the same free-energy construction applied to measures over paths. The split is not assumed as a primitive mechanical dichotomy ("kinetic" and "potential" energy as independent inputs). It is induced by self-consistent coarse-graining.

Kinetic component. The null dynamics determines which transitions between represented states are admissible and at what rate. The transport cost T_s measures deviation from those admissible null transitions. In a Hamiltonian system this recovers the familiar kinetic energy (cost of motion given the metric); in an open or stochastic system it gives the appropriate path-space transport term.

Potential component. The total effective potential is

$$V_s^{\text{eff}}(x) = V_s^{(0)}(x) + K_s(x), \tag{B5}$$

with inherited null potential $V_s^{(0)}$ and visible deformation K_s. Effective potentials produced by RG integration of hidden degrees of freedom are inherited null-potential components, not primitive energies.

Temperature. The scale M relates the two components: it is the ratio of potential to kinetic cost in the free-energy functional. In the standard thermodynamic context, $M = k_B T$. In more general settings, M is the scale parameter of the law-content.

4. Connection to the free-energy principle and MDL

Free Energy Principle. Friston's free-energy principle [16, 18] proposes that adaptive biological systems minimise a variational free energy $\mathcal{F} = \mathbb{E}_q[\log q(\mathbf{s}) - \log p(\mathbf{s}, \mu)]$, where $\mathbf{s}$ are hidden causes, μ are sensory signals, and q is the agent's internal model. This equals $D_{\text{KL}}(q \| p(\cdot|\mu)) - \log p(\mu)$: precisely the forced law-form (10) with $q_s = p(\cdot|\mu)$ as null law and $K = -\log p(\mu)$ as visible generator. The FEP is the free-energy law-form applied to the agent's own Markov blanket [16], which is a Markov blanket partition of the state space—a special case of the closure condition (2) restricted to the agent's sensory-motor boundary. The present derivation provides the axiomatic grounding the FEP has sought: the variational free energy is not a modelling choice but the unique scalar comparison forced by staged elimination.

Minimum description length. Rissanen's minimum description length principle [6] selects models by the code length of the model plus the code length of the data given the model. In the closure-residue framework, the code length of the model is K_s (the cost of retained complexity above the null background) and the code length of the data given the model is $M\, D_{\text{KL}}(p \| \gamma_s)$ (the residual complexity relative to the model's distribution γ_s). The MDL-optimal model is therefore the minimiser of the forced law-functional (10): MDL is staged-elimination inference. The "minimum description length" is the minimum cost of retained complexity, which is exactly what $\mathcal{F}_s$ measures.

Appendix C: Closure residue algebra

1. Composition proof

Proposition 5 (Closure-residue composition)**.** *For the scale tower $X_0 \xrightarrow{C_{01}} X_1 \xrightarrow{C_{12}} X_2$ with generators G_0, G_1, G_2, the two-step closure residue satisfies*

$$\Delta_{C02} = C_{12}\, \Delta_{C01} + \Delta_{C12}\, C_{01}. \tag{C1}$$

Proof. Expand $\Delta_{C02} = C_{12}C_{01}G_0 - G_2C_{12}C_{01}$ and add and subtract $C_{12}G_1C_{01}$:

$$\begin{aligned} \Delta_{C02} &= C_{12}(C_{01}G_0 - G_1C_{01}) + (C_{12}G_1 - G_2C_{12})C_{01} \\ &= C_{12}\, \Delta_{C01} + \Delta_{C12}\, C_{01}. \end{aligned}$$

□

The identity holds for linear generators on arbitrary vector spaces. For stochastic kernels it holds because they act linearly on signed measures. For differentiable nonlinear coarse maps and vector-field generators, the corresponding tangent-map identity is $\Delta_{C02}(x) = TC_{12} \cdot \Delta_{C01}(x) + \Delta_{C12}(C_{01}x)$. Numerical verification on 3×3 random matrices gives residual 3.93×10^{-17}.

Cocycle interpretation. The composition law (C1) has the structure of a 1-cocycle in the cochain complex of coarse-graining maps [26, 28]. Define the coboundary operator $(\delta R)_{02} = C_{12}R_{01} - R_{02} + R_{12}C_{01}$. The composition identity says that Δ_C is a closed 1-cochain: the residue distributes coherently over scale composition. This is why residue does not merely add as scalar noise: it is carried by the subsequent coarse-graining maps.

2. Canonical residue decomposition

Let $C\colon V \to W$ be a coarse-graining map and $C^+\colon W \to V$ a right-inverse, $CC^+ = I_W$. Define the retained projection $P = C^+C$ and hidden projection $Q = I_V - P$. Let G_μ and G_M be the fine and coarse generators.

Proposition 6 (Canonical residue decomposition)**.** *The closure residue $\Delta_C = CG_\mu - G_MC$ decomposes as*

$$\Delta_C = AC + CG_\mu Q, \tag{C2}$$

where $A = CG_\mu C^+ - G_M\colon W \to W$ is the retained mismatch *and $CG_\mu Q\colon V \to W$ is the* hidden coupling.

Proof. Substitute $I_V = C^+C + Q$:

$$CG_\mu = CG_\mu C^+C + CG_\mu Q.$$

Subtracting G_MC gives $\Delta_C = (CG_\mu C^+ - G_M)C + CG_\mu Q = AC + CG_\mu Q$. □

The decomposition is independent of the choice of C^+: any two right-inverses agree on $\operatorname{im} C$, and their difference lies in $\ker C$, which is annihilated by C in the first term and by Q in the second.

Retained modes. The map $A = CG_\mu C^+ - G_M$ acts entirely on the coarse space W. Its structure relative to the eigenbasis of G_M classifies the retained residue into three exhaustive types:

R1. Scalar retained residue: $A = \lambda I_W$. The shift $G_M \mapsto G_M + \lambda I_W$ absorbs the residue; no new structure is introduced. This is renormalisation-group coupling flow: the running of a coupling constant is precisely a scale-by-scale scalar absorption of retained residue into the null potential.

R2. Diagonal retained residue: A diagonal in the G_M-eigenbasis, non-scalar. The residue modifies the eigenvalue spectrum of G_M—altering transition rates between coarse modes—without coupling them. This is kinetic or frictional correction in open-system dynamics.

R3. Off-diagonal retained residue: A has off-diagonal entries in the G_M-eigenbasis. New couplings between distinct coarse modes are introduced; the retained structure acquires correlations not present in G_M. This is the appearance of an emergent order parameter or new field [21, 22].

Hidden modes. The term $CG_\mu Q$ couples the retained evolution to the hidden subspace. The exact Nakajima-Zwanzig identity [9, 17] applied to the splitting (P, Q) gives

$$\frac{d}{dt}Px(t) = PLPx(t) + \int_0^t PLe^{(t-s)QL}QL\,Px(s)\,ds + PLe^{tQL}Qx(0), \tag{C3}$$

where L is the full generator. The first term is the closed retained part; the remaining two are the two hidden modes:

H1. Memory kernel (nonlocal residue). The integral $\int_0^t K(t-s)\,Px(s)\,ds$ with $K(\tau) = PLe^{\tau QL}QL$ captures the systematic influence of the past retained state via the hidden degrees of freedom. No Markovian generator on W alone can reproduce this term; the coarse description is intrinsically non-Markovian.

H2. Initial-condition noise (incoherent residue). The term $PLe^{tQL}Qx(0)$ depends on the initial hidden state $Qx(0)$, which is uncontrolled relative to any preparation of $Px(0)$. This is the fluctuating force in the Langevin sense [24, 25].

Under scale separation ($\gamma \gg |\dot{u}/u|$) the memory kernel contracts to a scalar, converting H1 into R1; the explicit calculation is in Appendix D.

Geometric modes. Two further modes arise from the geometry of the fibration $C\colon V \to W$, independently of the generator structure.

G1. Reconstruction ambiguity (gauge residue). The right-inverse C^+ is not unique: any $\tilde{C}^+ = C^+ + \Phi$ with $C\Phi = 0$ is also a right-inverse. The field Φ defines a connection on the fibration $V \to W$; a non-flat connection produces holonomy. When a coarse path is lifted to the fine level, different choices of section yield different fine trajectories above the same coarse one. This is gauge freedom [26, 27].

G2. Spectral incompatibility (spectral residue). Let C_1, C_2 be two independent coarse-grainings of the same fine system with retained projectors $P_1 = C_1^+C_1$ and $P_2 = C_2^+C_2$. If $[P_1, P_2] \neq 0$, no classical probability distribution on V simultaneously represents both coarse statistics: joint representability fails [28, 29]. When the fine-level dynamics itself forces non-commuting retained projectors—as in the spectral fixed-point analysis of Section VI—this residue is unavoidable.

Corollary 1 (Exhaustiveness)**.** *Every mode of unabsorbed closure residue falls into one of the seven classes **R1**, **R2**, **R3** (retained), **H1**, **H2** (hidden), or **G1**, **G2** (geometric).*

The decomposition (C2) partitions Δ_C; R1–R3 exhaust the spectral classification of A; H1–H2 exhaust the exact Nakajima-Zwanzig splitting of the hidden component; G1–G2 exhaust the fibration geometry. No residue can fall outside these seven classes.

Appendix D: Worked examples

1. Ising decimation: RG as generator flow

Decimate the middle spin in a three-spin Ising block with generator $K(s_1, e, s_2) = -J(s_1 e + e s_2)$, retaining variables (s_1, s_2). The effective generator is given by contraction closure (Proposition 4):

$$K_{\text{eff}}(s_1, s_2) = -M \log\left[2\cosh\left(\frac{J(s_1 + s_2)}{M}\right)\right]. \tag{D1}$$

Projecting onto the retained basis $\{1, s_1 s_2\}$ gives the one-step coupling flow

$$J' = \frac{M}{2}\log\cosh\left(\frac{2J}{M}\right). \tag{D2}$$

For the full one-dimensional chain, repeated decimation gives

$$\tanh k' = \tanh^2 k, \qquad k = J/M, \tag{D3}$$

with all finite-temperature trajectories flowing to $k^* = 0$.

The free-energy law-form is preserved at every step. Only the coupling coordinate k changes. This confirms the contraction-closure property: RG is generator flow within the invariant free-energy family, not a modification of the family itself. The logarithm appears in (D1) for the same reason it appears in the law-content: it is the generator flow inside the invariant form.

In tanh coordinates $x = \tanh k$, the recursion reduces to the squaring map $x \mapsto x^2$: a linear (self-similar) map in log-coordinates, the same conformal scaling structure as the spiral in Fig. 4.

2. Two-mode open system: nonlocal closure residue

Let the fine state be (u, v) with u retained and v hidden. Take the generator

$$G_\mu = \begin{pmatrix} 0 & a \\ -b & -\gamma \end{pmatrix}, \qquad \gamma > 0. \tag{D4}$$

The coarse map is $C = (1, 0)$. For any scalar macro-generator G_M,

$$\Delta_C = CG_\mu - G_M C = (-G_M,\ a). \tag{D5}$$

No choice of scalar G_M eliminates the a component when $a \neq 0$: this is nonlocal closure residue (mode H1 of Proposition 6).

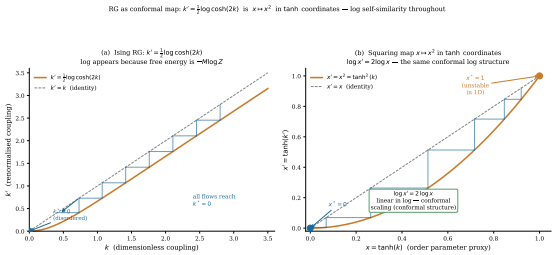

FIG. 10. **Ising RG as conformal map (appendix figure).** *Left:* Cobweb diagram for the 1D Ising recursion $k' = \frac{1}{2}\log\cosh(2k)$. *Right:* In tanh coordinates the recursion is the squaring map $x \mapsto x^2$, which is linear in log-coordinates. The logarithm appears because it is the generator flow inside the invariant free-energy form.

Solving for $v(t)$ and substituting into the u-equation gives

$$\dot{u}(t) = ae^{-\gamma t}v(0) - ab\int_0^t e^{-\gamma(t-s)}u(s)\,ds. \tag{D6}$$

The first term is transient noise from the hidden initial condition; the second is an exponential memory kernel. This is the Nakajima-Zwanzig structure [9, 17] at the level of the two-mode example.

In the scale-separated limit $\gamma \gg |\dot{u}/u|$, the kernel contracts:

$$\int_0^t e^{-\gamma(t-s)}u(s)\,ds \approx \frac{1}{\gamma}\,u(t), \tag{D7}$$

giving $\dot{u}(t) \approx -(ab/\gamma)\,u(t)$. The Markovian generator $G_M \approx -ab/\gamma$ is the scale-separated compression of the nonlocal residue into an absorbable scalar term. This illustrates mode 3 (nonlocal residue) transitioning to mode 1 (absorbable scalar) under scale separation: the microscopic mechanism behind nearly all effective dissipative dynamics.

The full Nakajima-Zwanzig identity is:

$$\frac{d}{dt}Px(t) = PLPx(t) + \int_0^t PLe^{(t-s)QL}QL\,Px(s)\,ds + PLe^{tQL}Qx(0), \tag{D8}$$

where P projects onto retained observables and $Q = 1 - P$. The first term is the closed retained generator; the integral is the memory kernel; the final term is noise from initial hidden-variable values. These correspond precisely to modes R1, H1, and H2 of the canonical decomposition (Proposition 6).

Appendix E: Equilibrium, symmetries, and conserved algebras

1. Stationarity versus equilibrium

The standard definition of equilibrium is stationarity: $G_s\gamma_s = 0$. This is necessary but not sufficient. A state

can be stationary within a model while maintaining itself through hidden fluxes, unresolved memory, or scale-specific cancellations that do not survive a change of description.

In the present framework, equilibrium requires additionally that the stationary law closes without residue:

$$G_s\gamma_s = 0 \quad \text{and} \quad \Delta_C\,\gamma_s = 0. \tag{E1}$$

A driven steady state is stationary ($G_s\gamma_s = 0$) but not in equilibrium because it requires a continuous flow of hidden variables (heat baths, driving forces, environmental gradients) that would be visible at a finer description; this flow appears as $\Delta_C\,\gamma_s \neq 0$ at that finer level.

The null law q_s is the equilibrium law of the scale-defining constraints: it is the minimally informative stationary law compatible with the current partition. The Gibbs law $\gamma_s \propto \exp[-(K_s + V_s^{(0)})/M]$ is the equilibrium of the full set of retained constraints. Both are equilibria; the difference is which constraints are enforced.

2. Entropy as absorbable structure

When hidden structure is eliminated, it either becomes part of the null potential $V_s^{(0)}$ (absorbed into the background, contributing to thermodynamic work terms) or leaves closure residue Δ_s.

The absorbed part is the entropy of the hidden conditional:

$$\Delta V_s^{(0)} = -M \log Z_{\text{conditional}} = \langle K_{\text{hidden}} \rangle + M H_{\text{conditional}}, \tag{E2}$$

where $H_{\text{conditional}}$ is the Shannon entropy of the hidden variables conditional on the retained ones.

Entropy is therefore not a measure of ignorance in an epistemological sense. It is the scalar bookkeeping of structure that can be absorbed into the null potential without closure residue. Entropy derivations from maximum-entropy and large-deviation principles [1, 8, 49] are special cases of this absorption construction. The thermodynamic interpretation of entropy as absorbed hidden structure aligns with stochastic-thermodynamic treatments of driven steady states [30, 31]: the irreversible work done on a driven system is precisely the rate at which the hidden flux generates closure residue that cannot be absorbed into a stationary null law.

3. Partition-relative symmetry

At a fixed partition, the theory has a family of horizontal symmetries: coordinate changes, basis rotations, canonical transformations. These preserve the set of distinctions the theory makes.

Changing the partition changes which symmetries exist. A coarse-graining C declares fine distinctions inside fibres to be invisible. Transformations g satisfying $Cg = C$ move within fibres and are gauge-like from the coarse point of view.

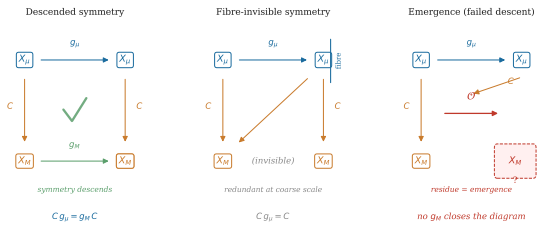

FIG. 11. **Transported symmetry, fibre-invisible symmetry, and failed closure (appendix figure).** *Left:* Fine symmetry satisfying $Cg_\mu = g_M C$; transported symmetry. *Centre:* Fine transformation satisfying $Cg_\mu = C$; fibre-invisible, becomes gauge redundancy. *Right:* No g_M closes the diagram; failed closure—residue is emergence.

Definition 2 (Symmetry transport)**.** *A fine symmetry g_μ is partition-covariant if there exists g_M with $Cg_\mu = g_M C$. It is gauge-like from the coarse point of view if $Cg_\mu = C$. Closure fails if no g_M closes the commuting square; the symmetry is fibre-invisible or produces emergence at the coarser level.*

The three cases produce:

1. $Cg_\mu = g_M C$ for some g_M: the symmetry is transported symmetry and remains a symmetry of the coarser theory.

2. $Cg_\mu = C$: the transformation is fibre-invisible, a gauge redundancy at the coarse level.

3. No g_M makes the diagram commute: failed closure. The closure residue of the symmetry is emergent structure at the coarser scale.

Emergence is the specific algebraic failure of the commuting square, not a vague appeal to novelty.

Symmetry breaking versus symmetry emergence. Symmetry *breaking* refers to a symmetric law whose selected state is not symmetric. Symmetry *emergence* (case 3 above) refers to a fine symmetry that fails to transport to the coarser level. Both are possible simultaneously.

4. Dual closure residue and conservation laws

The dual closure residue

$$\Delta_C{}^\dagger = G_\mu^\dagger C^* - C^* G_M^\dagger \tag{E3}$$

measures the failure of a macro integral of motion to lift to a true micro integral. For $G_M^\dagger I_M = 0$,

$$G_\mu^\dagger C^* I_M = \Delta_C{}^\dagger I_M. \tag{E4}$$

A macro conservation law $\partial_t\langle I_M\rangle = 0$ holds at the coarse level, but the lifted observable $C^* I_M$ is not conserved at the fine level unless ${\Delta_C}^\dagger I_M = 0$.

The set of integrals at scale s forms an algebra (closed under Poisson bracket or operator commutator). Changing scale changes which observables are available as integrals. Emergence is the appearance, disappearance, or incompatibility of conserved algebras under partition change: the closure residue of the transport of integrals of motion.

Integrable systems and MBL. In integrable systems, a large algebra of local integrals of motion survives arbitrary coarse-grainings in certain representations (the action variables of a classically integrable system) [54]. In many-body localised phases, approximate local integrals of motion are the effective conserved quantities at each scale of the disorder structure. The dual closure residue ${\Delta_C}^\dagger I_s$ measures the rate at which a local integral fails to remain conserved under vertical coarse-graining. Small $\|{\Delta_C}^\dagger I_s\|$ is the precise condition for approximate closure of a conserved quantity across scale. This connects to the causal-emergence framework [22]: a coarse description has higher causal power when its integrals of motion close more tightly under the coarse dynamics, i.e. when $\|{\Delta_C}^\dagger I_s\|$ is small.

Appendix F: Reconstruction, gauge, holonomy, and spectral residue

1. Affine reconstruction freedom

A coarse-graining $C : X_\mu \to X_M$ discards distinctions. Recovering them requires a reconstruction map $R : X_M \to X_\mu$ satisfying

$$CR = I_M. \tag{F1}$$

If R is one such map and $N : X_M \to X_\mu$ satisfies $CN = 0$ (N maps into the kernel of C), then $R' = R + N$ is also a valid reconstruction.

Proposition 7 (Affine reconstruction bundle)**.** *For a fixed linear coarse map C, the set of right inverses $\mathrm{Sec}(C) = \{R : CR = I_M\}$ is either empty or an affine space modelled on $\mathrm{Vert}(C) = \{N : CN = 0\}$.*

Proof. If $R, R' \in \mathrm{Sec}(C)$, then $C(R'-R) = 0$, so $R'-R \in \mathrm{Vert}(C)$. Conversely, $C(R+N) = CR+CN = I_M+0 = I_M$ for any $N \in \mathrm{Vert}(C)$. □

The hidden object is not a catalogue of primitive values; it is the affine space of admissible liftings.

2. Gauge as self-similar reconstruction freedom

Two reconstructions R and $R' = R + N$ are *gauge-equivalent* if both produce the same coarse law: the difference N is physically redundant given the coarse law-content.

Definition 3 (Gauge)**.** *Gauge freedom is reconstruction freedom constrained by preservation of coarse law-content.*

Gauge invariance in electrodynamics, Yang-Mills theory, or gravity corresponds to the physical statement that coarse predictions are invariant under the relevant class of fine-level rearrangements. Principal-bundle structure requires additionally a group $G_C = \{U : CU = C\}$ acting freely and transitively on admissible sections. The affine bundle (Proposition 7) is the underlying structure; naming G_C gives the principal bundle.

3. Holonomy from imperfect reconstruction transport

To compare reconstruction choices at neighbouring representations r^i in a space of representations $\mathcal{R}$, a connection is needed. Let

$$D_i = \partial_i + A_i, \tag{F2}$$

where A_i encodes how admissible reconstruction choices rotate as the representation changes. The curvature is

$$\Omega_{ij} = [D_i, D_j] = \partial_i A_j - \partial_j A_i + [A_i, A_j]. \tag{F3}$$

Around a closed loop γ in representation space, the holonomy is

$$U_\gamma = \mathcal{P}\exp\oint_\gamma A. \tag{F4}$$

When a compact one-dimensional reversible sector is isolated, this reduces to $U_\gamma = e^{i\theta_\gamma}$: the phase accumulated by transporting a reconstruction choice around a closed loop.

Perfect self-similarity of reconstruction (exact gauge invariance) gives flat connection: $\Omega_{ij} = 0$ and $U_\gamma = 1$. Imperfect self-similarity gives nonzero curvature and nontrivial holonomy. The Berry phase and its non-Abelian extension [27, 57, 58] are recovered as special cases in which the space of representations is the space of adiabatic parameters and the reconstruction fibre is a Hilbert-space bundle.

Closure residue and curvature are distinct objects. Closure residue ($\Delta_C \neq 0$) is closure failure at the representational level. Curvature ($\Omega \neq 0$) appears only after a connection on reconstruction choices has been specified. Do not equate them.

4. Noncommuting erasures and spectral closure residue

Spectral closure residue can arise from classical coarse-graining without invoking quantum mechanics. Let the

fine space be $\Omega = \{1, 2, 3\}$. Description A identifies $\{1, 2\}$ and leaves $\{3\}$ alone:

$$\Pi_A = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{F5}$$

Description B leaves $\{1\}$ alone and identifies $\{2, 3\}$:

$$\Pi_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix}. \tag{F6}$$

Both operations are classical averaging. Nevertheless,

$$[\Pi_A, \Pi_B] = \begin{pmatrix} 0 & -1/4 & 1/4 \\ 1/4 & 0 & -1/4 \\ -1/4 & 1/4 & 0 \end{pmatrix} \neq 0, \tag{F7}$$

with Frobenius norm 0.612. Incompatible lossy descriptions produce noncommuting projection-reconstruction operations when their erased distinctions are incompatible. The commutator norm is the finite spectral closure residue of the pair of descriptions.

When descriptions are spectrally incompatible, there is no single joint distribution over the primitive properties visible to both. The natural calculus is then states as positive normalised functionals over a noncommutative algebra—in finite dimensions resembling density-operator calculus with sequential update taking the projection form $\rho \mapsto \Pi_A \rho \Pi_A / \operatorname{Tr}(\rho \Pi_A)$. This paper does not derive the Born rule [59]; it locates the structural reason to expect noncommutative probability before additional axioms are supplied.

5. Infinitesimal residue density: full derivation

Following (22)–(25), expand the finite-step closure residue:

$$C_{s \to s+\delta s}\, G_s = G_s + \delta s\, L_s G_s + O(\delta s^2), \tag{F8}$$

$$\begin{aligned} G_{s+\delta s}\, C_{s \to s+\delta s} &= (G_s + \delta s\, \dot{G}_s)(I + \delta s\, L_s) + O(\delta s^2) \\ &= G_s + \delta s(\dot{G}_s + G_s L_s) + O(\delta s^2). \end{aligned} \tag{F9}$$

Subtracting:

$$\begin{aligned} \Delta_C{}^{s \to s+\delta s} &= \delta s(L_s G_s - G_s L_s - \dot{G}_s) + O(\delta s^2) \\ &= \delta s([L_s, G_s] - \dot{G}_s) + O(\delta s^2). \end{aligned} \tag{F10}$$

The residue density is $\omega_s = [L_s, G_s] - \dot{G}_s$. If $D_s = \partial_s - \operatorname{ad}(L_s)$ is the covariant scale-derivative, this is $\omega_s = -D_s G_s$. Exact differential closure is $D_s G_s = 0$: the generator is parallel-transported by the coarse-graining connection and leaves no residue.

6. Two-level spectral calculation

For $G_s = \frac{\Delta}{2}\sigma_z$ and $\omega_s = \kappa\sigma_x$, define the spectral connection (off-diagonal residue over spectral gap):

$$B_s = i\frac{\kappa}{\Delta}\,\sigma_y. \tag{F11}$$

Check: $[B_s, G_s] = [i(\kappa/\Delta)\sigma_y,\ (\Delta/2)\sigma_z] = -\kappa\sigma_x = -\omega_s$. ✓

With $\Pi_+ = (1 + \sigma_z)/2$, the transported projector is

$$\begin{aligned} \Pi_+(s + \delta s) &= \Pi_+(s) + \delta s[B_s, \Pi_+] + O(\delta s^2) \\ &= \Pi_+(s) - \delta s\,\frac{\kappa}{\Delta}\,\sigma_x + O(\delta s^2). \end{aligned} \tag{F12}$$

Therefore

$$[\Pi_+(s),\ \Pi_+(s + \delta s)] = -i\,\delta s\,\frac{\kappa}{\Delta}\,\sigma_y + O(\delta s^2), \tag{F13}$$

and

$$\lim_{\delta s \to 0} \frac{[\Pi_+(s),\ \Pi_+(s + \delta s)]}{\delta s} = -i\frac{\kappa}{\Delta}\,\sigma_y. \tag{F14}$$

The commutator density is finite; the raw commutator vanishes. The dimensionless spectral residue ratio is $\chi_s = \kappa/\Delta$. When $\chi_s \to 0$, spectral frames become asymptotically compatible (classical closure). When $\chi_s \to \chi^* \neq 0$, adjacent-scale descriptions differ infinitesimally but their spectral-frame rotation rate is finite and scale-stable: self-similar spectral residue.

Appendix G: Quantum-like fixed points

1. Self-similar residue dynamics

The residue density ω_s is itself subject to scale evolution. Let $\mathcal{R}_s$ be the space of residue densities at scale s and $\mathcal{S}_s : \mathcal{R}_s \to \mathcal{R}_{s+ds}$ the residue RG map.

Classical closure corresponds to the trivial fixed point $\omega^* = 0$: the residue decays under scale evolution and the system achieves classical autonomy in the limit. Quantum-like closure corresponds to a nontrivial fixed point:

$$\omega^* \neq 0, \qquad \mathcal{S}(\omega^*) = \omega^*. \tag{G1}$$

At this fixed point the failure of classical closure is itself governed by a self-similar law. The spectral noncommutativity produced by the off-diagonal part of ω^* is not a transient perturbation; it is a stable, scale-invariant feature of the algebra of descriptions.

More general self-similar residue structures include: anomalous scaling ($\omega_s \sim b^{-\Delta}\omega^*$), limit cycles ($\omega_{s+T} = \omega_s$), state-dependent fixed structures ($\omega_s = \omega_s[\rho]$), non-Abelian curvature ($F_{ab} \neq 0$), and nonassociative higher residue ($[A, B, C] \neq 0$).

2. Operational $\hbar$ as fixed-point closure residue

When spectral closure residue reaches the nontrivial fixed point, the off-diagonal residue density induces a stable commutator algebra on observable pairs. Define the *operational closure-residue scale* $\hbar_{\mathrm{op}}$: the coefficient converting stable spectral residue into an effective commutator algebra.

For conjugate finite descriptions Q and P, classical compatibility gives $[Q, P] = 0$. At the quantum fixed point, $[Q, P] = i\hbar_{\mathrm{op}}\,\Omega(Q, P)$ for a symplectic form Ω. Ordinary quantum mechanics is the special case in which $\hbar_{\mathrm{op}}$ becomes scalar, central, and universal:

$$\hbar_{\mathrm{op}}(s, \rho, C, R) \to \hbar\, I. \tag{G2}$$

A small loop in conjugate representation space returns with phase

$$\exp\!\left(-\frac{i}{\hbar}\,ab\right), \tag{G3}$$

where ab is the enclosed phase-space action area. $\hbar$ is the quantum of action because it is the fixed-point scale of irreducible representational residue. This identifies the structural role of $\hbar$; it does not determine the numerical value, which is set by physical content not representational calculus.

3. Classification of self-similar residue regimes

Table I organises the possible fixed-point and self-similar residue structures. Ordinary quantum mechanics is not the most general noncommutative theory of finite description; it is the unusually rigid case where spectral closure residue becomes scalar, central, universal, linear, associative, and Hilbert-representable.

4. GKSL form as Markovian quantum closure residue

When the spectral closure residue reaches the quantum fixed point ($\hbar_{\mathrm{op}} \to \hbar I$) and the residue dynamics is required to be *Markovian*—no memory of past scales—the generator of the residue semigroup is completely constrained. Gorini, Kossakowski, and Sudarshan [33] and Lindblad [32] proved independently that the most general generator of a completely positive trace-preserving semigroup on an N-level system has the form

$$\frac{d\rho}{dt} = -\frac{i}{\hbar}[H, \rho] + \sum_k \Big(L_k \rho L_k^\dagger - \tfrac{1}{2}\{L_k^\dagger L_k, \rho\}\Big), \tag{G4}$$

where H is the Hamiltonian (the spectral generator G_s), the L_k are jump operators, and $\{A, B\} = AB + BA$ is the anticommutator.

In the closure-residue framework this is the unique Markovian quantum form:

- The Hamiltonian commutator $-i[H, \rho]/\hbar$ is the spectral generator with the fixed-point residue $\chi^* = \hbar/\Delta$ folded in.
- The Lindblad terms $\sum_k (L_k \rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\})$ are the incoherent residue modes—transitions between spectral sectors that do not close as coherent off-diagonal rotation.

The GKSL equation is not a phenomenological model of dissipation; it is the algebraic consequence of requiring that the quantum closure-residue semigroup be completely positive and Markovian. More general residue structures (memory, non-Markovian, non-CP) arise when these requirements are relaxed, recovering the Nakajima-Zwanzig hierarchy (Appendix D) at the quantum level.

The Lindblad operators L_k are the jump operators of the residue: each one corresponds to a coherent or incoherent residue mode at the quantum fixed point. When all $L_k = 0$, the GKSL equation reduces to the Schrödinger (Liouville-von Neumann) equation: the fully coherent quantum fixed point with no incoherent residue.

5. Nonclaims

The framework:

- Identifies the structural setting for quantum-like behaviour as the nontrivial fixed point of self-similar spectral closure residue.
- Shows why noncommutative probability is the natural calculus when finite scale-fixings are spectrally incompatible.
- Identifies $\hbar$ as the universal scalar fixed-point scale of spectral closure residue.
- Identifies no-go theorems (Bell [34], Kochen-Specker [29, 35]) as obstructions to a global common spectral frame—the impossibility of making all reconstruction choices jointly flat.

The framework does not:

- Derive the Born rule [59].
- Reconstruct Hilbert space from first principles.
- Determine the numerical value of $\hbar$.
- Prove that all non-neutral scale fixing is quantum; more general self-similar residue regimes are possible (Table I).
- Claim that all commutator structures reduce to Hilbert-space quantum mechanics; standard quantum mechanics is the rigid fixed point, not the most general possible noncommutative theory.

TABLE I. Self-similar residue regimes. Each row corresponds to a distinct scale-stable structure of the spectral closure residue. Ordinary quantum mechanics is the rigid case in which the residue is scalar, central, universal, linear, associative, and Hilbert-representable.

Residue flow	Algebraic form	Regime
$\omega_s \to 0$	$[A, B] \to 0$	Classical closure
$\omega_s \to \omega^*$, $\hbar_{\rm op} \to \hbar I$	$[A, B] = i\hbar\,\Omega(A, B)$	Ordinary quantum mechanics
$\omega_s \to \omega^*$, $\hbar_{\rm op} \not\propto I$	$[A, B] = i\,\Theta(A, B)$	Noncanonical quantum-like theory
$\omega_s \sim b^{-\Delta}\omega^*$	$[A, B]_s = i\hbar_s\,\Omega(A, B)$	Scale-anomalous quantum theory
$\omega_{s+T} = \omega_s$	log-periodic commutator	Discrete-scale quantum-like theory
$\omega_s = \omega_s[\rho]$	$[A, B]_\rho = i\,\Omega_\rho(A, B)$	State-dependent / contextual theory
$F_{ab} \neq 0$	$[D_a, D_b] = F_{ab}$	Holonomic / gauge-like theory
$[A, B, C] \neq 0$	nonzero associator	Higher / nonassociative residue theory

The points where additional axioms would enter are: the admissible structure group on reconstruction choices, the algebra of observables, and the regularity of the residue flow. The present contribution is to have identified the structural reason to expect projector-noncommutativity, not to have derived it from scratch. Gleason's theorem [59] and related results show what additional assumptions can do once the appropriate projector structure is in place.

Appendix H: Scaling laws from closure-residue decay

This appendix provides technical details for the closure-residue interpretation of neural scaling laws and interpretability phase transitions.

1. Cross-entropy as KL residue

The training of a neural network can be viewed formally as a finite-description closure problem. Given a data distribution $p_{\rm data}$ and a parametrised model family p_θ, the standard cross-entropy loss is

$$\mathcal{L}(\theta) = -\mathbb{E}_{p_{\rm data}} \log p_\theta. \tag{H1}$$

This decomposes via standard identities into the irreducible entropy of the data generating process and the Kullback-Leibler divergence between the data and the model,

$$\mathcal{L}(\theta) = H(p_{\rm data}) + D_{\rm KL}(p_{\rm data} \| p_\theta). \tag{H2}$$

The $D_{\rm KL}$ term is the scalar representational residue: the excess statistical content of the data law that cannot be absorbed into the finite representation p_θ. The irreducible entropy $H(p_{\rm data})$ is the null statistical content, fixed by the data source independently of model choice.

2. Residue-mode decomposition

Consider a representation-scale parameter s. The closure residue can be decomposed into an eigenbasis of modes,

$$\Delta_s = \sum_k a_k(s)\phi_k. \tag{H3}$$

Linearising the residue renormalisation group flow near a closure fixed point yields

$$\partial_s a_k = -\lambda_k a_k, \tag{H4}$$

where λ_k are the corresponding eigenvalues. The solution implies exponential decay in the scale parameter,

$$a_k(s) = a_k(0)e^{-\lambda_k s}. \tag{H5}$$

By defining the scale parameter logarithmically with respect to a physical model resource, $s = \log N$ (where N is parameter count or capacity), we obtain

$$a_k(N) = a_k(0)N^{-\lambda_k}. \tag{H6}$$

If the total loss is dominated by the leading (slowest-decaying) residue mode, the expected scaling law emerges [41]:

$$\mathcal{L}(N) - \mathcal{L}_\infty \sim AN^{-\alpha}. \tag{H7}$$

3. Multi-axis scaling

In practice, model capacity, dataset size, and compute define a coupled system of representation scales: $s_N = \log N$, $s_D = \log D$, and $s_C = \log C$. The residue flow then depends on multiple axes,

$$\mathcal{L}(N, D) - \mathcal{L}_\infty \approx AN^{-\alpha} + BD^{-\beta} + E_{\rm int}(N, D). \tag{H8}$$

Here, the $AN^{-\alpha}$ term is the capacity closure residue, the $BD^{-\beta}$ term is the sampling (or null-law) residue, and $E_{\rm int}$ represents the interaction residue stemming from imperfect factorisation between the capacity and data scales.

4. Compute-optimal balance

Following the derivation of compute-optimal scaling [42], we apply a simple compute constraint $C \sim ND$. We wish to minimise the total residue

$$AN^{-\alpha} + BD^{-\beta} \tag{H9}$$

subject to $ND = C$. Substituting $D = C/N$, we obtain

$$\mathcal{L}(N) \approx AN^{-\alpha} + BC^{-\beta}N^{\beta}. \tag{H10}$$

Taking the derivative with respect to N and setting to zero yields the optimal balance,

$$\alpha AN^{-\alpha} \sim \beta BD^{-\beta}. \tag{H11}$$

This provides a closure-residue interpretation for compute-optimal scaling: the allocation of resources is optimal when the marginal reduction of capacity residue exactly balances the marginal reduction of sampling residue. Neither should dominate.

5. Circuit formation as residue-class transition

The closure calculus predicts that the emergence of discrete structural phenomena—such as induction heads, feature splitting, or superposition [46, 60]—corresponds to a transition in the class of the residue.

Defining a smooth scaling residual,

$$R(N, D) = \mathcal{L}_{\text{obs}}(N, D) - \mathcal{L}_{\text{fit}}(N, D), \tag{H12}$$

provides a testable empirical programme: mechanistic circuit formation should occur near structured features of $R(N, D)$. Specifically, local extrema, knees, exponent changes, or curvature changes in the log-log loss curve should mark the points where residue transitions from incoherent/nonlocal forms into coherent visible structure.

6. Singular learning theory and residue eigenvalues

Watanabe's singular learning theory [43, 44] provides an independent derivation of power-law scaling that deepens the closure-residue account. For a parametric model p_θ trained on n samples from p_{data}, the average generalisation loss satisfies

$$\mathbb{E}[\mathcal{L}_n] = \mathcal{L}_\infty + \frac{\lambda}{n} + o(1/n), \tag{H13}$$

where λ is the *real log-canonical threshold* (RLCT)—a rational number determined by the algebraic geometry of the zero set of $D_{\text{KL}}(p_{\text{data}} \| p_\theta)$. With sample-size $n \sim N^\gamma$ for some scaling exponent γ, this gives $\mathcal{L}(N) - \mathcal{L}_\infty \sim AN^{-\gamma}$. The RLCT is the smallest eigenvalue in the Newton polyhedron of the KL geometry near its singular points; it is the RG eigenvalue of the dominant closure-residue mode in the present framework.

The connection is precise: the residue eigenvalues λ_k in Eq. (H13) are the eigenvalues governing residue-mode decay near closure fixed points. When the model class is nonsingular (regular model), all eigenvalues are integer-valued and the RLCT equals half the parameter dimension—recovering standard parametric rates. Singular models (neural networks, mixture models, hidden Markov models) have non-integer RLCT, producing sub-linear exponents. The closure-residue framework provides a dynamical-systems interpretation of the SLT result: the RLCT measures the depth of the singularity in the KL geometry, which controls how slowly the dominant residue mode decays under scale evolution.

The WAIC [44] is the practical criterion: it is an asymptotically unbiased estimator of the generalisation loss that automatically detects singular model structure. In the closure-residue language, WAIC estimates the residue remaining after training.

7. Emergent abilities as residue-class transitions

Discrete qualitative transitions in LLM capabilities—the sudden appearance of multi-step reasoning, chain-of-thought, or in-context learning at specific scale thresholds [47]—are predicted by the closure-residue framework to be residue-class transitions: points where the dominant residue mode changes algebraic type. Below the transition the dominant residue is incoherent or nonlocal; above it a coherent residue mode closes as an internal transition operator (an induction head or a reasoning circuit), producing the observed qualitative jump.

This is distinct from the SLT account: SLT describes smooth power-law decay within a fixed residue class, while emergent abilities mark the boundary between residue classes. The two accounts are complementary: smooth scaling is self-similar decay within a class; emergent abilities are the transitions between classes. The joint prediction is that the scaling curve $\mathcal{L}(N)$ should be *piecewise power-law*: locally smooth with distinct exponents on either side of a residue-class transition, with the transition visible as a knee or exponent change in the residual $R(N, D)$.

8. Relation to existing literature

The smooth power-law scaling of loss with model and data size was established empirically by Kaplan *et al.* [41] and subsequently refined by compute-optimal analyses [42]. These empirical laws have been studied in terms of data-limited versus parameter-limited regimes, but the present framework offers a structural account: each scaling axis corresponds to an independent residue-decay mode, and compute-optimal balance is the condition that no single mode dominates. The singular-learning-theory

perspective [43] identifies the power-law exponent with an algebraic invariant of the model class, providing rigorous bounds on achievable scaling rates.

Mechanistic interpretability has documented a range of discrete structural transitions in trained networks—induction heads [45], superposition [46], and feature splitting—that appear as sharp qualitative changes rather than smooth variations [47]. From the closure-residue perspective, these are residue-class transitions: the dominant residue mode changes algebraic type (from nonlocal/incoherent to coherent/transition), producing the observed discontinuous character. The present framework predicts that such transitions should be visible as structured features in the residual $R(N, D)$ defined above, and thus provides a quantitative bridge between the scaling-laws literature and the mechanistic-interpretability literature.

9. Empirical prediction

The central empirical prediction of the closure-residue framework is:

> *Structured deviations from smooth scaling residuals $R(N, D)$ should correlate with mechanistic reorganisation events—circuit formation, feature splitting, induction-head emergence, and superposition reorganisations—rather than occurring randomly along the scaling curve.*

Smooth scaling is the self-similar decay of residue modes in the absence of residue-class transitions. Deviations from smooth scaling are not noise by default; they mark points where the algebraic class of the dominant residue changes. This prediction is falsifiable: if mechanistic transitions occur at randomly distributed points along the scaling curve, the framework's identification of residue-class transitions with observable scaling features would be refuted.

Appendix I: Compression costs and mass-like parameters

The cost-function structure of K_s identifies where mass-like parameters enter the closure calculus. The structural mapping is:

$$K_s = \sum_a g_a A_a \tag{I1}$$

$$A_{\text{comp}} \xrightarrow{\text{enforced}} g_{\text{comp}} \in K_s \xrightarrow{\text{if closed}} m_{\text{eff}}, \tag{I2}$$

where A_{comp} is the retained compression variable, g_{comp} its conjugate cost, and m_{eff} the resulting mass-like coefficient.

A localised excitation is a compressed pattern of antecedent structure that remains identifiable after coarse-graining. If the finite description retains a variable A_{comp} measuring the persistence of such a compression, the conjugate coefficient g_{comp} entering K_s is the cost of maintaining it. When this cost:

1. closes under the coarse-graining (the cost survives as a well-defined feature of the coarse description),
2. is a scalar (not a matrix or tensor in the spectral frame),
3. is transportable (the same cost appears consistently across representation changes),

it has the structural role of a mass-like parameter. In this limited sense, mass is compression after closure: not the compression pattern itself, but the cost of maintaining a self-closing compression against the null law.

To obtain the familiar local form of a mass term would require the following additional steps:

1. Derive the cost-of-complexity function explicitly from the null-law recursion.
2. Obtain its differential form by expanding the residue triage in δs.
3. Lift the construction to path/action space (Appendix B).
4. Take the relevant small-deformation (gradient-expansion) limit.

These steps belong to a more developed treatment of the path-space version of the theory. The present contribution is more limited:

> The framework identifies where mass-like coefficients enter the closure calculus: as closed scalar costs of retained compression.

This identification is independent of the specific physical content (field type, spacetime symmetry, gauge structure) that would determine the numerical value of the mass. Those inputs enter at the step of specifying the observable A_{comp} and the admissible class $\mathcal{N}_s$, not at the level of the general closure calculus.

Appendix J: Action-angle coordinates and the origin of $\mathcal{K}_s$

This appendix makes explicit the claim in Section II that K_s is forced by the coarse-graining's action on the symmetry structure of G_μ. Section J 1 gives a four-state Markov chain toy model that illustrates each fate explicitly (Fig. 3). Sections J 2–J 6 give the continuum version in action-angle coordinates. We work with integrable and near-integrable generators, where the construction is concrete.

1. Four-state Markov chain: an explicit toy model

Consider a continuous-time Markov chain on four states $\{1,2,3,4\}$, grouped by the coarse-graining C into two coarse states $\alpha = \{1,2\}$ and $\beta = \{3,4\}$ (Fig. 3b). The fine generator is

$$G_\mu = k_f L_{\text{intra}} + k_s L_{\text{inter}} + \varepsilon L_{\text{diag}}, \tag{J1}$$

where L_{intra} drives $1 \leftrightarrow 2$ and $3 \leftrightarrow 4$ at rate k_f (fast intra-partition equilibration); L_{inter} drives $1 \leftrightarrow 3$ and $2 \leftrightarrow 4$ at rate k_s (slow inter-partition flow); and L_{diag} drives $1 \leftrightarrow 4$ and $2 \leftrightarrow 3$ at rate ε (symmetry-breaking diagonal transitions).

Fate of k_f (exact symmetry $\to$ null law). In the limit $k_f \to \infty$ with k_s, ε fixed, states within each partition equilibrate instantly. The stationary weight within α is uniform: $q(1|\alpha) = q(2|\alpha) = \frac{1}{2}$, and likewise for β. This intra-partition flat distribution is the null law at the coarse scale. The coarse generator G_M is independent of k_f; the fast transitions are absorbed entirely into q_s via the recursion (4).

Fate of k_s (demoted symmetry $\to$ K_s). The inter-partition transitions $1 \leftrightarrow 3$ and $2 \leftrightarrow 4$ both contribute equally to $\alpha \leftrightarrow \beta$ flow. Projecting through C gives a coarse generator with rate k_s each way: $G_M = k_s(L_{\alpha\beta} + L_{\beta\alpha})$. This flow is representable in coarse variables. In the free-energy form, the log-ratio of coarse stationary weights gives $K_s = g\,\mathbf{1}_\beta$ with $g = \log(k_s^+/k_s^-)$ when rates are asymmetric; at detailed balance, $g = 0$ and q_s absorbs the k_s structure as well.

Fate of ε (broken symmetry $\to \Delta_s$). The diagonal transitions $1 \to 4$ and $2 \to 3$ move between partitions but in a way that breaks the $\alpha \leftrightarrow \beta$ symmetry: $1 \to 4$ increases the β-side occupation of state 4 relative to state 3. This asymmetry cannot be expressed in any functional of the coarse occupation probabilities $p(\alpha)$ and $p(\beta)$ alone. Computing $\Delta_C = CG_\mu - G_M C$ gives a non-zero matrix whose entries are proportional to ε; it cannot be set to zero by any choice of G_M. This is the closure residue.

The three fates are therefore structural, not incidental: they follow directly from how the coarse-graining partition interacts with the symmetry structure of the fine generator.

2. Integrable generators in action-angle form

Let the fine-scale generator be integrable: there exist canonical coordinates (I_a, θ_a), $a = 1, \ldots, n$, such that the generator takes the form

$$G_\mu = \sum_a \omega_a(I)\, \partial_{\theta_a}, \tag{J2}$$

where the frequencies $\omega_a(I) = \partial H/\partial I_a$ depend only on the actions and $H(I)$ is the Hamiltonian in action variables. The actions I_a are exact conserved quantities: $\{I_a, H\} = 0$ for all a. The fine stationary state is

$$\mu \propto \exp[-\beta H(I)]\,, \tag{J3}$$

depending only on the actions.

3. Coarse-graining that averages over angles

Consider a coarse-graining C that marginalises over all angle coordinates θ while retaining all actions I. The pushforward of μ is

$$C_*\mu(I) = \int_{[0,2\pi)^n} \mu(I,\theta)\, \frac{d^n\theta}{(2\pi)^n} \propto \exp[-\beta H(I)]\,. \tag{J4}$$

The coarse stationary state has the same form as the fine one. The coarse generator G_M drives toward this state; $K_s = \beta H(I)$ is the cost function of retained complexity, with each action I_a entering with conjugate coefficient $g_a = \beta\omega_a$. The null law at the coarse scale is the uniform measure over action space (the image of the flat measure under C); all structure in γ_s above this background is accounted for by K_s.

4. Coarse-graining that additionally marginalises some actions

Now let C also marginalise over a subset of actions, say $I_2, \ldots, I_n$, retaining only I_1. The coarse stationary state is

$$C_*\mu(I_1) \propto \exp[-\beta\omega_1 I_1] \cdot Z_{2,\ldots,n}, \tag{J5}$$

where $Z_{2,\ldots,n} = \int \exp[-\beta \sum_{a\geq 2} \omega_a I_a]\, dI_2 \cdots dI_n$ is a constant. Only I_1 remains as a visible constraint; $K_s = \beta\omega_1 I_1$. The erased actions $I_2, \ldots, I_n$ have been absorbed into the null law via the null-law recursion (4): their Boltzmann factors become part of q_s, contributing to the null potential $V_s^{(0)}$.

5. Coarse-graining that mixes actions: the coupling angle

The interesting case is a coarse-graining that retains a linear combination of actions, $\hat{I} = \alpha I_1 + (1-\alpha) I_2$, for some $\alpha \in (0,1)$. This occurs when the coarse observable is not aligned with any single mode of the fine generator — a generic situation in many-body and field-theoretic coarse-grainings. The coarse stationary state concentrates on $\hat{I}$-level sets; the effective frequency is

$$\hat{\omega} = \alpha\omega_1 + (1-\alpha)\omega_2, \tag{J6}$$

a convex combination of the fine frequencies. The coefficient entering K_s is $g_{\hat{I}} = \beta\hat{\omega}$.

The mixing angle α is determined geometrically by the coarse-graining: it is the projection of the coarse observable $\hat{I}$ onto the action basis of the fine generator. Modes orthogonal to $\hat{I}$ in action space are erased by C; their Boltzmann weights fold into q_s. Modes not exactly aligned but with nonzero projection onto $\hat{I}$ contribute to K_s with coefficients set by the projection angle.

This is the mechanism behind the Legendre structure of K_s: the conjugate pairs (A_a, g_a) in the coarse law correspond precisely to the eigendirections of the fine generator that C retains (exactly or approximately), and the coefficients g_a are determined by the geometry of the coarse-graining in the symmetry space of G_μ.

6. Near-integrable corrections

When the fine generator is near-integrable — $G_\mu = G_0 + \epsilon V$ where G_0 is integrable — the actions I_a acquire slow drift under V. To lowest order in ϵ, the coarse-graining analysis proceeds as above with corrected effective frequencies $\omega_a \to \omega_a + \epsilon\delta\omega_a$. The corrections $\delta\omega_a$ are the secular parts of V in the action-angle frame of G_0; they enter the coefficients g_a in K_s and are otherwise invisible to the coarse description. Resonant corrections ($k \cdot \omega = 0$ for integer vector k) generate closure residue: they produce terms that mix action-angle sectors and cannot be absorbed as a modified K_s without residue.

Appendix K: Constructive demonstration: twelve-state Markov chain

This appendix provides the full specification of the toy model used in Section IV, tabulates all candidate coarse-grainings, and contains the detailed figures.

Generator convention

The antecedent generator $G_\mu \in \mathbb{R}^{n\times n}$ obeys the column-stochastic convention

$$G_\mu[i,j] \geq 0 \ (i \neq j), \qquad G_\mu[j,j] = -\sum_{i\neq j} G_\mu[i,j], \quad \text{(K1)}$$

so that $\dot{p} = G_\mu p$ preserves normalisation and non-negativity for any initial probability vector p.

Model specification

The antecedent system is a 12-state CTMC with three blocks: $A = \{0,1,2,3\}$, $B = \{4,5,6,7\}$, $C = \{8,9,10,11\}$. Each block has a *hub-and-spoke* internal structure: state 0 (resp. 4, 8) is the hub of block A (resp. B, C), connected to three spokes. Hub-to-spoke rate: $r_{\text{hub}} = 30$ (outward); spoke-to-hub and spoke-to-spoke rates: $r_{\text{spoke}} = 12$ (inward/lateral). These fast intra-block rates produce within-block equilibration on timescale $\sim 1/r_{\text{hub}} \approx 0.03$, so each block appears as a single effective state at the inter-block timescale.

The hub ratio $r_{\text{hub}}/r_{\text{spoke}} = 2.5$ is chosen so that spokes carry more stationary mass than hubs: spoke states are *more probable but less connected*. This causes variance-based (fast-mode) compression, which assigns large variance to high-mass spokes, to group hub and spoke states across different blocks rather than keeping each block intact.

Slow inter-block connections link the hubs: $A \leftrightarrow B$ and $B \leftrightarrow C$ at rate $r_{\text{inter}} = 0.4$; a very slow direct $A \leftrightarrow C$ hub-to-hub path at rate $r_{\text{direct}} = 0.05$. Under these rates alone, topology correctly suggests B mediates all A–C flux.

The *hidden shortcut*: spoke state $1 \in A$ is directly connected to spoke state $9 \in C$ at forward rate $r_h = 2.0$ and backward rate 1.4. This connection is invisible in the nominal block graph. The effective A-to-C transport rate, accounting for the stationary fraction $\pi_{1|A} \approx 0.29$ of block-A mass in state 1, is

$$r_{AC}^{\text{eff}} \approx r_{\text{direct}} + \pi_{1|A} \cdot r_h \approx 0.05 + 0.29 \times 2.0 \approx 0.64, \quad \text{(K2)}$$

which exceeds both $r_{AB} = r_{BC} = 0.40$. The hidden shortcut therefore makes A and C dynamically more proximate than either is to B, despite B topologically separating them (Fig. 12).

The model is designed so that: (i) the topologically obvious partition $\{A,B\} \mid \{C\}$ is wrong; (ii) the correct partition $\{A,C\} \mid \{B\}$ is identifiable by $J(C)$; (iii) fast-mode compression fails for a mechanistically clear reason; and (iv) the residue column profile σ_j of the wrong partition points directly to the shortcut endpoints.

Aggregation matrices

The coarse-graining C is a binary aggregation matrix: $C_{\alpha i} = 1$ if antecedent state i belongs to coarse group α, and 0 otherwise. Exactly one entry per column equals 1. C is *not* row-stochastic; it maps probability vectors by $P = Cp$ (conserving normalisation since $\mathbf{1}^\top C = \mathbf{1}^\top$).

The stationary-weighted right inverse satisfying $CC^+ = I_{n_M}$ is

$$C^+ = WC^\top \left(CWC^\top\right)^{-1}, \qquad W = \text{diag}(\pi_\mu). \quad \text{(K3)}$$

All candidate coarse-grainings

Table II reports the closure loss and prediction error at $t = 5$ for all candidate 2-state coarse-grainings considered, plus the promoted 3-state partition. "Fast-mode" clusters on the fastest non-trivial eigenvector of G_μ (largest $|\lambda|$ above a threshold), which captures within-block hub/spoke structure rather than the inter-block slow variable. "Slow-mode heuristic" clusters on the

slowest non-trivial eigenvector; this also fails here because the hidden shortcut contaminates the slow spectrum (Fig. 13).

TABLE II. Closure loss $\|\Delta_C\|_{W,F}$ and mean L^1 prediction error at $t = 5$ (averaged over 50 random initial distributions) for all candidate coarse-grainings.

Coarse-graining	$\|\Delta_C\|_{W,F}$	PE($t = 5$)
$\{A, C\} \mid \{B\}$	0.26	6.1×10^{-4}
$\{A\} \mid \{B, C\}$	0.90	0.044
$\{A, B\} \mid \{C\}$	0.93	0.044
Fast-mode (fastest eigvec)	30.9	0.048
Slow-mode heuristic	32.1	0.069
Promoted $\{1, 9\}$ (3-state)	13.6	0.047

The "promoted $\{1, 9\}$" row corresponds to splitting state 1 from block A and state 9 from block C into a new third coarse variable, forming the partition $\{A\backslash 1,\, B\} \mid \{C\backslash 9\} \mid \{1, 9\}$. This partition directly encodes the shortcut pair as a visible coarse state, and the best-fit $G^*_M(C)$ for it shows explicit transition rates between $\{1, 9\}$ and both of the other coarse groups. The high closure loss (13.6) reflects that the partition is not a clean dynamical grouping: it severs states 1 and 9 from their host blocks, creating large residue for the remaining block-membership transitions. The residue signal is therefore correctly interpreted not as "isolate states 1 and 9" but as "states 1 and 9 link blocks A and C, so merge A and C."

Detailed results

R1. Closure-optimised basis recovers the dynamically correct grouping. $\{A, C\} \mid \{B\}$ achieves weighted closure loss 0.26 and prediction error 6.1×10^{-4} at $t = 5$. Topology-guided alternatives score 0.90–0.93 on closure loss and 0.044 on prediction error—a factor of 72× worse. Fast-mode compression yields closure loss 30.9: two orders of magnitude above the optimum. The slow-mode heuristic (closure-learned basis) yields 32.1 (Fig. 13; see below).

R2. Residue anatomy identifies hidden coupling states. The weighted column profile σ_j of $\{A, B\} \mid \{C\}$ shows two anomalous antecedent states: $\sigma_1 \approx 0.57$ and $\sigma_9 \approx 0.47$, separated from the next-largest contributor by a factor of ∼6 (Fig. 14). These are precisely the hidden-shortcut endpoints. No information about the shortcut is used in computing σ_j; the large values emerge purely from the mismatch between antecedent dynamics and the coarse law imposed by $\{A, B\} \mid \{C\}$.

R3. Spectral heuristics fail because the hidden coupling contaminates the slow spectrum. The hidden shortcut perturbs the within-A stationary distribution and generates an additional slow mode—hub/spoke relaxation modulated by the escape channel—that occupies the spectral slot used for thresholding. The resulting partition (hubs in one group, spokes in another, across all three blocks) has closure loss 32.1, worse than any topology-guided partition. The direct criterion $J(C)$, evaluated on explicit candidate partitions, is unaffected by this spectral contamination and correctly ranks all candidates (Fig. 13).

R4. Transport identity holds at machine precision. For the three-scale tower $12 \to 3 \to 2$ (antecedent blocks, then $\{A, C\} \mid \{B\}$), the residue composition identity

$$\Delta_{C02} = C_{12}\,\Delta_{C01} + \Delta_{C12}\,C_{01} \tag{K4}$$

holds to residual 2.83×10^{-15}—machine precision—verifying that residue composition is an exact algebraic identity, not an approximation (Fig. 15).

Connection to existing methods

The generator-closing criterion unifies several apparently distinct representation-learning approaches.

Koopman-based methods [61–63] seek invariant subspaces of the Koopman operator, which is the dual condition (18). Mori–Zwanzig projection [9, 10, 17] decomposes dynamics into a projected and a hidden part; the hidden part is precisely $CG_\mu Q$ where $Q = I - C^+C$ ($C^+ = WC^\top(CWC^\top)^{-1}$ is the stationary-weighted right inverse satisfying $CC^+ = I_M$). Spectral clustering and diffusion maps [64, 65] cluster states by slow eigenvectors, approximating invariant subspaces. Slow-feature analysis [66] minimises temporal change in observables, corresponding to minimising $J(C)$ over observable-valued C. Transfer-operator methods [67, 68] identify metastable sets as approximate invariant subspaces.

These are not separate heuristics; they are each a specialisation of the condition: minimise the closure residue of the generator under a finite representation. The present framework adds two features absent individually: a single criterion $J(C)$ applicable regardless of representation type, and a structured column-profile decomposition σ_j that diagnoses what the representation has forgotten.

[1] E. T. Jaynes, Physical Review **106**, 620 (1957).
[2] E. T. Jaynes, Physical Review **108**, 171 (1957).
[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley, 2006).
[4] M. J. Wainwright and M. I. Jordan, Foundations and Trends in Machine Learning **1**, 1 (2008).

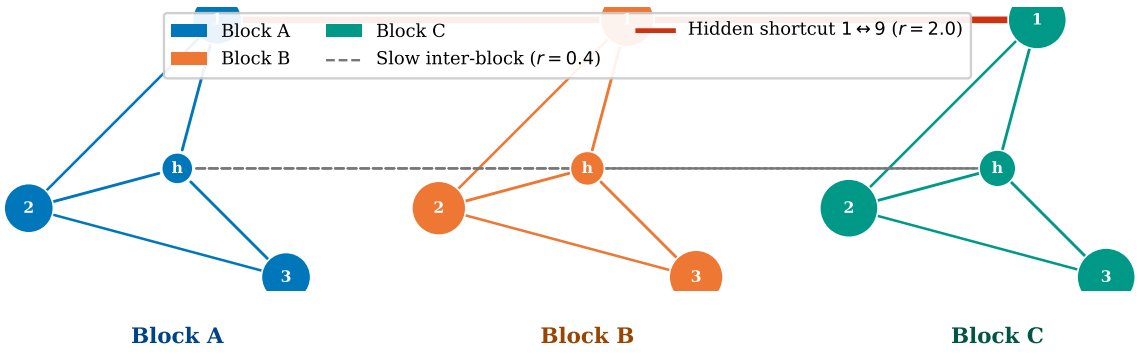

FIG. 12. **Antecedent Markov network (12 states, 3 blocks).** Node size proportional to stationary probability π_j; spoke states are more probable than hubs. Intra-block edges (coloured, solid) carry fast rates $r_{\text{hub}} = 30$ (hub $\to$ spoke) and $r_{\text{spoke}} = 12$ (spoke $\to$ hub and lateral), producing rapid within-block equilibration. Slow inter-block hub-to-hub edges (grey, dashed) carry $r_{\text{inter}} = 0.4$. The hidden shortcut state $1 \to$ state 9 (red, rate 2.0) creates an effective $A \leftrightarrow C$ coupling of $\approx$0.64, exceeding r_{inter}. Block membership is not visible in the topology: B lies between A and C in graph distance, yet A and C are dynamically closer.

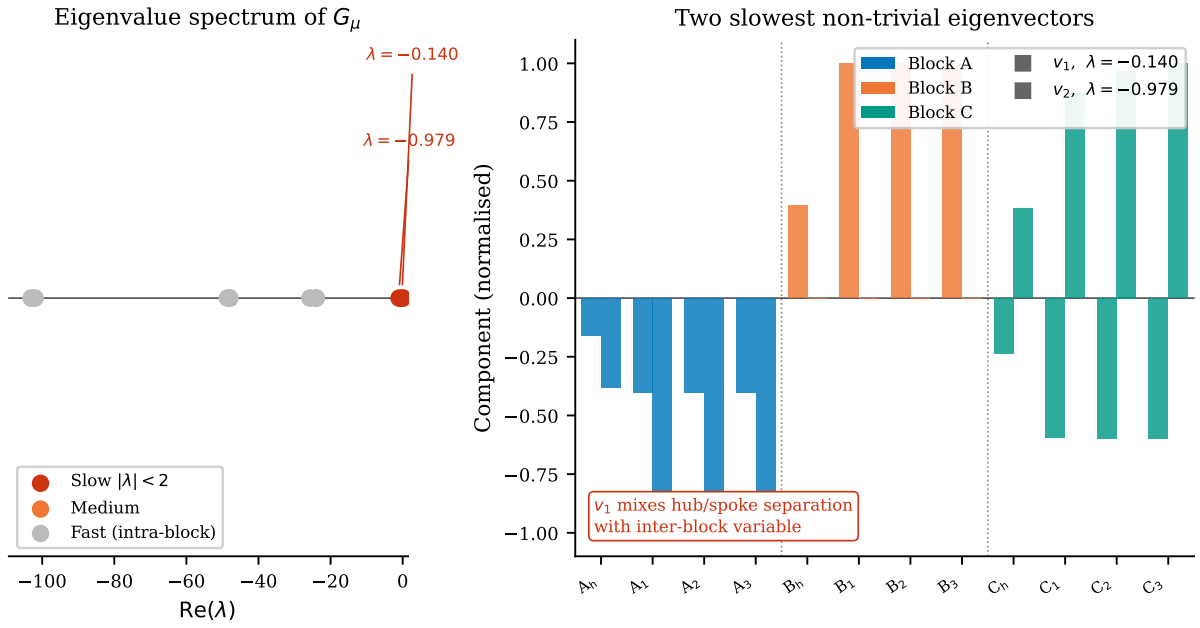

FIG. 13. **Eigenvalue spectrum and slow eigenvectors of G_μ.** *Left:* The 12 eigenvalues on the real line. Two slow modes ($\lambda \approx -0.14$ and $\lambda \approx -0.98$, red) are separated by a spectral gap from the fast intra-block modes (grey). *Right:* Components of the two slowest eigenvectors over all 12 states, coloured by block membership. The slowest mode ($\lambda \approx -0.14$) is contaminated by the within-block hub/spoke asymmetry induced by the hidden shortcut; it separates spoke states from hub states across all three blocks rather than separating $\{A, C\}$ from $\{B\}$. A threshold-based heuristic using this eigenvector therefore fails to recover $\{A, C\} \mid \{B\}$, while the direct criterion $J(C)$ is unaffected.

[5] N. Tishby, F. C. Pereira, and W. Bialek, in *37th Annual Allerton Conference on Communication, Control, and Computing* (1999).
[6] J. Rissanen, Automatica **14**, 465 (1978).
[7] J. E. Shore and R. W. Johnson, IEEE Transactions on Information Theory **26**, 26 (1980).
[8] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Reviews of Modern Physics **85**, 1115 (2013).
[9] R. Zwanzig, Physical Review **124**, 983 (1961).
[10] H. Mori, Progress of Theoretical Physics **33**, 423 (1965).
[11] K. G. Wilson, Physical Review B **4**, 3174 (1971).
[12] L. P. Kadanoff, Physics Physique Fizika **2**, 263 (1966).
[13] S. Weinberg, Physica A: Statistical Mechanics and its Applications **96**, 327 (1979).
[14] H. Georgi, Annual Review of Nuclear and Particle Science **43**, 209 (1993).
[15] C. R. Shalizi and J. P. Crutchfield, Journal of Statistical Physics **104**, 817 (2001).

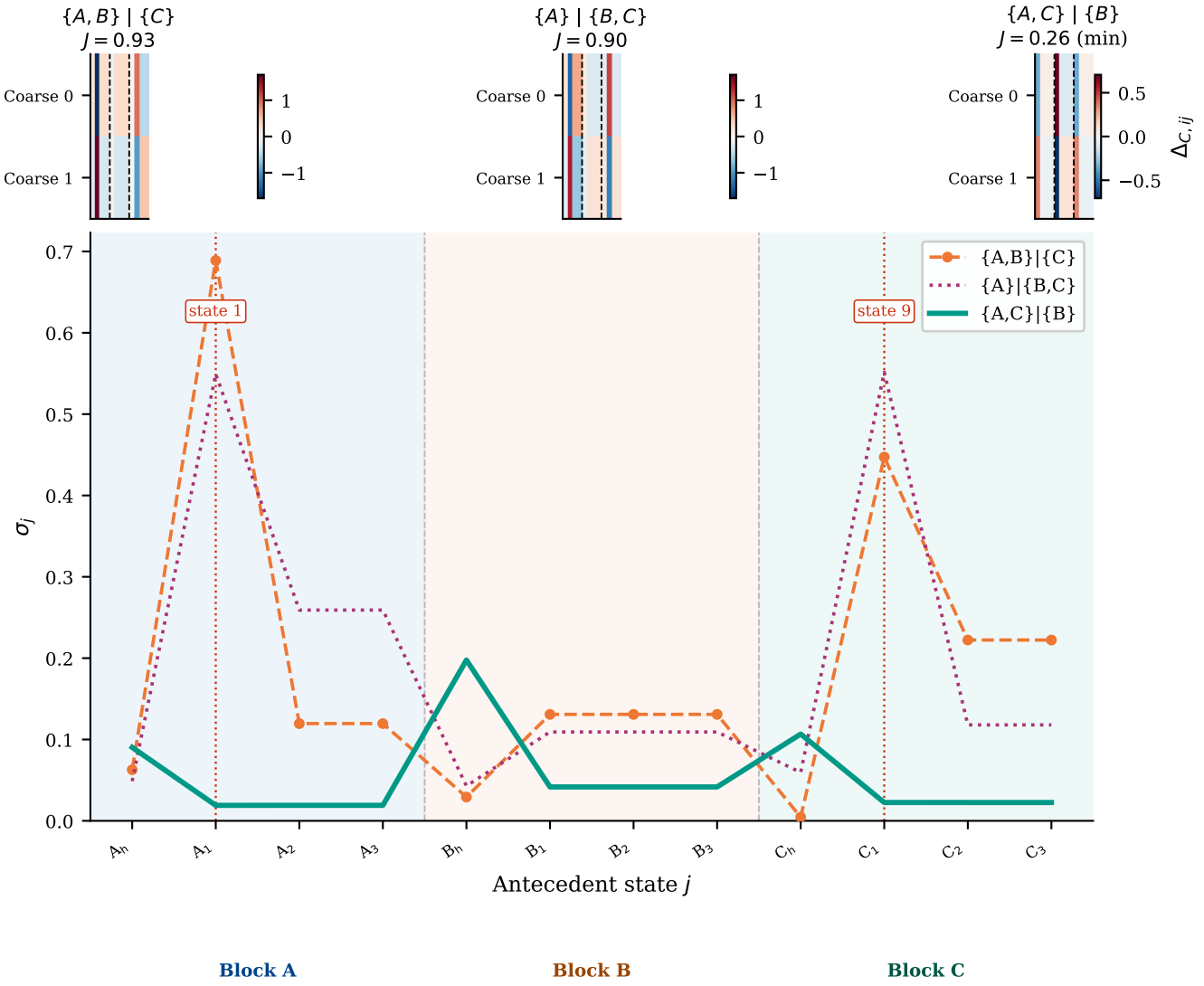

FIG. 14. **Closure residue anatomy for candidate bases.** *Top row:* Heatmaps of $\Delta_C = CG_\mu - G_M^* C$ for the three binary block partitions. The two topology-guided partitions carry large, spatially coherent residue; $\{A, C\} \mid \{B\}$ carries small, diffuse residue (closure losses 0.26 vs. 0.93 and 0.90). *Bottom panel:* Weighted column profile σ_j for all three 2-state partitions. For $\{A, B\} \mid \{C\}$, states 1 and 9 stand out by a factor of $\sim$6 above the next contributor: these are the hidden-shortcut endpoints, identified by residue anatomy without prior knowledge of the shortcut. For $\{A, C\} \mid \{B\}$, the profile is uniformly small and structureless, indicating an adequate representation.

[16] K. J. Friston, arXiv 10.48550/arXiv.1906.10184 (2019).
[17] S. Nakajima, Progress of Theoretical Physics **20**, 948 (1958).
[18] K. J. Friston, Nature Reviews Neuroscience **11**, 127 (2010).
[19] M. Gerstenhaber, Annals of Mathematics **79**, 59 (1964).
[20] J. Polchinski, Nuclear Physics B **231**, 269 (1984).
[21] P. W. Anderson, Science **177**, 393 (1972).
[22] E. P. Hoel, L. Albantakis, and G. Tononi, Proceedings of the National Academy of Sciences **110**, 19790 (2013).
[23] H. Grabert, *Projection Operator Techniques in Nonequilibrium Statistical Mechanics*, Springer Tracts in Modern Physics, Vol. 95 (Springer Berlin Heidelberg, 1982).
[24] R. P. Feynman and F. L. J. Vernon, Annals of Physics **24**, 118 (1963).
[25] A. O. Caldeira and A. J. Leggett, Annals of Physics **149**, 374 (1983).
[26] M. Nakahara, *Geometry, Topology and Physics* (Institute of Physics Publishing, 2003).
[27] F. Wilczek and A. Zee, Physical Review Letters **52**, 2111 (1984).
[28] S. Abramsky and A. Brandenburger, New Journal of Physics **13**, 113036 (2011).
[29] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-Å. Larsson, Reviews of Modern Physics **94**, 045007 (2022).
[30] U. Seifert, Reports on Progress in Physics **75**, 126001 (2012).

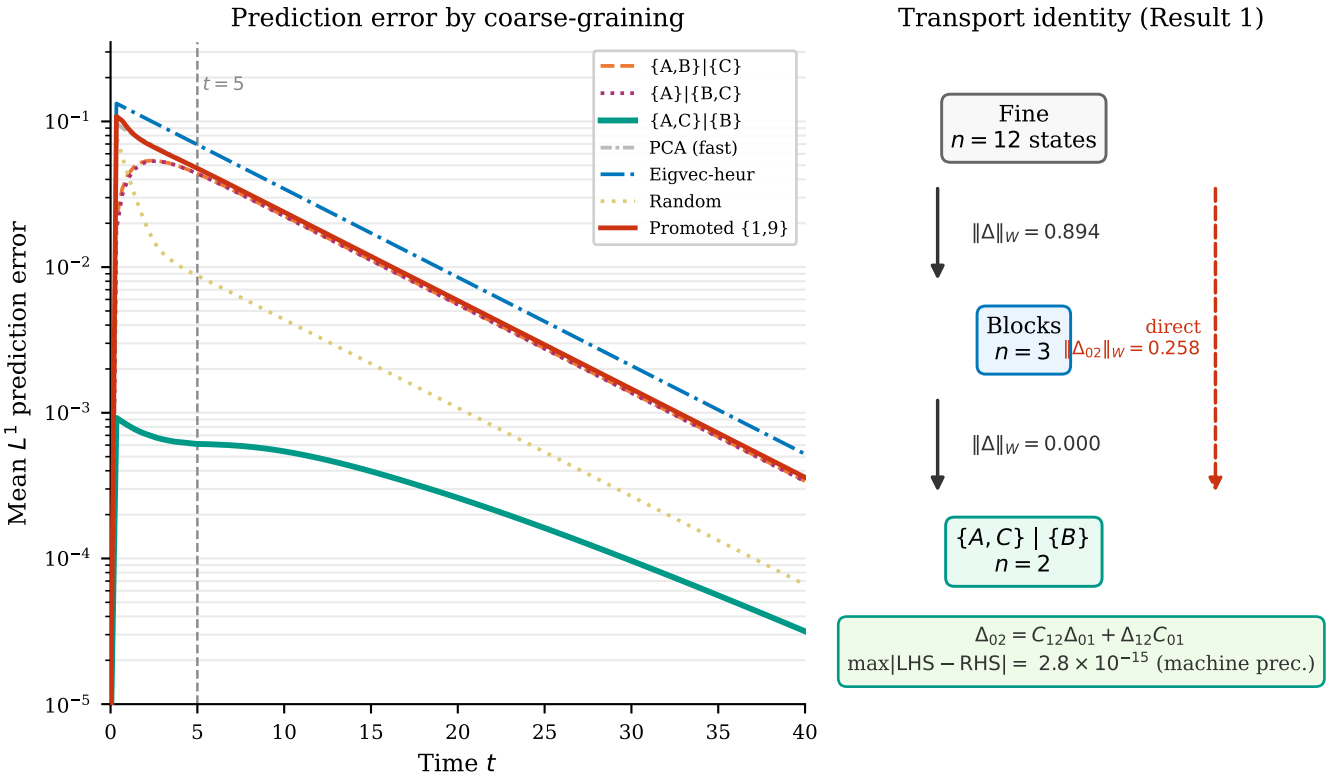

FIG. 15. **Prediction accuracy and the transport identity.** *Left:* Mean L^1 prediction error $\mathbb{E}_\rho[\|Ce^{tG_\mu}\rho - e^{tG^*_M}C\rho\|_1]$ versus time for all candidate coarse-grainings (averaged over 50 random initial distributions ρ). $\{A, C\} \mid \{B\}$ (teal, solid) outperforms all 2-state alternatives by 65–80× at $t = 5$ and improves further at $t = 20$. The fast-mode heuristic (light grey) and slow-mode heuristic (blue, dot-dash) perform no better than the wrong topological partitions, confirming that spectral contamination defeats both heuristics while $J(C)$ is unaffected. *Right:* The three-scale transport identity $\Delta_{C02} = C_{12}\Delta_{C01} + \Delta_{C12}C_{01}$ (Result 3) verified numerically for the tower $12 \to 3 \to 2$; residual 2.83×10^{-15} at machine precision.

[31] C. Jarzynski, Physical Review Letters **78**, 2690 (1997).
[32] G. Lindblad, Communications in Mathematical Physics **48**, 119 (1976).
[33] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Journal of Mathematical Physics **17**, 821 (1976).
[34] J. S. Bell, Physics Physique Fizika **1**, 195 (1964).
[35] S. Kochen and E. P. Specker, Journal of Mathematics and Mechanics **17**, 59 (1967).
[36] W. H. Zurek, Reviews of Modern Physics **75**, 715 (2003).
[37] M. Schlosshauer, Reviews of Modern Physics **76**, 1267 (2004).
[38] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, 1930).
[39] F. Bayen, M. Flato, C. Frønsdal, A. Lichnerowicz, and D. Sternheimer, Annals of Physics **111**, 61 (1978).
[40] M. Kontsevich, Letters in Mathematical Physics **66**, 157 (2003).
[41] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, arXiv preprint arXiv:2001.08361 (2020).
[42] J. Hoffmann, S. Borgeaud, A. Mensch, E. Buchatskaya, T. Cai, E. Rutherford, D. de Las Casas, L. A. Hendricks, J. Welbl, A. Clark, *et al.*, arXiv preprint arXiv:2203.15556 (2022).
[43] S. Watanabe, *Algebraic Geometry and Statistical Learning Theory* (Cambridge University Press, 2009).
[44] S. Watanabe, Journal of Machine Learning Research **14**, 867 (2013).
[45] C. Olsson, N. Elhage, N. Nanda, N. Joseph, N. DasSarma, T. Henighan, B. Mann, A. Askell, Y. Bai, A. Chen, *et al.*, Transformer Circuits Thread (2022).
[46] N. Elhage, T. Hume, C. Olsson, N. Schiefer, T. Henighan, S. Kravec, Z. Hatfield-Dodds, R. Lasenby, D. Drain, C. Chen, *et al.*, arXiv preprint arXiv:2209.10652 (2022).
[47] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Zhou, M. Goel, X. Wang, H. W. Chung, *et al.*, Transactions on Machine Learning Research 10.48550/arXiv.2206.07682 (2022).
[48] C. Wetterich, Physics Letters B **301**, 90 (1993).
[49] H. Touchette, Physics Reports **478**, 1 (2009).
[50] C. Rovelli, International Journal of Theoretical Physics **35**, 1637 (1996).
[51] C. M. Caves, C. A. Fuchs, and R. Schack, Physical Review A **65**, 022305 (2002).
[52] I. Csiszar, Studia Scientiarum Mathematicarum Hungarica **2**, 299 (1967).
[53] J. C. Baez and T. Fritz, Theory and Applications of Categories **29**, 421 (2014).
[54] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*, 2nd

ed. (Cambridge University Press, 2004).
[55] G. Birkhoff and J. von Neumann, Annals of Mathematics **37**, 823 (1936).
[56] A. Hobson, Journal of Statistical Physics **1**, 383 (1969).
[57] M. V. Berry, Proceedings of the Royal Society of London. Series A **392**, 45 (1984).
[58] B. Simon, Physical Review Letters **51**, 2167 (1983).
[59] A. M. Gleason, Journal of Mathematics and Mechanics **6**, 885 (1957).
[60] N. Elhage, N. Nanda, C. Olsson, T. Henighan, N. Joseph, B. Mann, A. Askell, Y. Bai, A. Chen, N. Conerly, *et al.*, Transformer Circuits Thread (2021).
[61] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, Journal of Fluid Mechanics **641**, 115 (2009).
[62] P. J. Schmid, Journal of Fluid Mechanics **656**, 5 (2010).
[63] I. Mezić, Nonlinear Dynamics **41**, 309 (2005).
[64] R. R. Coifman and S. Lafon, Applied and Computational Harmonic Analysis **21**, 5 (2006).
[65] J. Shi and J. Malik, IEEE Transactions on Pattern Analysis and Machine Intelligence **22**, 888 (2000).
[66] L. Wiskott and T. J. Sejnowski, Neural Computation **14**, 715 (2002).
[67] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, Journal of Computational Physics **151**, 146 (1999).
[68] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte, Linear Algebra and its Applications **315**, 39 (2000).