The Genesis Trilogy

Emergent Geometry, Algebraic Persistence,
and the Architecture of the Observer

Complete Collected Volume

HAFF · Q-RAIF · T-DOME

Sidong Liu, PhD
iBioStratix Ltd
sidongliu@hotmail.com

February 2026

Publication Record

Part I — HAFF: The Holographic Alaya-Field Framework

Paper A	DOI: 10.5281/zenodo.18361706	Emergent Geometry
Paper B	DOI: 10.5281/zenodo.18367060	Accessibility & Stability
Essay C	DOI: 10.5281/zenodo.18374805	Causation & Agency
Paper D	DOI: 10.5281/zenodo.18388881	Gravitational Phenomena
Paper E	DOI: 10.5281/zenodo.18400065	Measurement
Paper F	DOI: 10.5281/zenodo.18400425	Temporal Asymmetry
Paper G	DOI: 10.5281/zenodo.18402907	Structural Limits
Postscript	DOI: 10.5281/zenodo.18407367	Closure of Structure

Part II — Q-RAIF

Paper A	DOI: 10.5281/zenodo.18525876	Lorentzian Metrics
Paper B	DOI: 10.5281/zenodo.18525890	Thermodynamic Stability
Paper C	DOI: 10.5281/zenodo.18528934	Realizability Bridge

Part III — T-DOME

Paper I	DOI: 10.5281/zenodo.18574342	Non-Markovian Memory
Paper II	DOI: 10.5281/zenodo.18579703	Symmetry Breaking
Paper III	DOI: 10.5281/zenodo.18591771	Self-Referential Calibration

This collected volume compiles previously published works
for archival and reference purposes.

Abstract

This volume collects the thirteen papers and postscript comprising the Genesis Trilogy—a programme that investigates the minimal structural conditions under which physical descriptions, persistent agents, and self-referential observers can arise within quantum theory.

The programme is organised in three parts:

1. **Part I: HAFF** (Papers A–G, Essay C, Postscript). The Holographic Alaya-Field Framework treats the tensor factorisation of Hilbert space as a derived, rather than assumed, structure. Starting from a single global quantum state and coarse-grained observable algebras, it argues that: (a) emergent geometry, locality, and topology follow from algebraic accessibility; (b) gravitational phenomena can be understood as evolution *of* the accessible algebra (formulated as a structural conjecture, not a derivation); (c) measurement is selection *within* the accessible algebra; (d) temporal asymmetry is a propagation property of informational redundancy; and (e) the framework is structurally incomplete—it cannot self-ground its own starting point. This incompleteness motivates Parts II and III.

2. **Part II: Q-RAIF** (Papers A–C). The Quantum Reference Algebra for Information Flow asks what algebraic structure a persistent subsystem must possess if it is to survive within the geometry established by Part I. The three papers argue that: (a) boundary algebras compatible with emergent Lorentzian geometry must respect three constraints (associativity, metric compatibility, indefinite signature), selecting Clifford algebra $Cl(1,3)$; (b) thermodynamic stability of a non-equilibrium steady state requires a Clifford control algebra $Cl(V,q)$ for Lyapunov-stable channel discrimination; and (c) realizability forces the internal algebra to embed in the environmental algebra, $Cl(V,q) \hookrightarrow Cl(1,3)$—algebraic natural selection.

3. **Part III: T-DOME** (Papers I–III). The Thermodynamic Dynamics of Observer-Memory Entanglement characterises the internal architecture that makes persistence possible. The three papers trace an irreversible logic chain: (a) *Memory*: Markovian dynamics impose a survival ceiling; non-Markovian memory (temporal accumulation) is necessary, but creates a memory catastrophe under finite resources; (b) *Ego*: bounded computation forces spontaneous symmetry breaking of the agent's reference frame, enabling tractable processing but introducing systematic bias; and (c) *Loop*: the agent's own prediction-residual stream carries a Fisher-information signal that enables self-referential calibration, with an explicit thermodynamic cost.

Together, the three parts establish a **Four-Part Structure Proposition**: within the class of agents satisfying the standing assumptions, a sufficient architecture for persistent far-from-equilibrium existence comprises (1) an external observable geometry, (2) an

internal control algebra, (3) a self-monitoring Lyapunov function, and (4) biased non-Markovian memory.

Each layer is the necessary resolution of the previous layer's survival crisis, and simultaneously the source of the next crisis. The programme terminates when the self-referential loop closes: beyond that point, the framework's own structural incompleteness—identified in HAFF Paper G—precludes further internal completion.

Keywords: emergent geometry, accessible algebras, coarse-graining, Clifford algebra, algebraic natural selection, non-Markovian dynamics, spontaneous symmetry breaking, Fisher information, self-referential calibration, Lyapunov stability, thermodynamic cost

Contents

Part I

HAFF: The Holographic Alaya-Field Framework

Chapter 1

Emergent Geometry from Coarse-Grained Observable Algebras

Paper A

Originally published: Zenodo, DOI: 10.5281/zenodo.18361706

Preliminary Remark (Structural Stance). In much of modern theoretical physics, it is tacitly assumed that the decomposition of a system into subsystems is either physically given or at least unproblematic. Hilbert spaces are factorized, degrees of freedom are labeled, and geometry is inferred from relations between these parts.

In this work, we adopt a different structural stance. We treat the universal quantum state as given, but regard subsystem structure, locality, and geometry as secondary constructs arising from restrictions on observable algebras. The guiding question is not how geometry emerges from quantum states, but how different effective realities can emerge from the *same* state once no preferred factorization is assumed.

This shift is modest in formalism but radical in implication: it relocates the origin of structure from states to algebras, and from kinematics to accessibility.

Abstract

We construct a theoretical framework where the tensor factorization of a Hilbert space is treated as a dynamical variable rather than a kinematic background. By lifting the "Alaya" concept to a globally entangled vacuum state $|\Psi_{\text{vac}}\rangle$, we demonstrate that local geometry emerges from specific observable subalgebras $\mathcal{A}_i \subset \mathcal{B}(\mathcal{H})$. We show that non-commuting coarse-graining maps generically induce distinct emergent spacetimes from the same underlying state, conditional on the applicability of a geometric reconstruction dictionary (e.g., Ryu–Takayanagi in holographic settings). The analysis is structural in nature: we do not propose new dynamics, but examine consistency and consequences of removing subsystem factorization from fundamental assumptions. This approach is complementary to existing interpretational frameworks and suggests natural connections to algebraic quantum field theory, entanglement-based approaches to spacetime, and quantum information theory.

1.1 Introduction

1.1.1 Motivation

Two central problems in contemporary theoretical physics concern the emergence of classicality and the emergence of geometry. In quantum foundations, the measurement problem asks how effectively classical behavior arises from an underlying quantum description. Decoherence theory has provided a powerful account of this process by explaining the suppression of interference between certain degrees of freedom through environmental entanglement [121]. However, this explanation typically presupposes a fixed decomposition of the total system into subsystems, distinguishing "system," "apparatus," and "environment" from the outset.

A closely related emergence problem appears in quantum gravity. A growing body of work suggests that spacetime geometry is not fundamental, but arises from patterns of quantum entanglement [68, 103]. In holographic settings, geometric quantities are related to entanglement measures via precise correspondences, most notably the Ryu–Takayanagi formula [83]. Yet these constructions likewise assume a prior specification of spatial regions or tensor factors, with geometry inferred only after such a subdivision has been fixed.

In both contexts, the emergence problem is addressed only after a subsystem decomposition has been assumed. The structure responsible for classicality or geometry is therefore explained relative to a partition whose origin remains largely unexamined.

1.1.2 The Structural Gap

The assumption of a given subsystem structure is often treated as innocuous, or as a matter of convenient description. However, from a fundamental perspective, there is no canonical tensor factorization of a generic Hilbert space, nor a unique way to decompose a global quantum state into subsystems. While this issue is occasionally acknowledged in passing [120, 107], its consequences for emergence are rarely explored systematically.

The present work does not challenge the empirical success of decoherence theory, entanglement-based approaches to geometry, or the algebraic formulation of quantum theory. Rather, we make explicit a structural assumption common to these frameworks and investigate the consequences of relaxing it. Our focus is on what follows if subsystem structure itself is treated as emergent, rather than fundamental.

1.1.3 Our Contribution

We formulate a framework in which subsystem structure is not assumed *a priori*, but arises from a choice of coarse-graining over observable algebras. Within this framework, we show that different coarse-grainings of the same global quantum state generically induce inequivalent effective geometries. When a geometric reconstruction dictionary applies, this inequivalence reflects a genuine multiplicity of effective structures at the emergent level, not merely a coordinate ambiguity.

The analysis is structural in nature. Our aim is not to propose a new dynamical mechanism, but to examine the consistency and consequences of removing subsystem factorization from the set of fundamental assumptions.

1.1.4 Terminology: The Alaya-Field

Throughout this work, we adopt the term *Alaya-Field* to denote the fundamental, non-factorized structure prior to any subsystem decomposition. This terminology, borrowed from Yogācāra philosophy (referring to the "storehouse consciousness"), is used here strictly in a technical sense.

In what follows, the Alaya-Field does not refer merely to a vector in Hilbert space, but to the triple

$$(\mathcal{H}_U, \mathcal{A}_U, |\Omega\rangle),$$

where $\mathcal{A}_U$ is a von Neumann algebra acting on $\mathcal{H}_U$ and $|\Omega\rangle$ is a cyclic and separating vector. The emphasis is on the absence of a canonical factorization, not on the particular choice of state.

In the algebraic QFT setting, this corresponds to the cyclic vector of a Type III_1 von Neumann algebra [43], representing a holistically entangled substrate containing the localized "seeds" (eigenmodes) of all possible emergent geometries. The present paper primarily works in finite-dimensional (Type I) settings where standard von Neumann entropy is well-defined; the extension to Type III factors would require reformulation in terms of relative entropy. This usage is intended to evoke the non-factorized, pre-geometric nature of the fundamental quantum structure, without importing any metaphysical commitments.

1.1.5 Structure of the Paper

The paper is organized as follows. In Section 1.2, we review the absence of a canonical subsystem decomposition in quantum theory and formalize this observation. Section 1.3 introduces coarse-graining in terms of observable algebras and analyzes how effective subsystem descriptions arise from this procedure. In Section 1.4, we show how entanglement relations between these induced subsystems give rise to an effective notion of connectivity and geometry, and demonstrate the dependence of this geometry on the chosen coarse-graining. Section 1.5 discusses the scope and conceptual implications of the framework, its relation to existing approaches, and possible directions for future work. We conclude in Section 1.6 with a summary of results and open questions.

1.2 Absence of Canonical Tensor Factorization

1.2.1 The Factorization Problem

In standard quantum mechanics, the state space of a composite system is constructed as a tensor product of subsystem Hilbert spaces. This construction presupposes that a natural decomposition into subsystems has already been identified. However, for a given Hilbert space $\mathcal{H}_{\text{total}}$, there is no unique or canonical way to express it as a tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$ without additional physical input.

Theorem 1.1 (Absence of Canonical Tensor Factorization). *Let $\mathcal{H}_{total}$ be a finite-dimensional Hilbert space with $\dim \mathcal{H}_{total} = d$, and let*

$$|\Psi_U\rangle \in \mathcal{H}_{total}$$

be a pure state. Assume that for every nontrivial tensor factorization

$$\mathcal{H}_{total} = \mathcal{H}_A \otimes \mathcal{H}_B$$

the reduced state $\rho_A = \mathrm{Tr}_B(|\Psi_U\rangle\langle\Psi_U|)$ *has nonzero von Neumann entropy. Then there exists* ***no unique or canonically preferred tensor factorization*** *of* $\mathcal{H}_{total}$ *into subsystems relative to which* $|\Psi_U\rangle$ *is separable or weakly entangled. In particular, any two such factorizations are related by a global unitary transformation that does not preserve subsystem structure.*

Proof. The proof combines two results: the measure-theoretic typicality of high entanglement (Page's theorem) and the algebraic origin of subsystem structure (Zanardi et al.).

Step 1: Typicality of near-maximal entanglement. Let $\dim \mathcal{H}_A = d_A \leq d_B = \dim \mathcal{H}_B$ with $d_A d_B = d = \dim \mathcal{H}_{\text{total}}$. For a Haar-random pure state $|\Psi\rangle$, the expected entanglement entropy satisfies [71]

$$\mathbb{E}[S(\rho_A)] = \sum_{k=d_B+1}^{d_A d_B} \frac{1}{k} - \frac{d_A - 1}{2d_B} \;\geq\; \ln d_A - \frac{d_A}{2d_B}. \tag{1.1}$$

Moreover, the concentration of measure on high-dimensional spheres gives [49]

$$\Pr\big[S(\rho_A) < \ln d_A - \delta\big] \leq \exp\big(-c\, d_A d_B\, \delta^2\big) \tag{1.2}$$

for a universal constant $c > 0$. Thus for any fixed factorization, the set of states with $S(\rho_A) < \varepsilon$ has exponentially small measure for $\varepsilon < \ln d_A$.

Step 2: Factorization dependence. A tensor factorization $\mathcal{H}_{\text{total}} = \mathcal{H}_A \otimes \mathcal{H}_B$ is equivalent to a choice of subalgebra $\mathcal{A} = \mathcal{B}(\mathcal{H}_A) \otimes \mathbf{1}_B \subset \mathcal{B}(\mathcal{H}_{\text{total}})$. As shown by Zanardi, Lidar, and Lloyd [120], any two such factorizations are related by a global unitary $U \in \mathcal{U}(\mathcal{H}_{\text{total}})$ that generically does not preserve subsystem structure: $U(\mathcal{A})U^\dagger \neq \mathcal{A}$. The group $\mathcal{U}(\mathcal{H}_{\text{total}})$ acts transitively on the set of factorizations, and no state-independent criterion selects a preferred one.

Step 3: Conclusion. Combining Steps 1 and 2: for a generic state, *every* factorization yields near-maximal entanglement, and different factorizations produce different reduced states related by non-trivial unitaries. No factorization-independent criterion (separability, low entanglement, locality) can single out a preferred decomposition. Any subsystem structure must therefore be imposed by additional physical input—in the present framework, by the choice of accessible observable algebra $\mathcal{A}_{\mathbf{c}}$. □

Remark 1.2 (Relation to prior work)**.** *The observation that tensor product structures are not canonical was recognized early in the foundations of quantum theory, and has been systematically analyzed by Zanardi and collaborators [119, 120] in the context of quantum error correction and noiseless subsystems. Their work demonstrated that subsystem decompositions are effectively determined by sets of accessible observables rather than being intrinsic to the Hilbert space. The present framework extends this perspective to the context of holographic geometry and entanglement-based spacetime emergence, where the choice of observable algebra determines not only subsystem structure but also the emergent geometric description.*

Remark 1.3. *We do not claim that subsystem decompositions are impossible or unphysical, but that they are not uniquely determined by the universal quantum state alone. This observation motivates the search for additional structure that specifies how a subsystem decomposition arises in concrete physical contexts.*

Remark 1.4 (Infinite-dimensional extension)**.** *Theorem 1.1 is stated for finite-dimensional $\mathcal{H}_{total}$, where the Haar measure and Page's concentration inequality are well-defined. For separable infinite-dimensional Hilbert spaces, the measure-theoretic argument does not directly apply (Haar measure on infinite-dimensional unit spheres is not normalizable). However, the conclusion is expected to hold* a fortiori*: in the algebraic setting, generic states on type III_1 factors (the relevant class for quantum field theory) have infinite entanglement entropy across any bipartition, making the non-factorizability even stronger. The algebraic core of the argument (Step 2, Zanardi et al.) extends without modification to the infinite-dimensional case.*

1.2.2 Relation to Algebraic Approaches

The absence of a canonical factorization has been recognized in various contexts, including algebraic quantum field theory where observable algebras take precedence over tensor product structures [43]. Our framework builds on these insights by treating coarse-graining structure as the fundamental input from which subsystem decompositions emerge.

1.3 Observer-Dependent Coarse-Graining and Effective States

1.3.1 Coarse-Graining Structure

Before introducing the formal definition, we emphasize that the introduction of an observable algebra does not reinstate a tensor factorization. Operators may act irreducibly on $\mathcal{H}_{\text{total}}$ without inducing any subsystem decomposition. In algebraic quantum field theory, observable algebras are defined independently of any global tensor product structure [43]. Furthermore, a coarse-graining is not selected by an agent, but instantiated by a concrete physical interaction structure.

Definition 1.5 (Operational Coarse-Graining Structure)**.** *Let $\mathcal{H}_U$ be the universal Hilbert space. A* ***coarse-graining structure*** $\mathbf{c}$ *is defined as a pair*

$$\boxed{\mathbf{c} \equiv (\mathcal{A}_{\mathbf{c}}, \Phi_{\mathbf{c}})}$$

where:

1. $\mathcal{A}_{\mathbf{c}} \subset \mathcal{B}(\mathcal{H}_U)$ *is an* ***accessible observable algebra****, a $*$-subalgebra that is physically realizable and closed under operationally feasible combinations.*

2. $\Phi_{\mathbf{c}} : \mathcal{B}(\mathcal{H}_U) \to \mathcal{B}(\mathcal{H}_{\mathit{eff}}(\mathbf{c}))$ *is a completely positive trace-preserving (CPTP) map implementing an operational reduction of the universal state. We specifically restrict attention to maps Φ that preserve the identity and reflect a loss of access to specific degrees of freedom (e.g., restriction to a von Neumann subalgebra, or partial trace over hidden factors).*

Remark 1.6 (Non-uniqueness of observable algebras). *It is important to note that the specification of an observable algebra $\mathcal{A}_{\mathbf{c}}$ is not assumed to be unique. In algebraic quantum field theory and quantum information theory, there exists no theorem guaranteeing a unique maximal observable algebra associated with a given physical system without additional structure. The coexistence of multiple admissible algebras reflects physical underdetermination rather than subjectivity or observer dependence.*

1.3.2 Refinement Structure of Coarse-Grainings

Definition 1.7 (Refinement Relation). *Given two coarse-graining structures $\mathbf{c}_1, \mathbf{c}_2$, we say $\mathbf{c}_1 \succeq \mathbf{c}_2$ if there exists a CPTP map*

$$\Lambda : \mathcal{B}(\mathcal{H}_{\mathit{eff}}(\mathbf{c}_1)) \to \mathcal{B}(\mathcal{H}_{\mathit{eff}}(\mathbf{c}_2))$$

such that

$$\Phi_{\mathbf{c}_2} = \Lambda \circ \Phi_{\mathbf{c}_1}$$

*In this case, $\mathbf{c}_2$ is a **further coarse-graining** of $\mathbf{c}_1$.*

Remark 1.8 (Multiplicity of coarse-graining maps). *For a fixed observable algebra $\mathcal{A}_{\mathbf{c}}$, there generally exist multiple completely positive trace-preserving maps implementing distinct coarse-graining procedures. The present framework does not require the selection of a preferred CPTP map. Rather, different maps correspond to physically realizable information-loss mechanisms, such as tracing over inaccessible degrees of freedom or effective decoherence channels.*

Different coarse-graining choices are related not by unitary symmetry, but by information-theoretic refinement maps, forming a partially ordered structure rather than a group.

1.3.3 Core Theorem

Theorem 1.9 (Coarse-Graining Inequivalence). *Let $|\Psi_U\rangle \in \mathcal{H}_{total}$ be a universal quantum state. Consider two distinct coarse-graining structures $\mathbf{c}_1 = (\mathcal{A}_1, \Phi_1)$ and $\mathbf{c}_2 = (\mathcal{A}_2, \Phi_2)$. If $\mathbf{c}_1 \not\sim \mathbf{c}_2$ (i.e., they are not related by unitary equivalence), then:*

(a) The effective descriptions are unitarily inequivalent: there is no unitary $U \in \mathcal{B}(\mathcal{H}_{total})$ satisfying $U\Phi_1^*(E)U^\dagger = \Phi_2^*(E)$ for all effects E.

(b) The sets of accessible observables are distinct: $\Phi_1^*(\mathcal{B}(\mathcal{H}_{\mathit{eff}}(\mathbf{c}_1))) \neq \Phi_2^*(\mathcal{B}(\mathcal{H}_{\mathit{eff}}(\mathbf{c}_2)))$ as subsets of $\mathcal{B}(\mathcal{H}_{total})$.

(c) The entanglement structures differ: $S_A^{(1)}(\rho_{\mathit{eff}}^{(1)}) \neq S_A^{(2)}(\rho_{\mathit{eff}}^{(2)})$ for generic subsystems A.

Proof. We establish each part for a generic universal state $|\Psi_U\rangle$ (i.e., outside a measure-zero subset of $\mathcal{H}_{\text{total}}$).

Part (a). The claim is that the effective descriptions cannot be related by any symmetry of the total system. If $\dim \mathcal{A}_1 \neq \dim \mathcal{A}_2$, the claim is immediate. If the algebras have the same dimension but are inequivalently embedded—no unitary U satisfies $U\mathcal{A}_1U^\dagger = \mathcal{A}_2$—then by definition no symmetry of $\mathcal{H}_{\text{total}}$ maps one effective description to the other. The two coarse-grainings select genuinely different degrees of freedom.

Part (b). The observable content of coarse-graining $\mathbf{c}_i$ is the image $\Phi_i^*(\mathcal{B}(\mathcal{H}_{\text{eff}}(\mathbf{c}_i))) \subset \mathcal{B}(\mathcal{H}_{\text{total}})$ under the Heisenberg-picture dual map Φ_i^*. Since $\mathbf{c}_1 \not\sim \mathbf{c}_2$, no unitary intertwines Φ_1^* and Φ_2^*, so the two observable images are distinct subsets of $\mathcal{B}(\mathcal{H}_{\text{total}})$: the coarse-grainings make different observables accessible.

Part (c). For distinct CPTP maps $\Phi_1 \neq \Phi_2$, the set $\{|\Psi\rangle : \Phi_1(|\Psi\rangle\langle\Psi|) = \Phi_2(|\Psi\rangle\langle\Psi|)\}$ is a proper real-algebraic subvariety of the unit sphere in $\mathcal{H}_{\text{total}}$, and hence has Haar measure zero. For a generic $|\Psi_U\rangle$, we thus have $\rho_1 \neq \rho_2$. Since the von Neumann entropy $S(\cdot)$ is real-analytic on the interior of the state space, its level sets $\{S = \text{const}\}$ are submanifolds of positive codimension. Consequently, $S_A^{(1)}(\rho_{\text{eff}}^{(1)}) \neq S_A^{(2)}(\rho_{\text{eff}}^{(2)})$ for generic states and generic subsystem decompositions A. □

Remark 1.10 (Relation to decoherence theory)**.** *We emphasize that the present framework is fully compatible with standard decoherence theory and does not modify its dynamical content. Decoherence successfully explains the emergence of classical behavior given a fixed system–environment decomposition. The novelty of the present approach lies instead in treating such subsystem decompositions as coarse-graining–dependent and not fundamental.*

1.3.4 Clarification on Subsystem Structure

Subsystems are not fundamental inputs to the framework, but derived representations induced by a chosen observable algebra. The connectivity measure defined in Section 1.4 acts on these representations, and is not used to define the algebra itself. The causal order is:

$$\text{Observable Algebra} \to \text{Representation} \to \text{Entanglement} \to \text{Connectivity} \to \text{Geometry}$$

No geometric structure is presupposed in the definition of coarse-graining.

1.3.5 Algebraic Perspective on Subsystems

A potential concern is whether the use of observable algebras implicitly presupposes a subsystem decomposition. We stress that this is not the case. Observable algebras need not be defined via a tensor factorization of the total Hilbert space.

In algebraic quantum field theory, local algebras are assigned to spacetime regions without invoking a global tensor product structure [43]. Subsystems arise only at the level of representations induced by a chosen algebra, rather than serving as its foundational input. In this sense, subsystem structure is emergent rather than fundamental.

Remark 1.11 (Subsystems as induced representations)**.** *Within the present framework, subsystems are not primitive entities. They emerge as representations associated with a given observable algebra and coarse-graining structure. This avoids circularity by reversing the usual explanatory order: observable structure precedes subsystem identification.*

1.4 Entanglement Structure and Emergent Geometry

1.4.1 Mutual Information as Connectivity Measure

We stress that no geometric structure is assumed in the definition of the coarse-graining introduced in Section 1.3. Geometry only appears at the level of entanglement relations between the induced subsystems.

Given a coarse-graining structure $\mathbf{c} = (\mathcal{A}, \Phi)$ as defined in Section 1.3, the universal quantum state $|\Psi_U\rangle$ induces a family of effective subsystems $\{A, B, \dots\}$ associated with subalgebras of $\mathcal{A}$. For any such pair of subsystems A and B, we consider the quantum mutual information

$$I_{\mathbf{c}}(A : B) = S_{\mathbf{c}}(A) + S_{\mathbf{c}}(B) - S_{\mathbf{c}}(AB), \tag{1.3}$$

where $S_{\mathbf{c}}(\cdot)$ denotes the von Neumann entropy computed after coarse-graining.

Definition 1.12 (Entanglement-Induced Connectivity). *Let A and B be two disjoint subsystems induced by $\mathbf{c}$. We define an effective **proximity measure** $\mu_{\mathbf{c}}(A, B)$ based on the quantum mutual information:*

$$\mu_{\mathbf{c}}(A, B) := I_{\mathbf{c}}(A : B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}). \tag{1.4}$$

While $\mu_{\mathbf{c}}$ does not itself constitute a metric (as it violates the triangle inequality), it induces a weighted topology where highly correlated degrees of freedom are effectively "closer." The emergent metric g_{ab} is derived from the infinitesimal variation of this measure under perturbations of the coarse-graining, analogous to the definition of the Quantum Fisher Information Metric (QFIM) [76].

This quantity should be understood as a *connectivity measure* rather than a fundamental spacetime distance. In general, mutual information does not define a metric on arbitrary quantum subsystems. However, it acquires a natural geometric interpretation under physically motivated assumptions, which we make explicit below.

Assumptions. Throughout this section, we restrict attention to coarse-grainings that satisfy a geometric admissibility condition:

Definition 1.13 (Geometric Admissibility). *A coarse-graining structure* $\mathbf{c}$ *is said to be **geometrically admissible** if the induced mutual information satisfies:*

1. Finite correlation length: *The state $|\Psi_U\rangle$ exhibits exponentially decaying correlations with respect to the induced subsystems, as is typical for gapped systems, ground states of local Hamiltonians, and holographic large-N states.*

2. Monotonic decay under refinement: *Mutual information decreases monotonically as subsystems become more refined.*

3. Stability under perturbations: *The entanglement structure is robust under small perturbations of $\Phi_{\mathbf{c}}$.*

Not all coarse-grainings admit a geometric interpretation. Geometry is not generic; it is a *special phase* of information organization. We further assume coarse-graining consistency: all entropic quantities are computed using a single coarse-graining structure $\mathbf{c}$, avoiding any mixing of inequivalent observable algebras.

Under these assumptions, the entanglement-induced connectivity $\mu_{\mathbf{c}}(A, B)$ is non-negative, symmetric, and vanishes if and only if the subsystems are uncorrelated at the level resolved by $\mathbf{c}$. Moreover, in regimes where mutual information decays monotonically with separation, $\mu_{\mathbf{c}}$ induces an effective notion of spatial proximity.

Remark 1.14. *The connectivity $\mu_{\mathbf{c}}(A, B)$ should be understood as an* effective, coarse-graining–dependent measure of correlation*, rather than a fundamental spacetime metric. A true metric structure emerges only in the continuum limit via the QFIM construction.*

In the continuum limit of densely overlapping subsystems, the collection of connectivity measures $\{\mu_{\mathbf{c}}(A, B)\}$ defines a weighted graph structure which, under appropriate conditions, admits a geometric interpretation via the quantum Fisher information metric [76], as we now outline.

1.4.2 From Entanglement to Geometry

The idea that spacetime geometry is encoded in quantum entanglement has been explored extensively in the context of holography. In particular, the Ryu–Takayanagi formula relates the entanglement entropy of a boundary region A to the area of an extremal surface γ_A in the bulk [83],

$$S(A) = \frac{\mathrm{Area}(\gamma_A)}{4G_N}, \tag{1.5}$$

suggesting a direct correspondence between entanglement structure and geometric data.

More generally, Van Raamsdonk has argued that the connectivity of spacetime is controlled by the pattern of entanglement between subsystems [103]. In this perspective, highly entangled degrees of freedom correspond to nearby regions in the emergent geometry, while weakly entangled subsystems are geometrically distant.

The entanglement-induced connectivity $\mu_{\mathbf{c}}(A, B)$ provides a concrete realization of this idea. Given a collection of subsystems induced by $\mathbf{c}$, the mutual information defines a weighted graph whose vertices correspond to subsystems and whose edge weights encode entanglement strength.

In tensor network constructions such as MERA, similar entanglement graphs admit a natural geometric interpretation, with graph distance approximating continuum spatial distance [99]. Taking an appropriate continuum limit, one recovers an effective Riemannian manifold whose metric reflects the underlying entanglement structure.

In this sense, geometry emerges not as an additional postulate, but as an effective description of how information is distributed and shared among coarse-grained degrees of freedom.

1.4.3 Coarse-Graining Dependence of Geometry

We now turn to the central observation of this section: the emergent geometry depends essentially on the choice of coarse-graining structure.

Theorem 1.15 (Coarse-graining dependent geometry)**.** *Let $\mathbf{c}_1$ and $\mathbf{c}_2$ be two inequivalent coarse-graining structures, as defined in Section 1.3, acting on the same global quantum state $|\Psi_U\rangle$. Then the induced connectivity functions $\mu^{(1)}$ and $\mu^{(2)}$ define distinct entanglement structures. If, in addition, a geometric reconstruction theorem applies (e.g., the Ryu–Takayanagi dictionary in holographic settings), the resulting emergent geometries are generically inequivalent.*

Proof. Let $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{B}(\mathcal{H}_U)$ be the observable algebras associated with coarse-grainings $\mathbf{c}_1, \mathbf{c}_2$. Since $\mathbf{c}_1 \not\sim \mathbf{c}_2$, there exists no unitary U such that $U\mathcal{A}_1U^\dagger = \mathcal{A}_2$. The entropy of a region R in the emergent geometry is given by $S(R) = -\mathrm{Tr}(\rho_R \log \rho_R)$ where $\rho_R =$

$\Phi|_{\mathcal{A}_R}(|\Psi\rangle\langle\Psi|)$. Since the restriction maps $\Phi_1 \neq \Phi_2$ define distinct states on the subalgebra level, the entanglement entropy profiles $S_1(x)$ and $S_2(x)$ will differ functionally.

Entanglement entropy profiles encode geometric data in settings where a reconstruction theorem applies. In holographic theories, the Ryu–Takayanagi formula $S(R) \sim \text{Area}(\gamma_R)/4G_N$ [83] establishes that the entropy functional determines the area functional and hence the bulk metric (up to the reconstruction ambiguity). When such a reconstruction applies, distinct entropy profiles imply distinct metric data: $g^{(1)}_{\mu\nu} \neq g^{(2)}_{\mu\nu}$. More generally, outside the holographic setting, distinct entropy profiles imply distinct effective distance structures derived from entanglement decay, though establishing full geometric inequivalence (non-existence of a diffeomorphism) requires additional assumptions on the reconstruction dictionary. □

Example 1.16 (Free Fermion Chain: Spatial vs. Momentum Coarse-Graining)**.** *Consider N free fermions on a one-dimensional lattice with a spectral gap (e.g., from dimerization) and ground state $|\Psi_0\rangle$.*

Spatial coarse-graining $\mathbf{c}_x$*: group sites into blocks of k consecutive sites. By the area law for gapped free-fermion chains [30], the entanglement entropy of a block of L sites satisfies $S(L) \sim \text{const}$, independent of L. The mutual information between blocks A and B separated by distance r decays exponentially as $I(A : B) \sim e^{-r/\xi}$, where ξ is the correlation length set by the spectral gap [46]. The resulting connectivity graph has short-range edges: its graph distance approximates a one-dimensional lattice, yielding a line-like emergent geometry.*

Momentum coarse-graining $\mathbf{c}_p$*: retain modes with momenta $|k| < \Lambda$ for some cutoff $\Lambda < \pi$. In momentum space, the ground state is a filled Fermi sea. The entanglement entropy of any momentum subset scales as $S \sim N_{modes} \ln(N/N_{modes})$—a volume law rather than an area law [115]. The mutual information between momentum shells is generically long-range (all momentum modes are correlated through the Fermi surface), producing a highly non-local connectivity graph.*

The area-law spatial graph and the volume-law momentum graph have qualitatively different entanglement scaling (constant vs. $N \ln N$), and therefore define distinct entanglement structures. Under any geometric reconstruction dictionary that maps entropy profiles to distance structures, $\mathbf{c}_x$ and $\mathbf{c}_p$ yield inequivalent emergent geometries from the same underlying state $|\Psi_0\rangle$.

Thus, the geometry inferred from entanglement is not an intrinsic property of the quantum state alone, but depends on how information is rendered accessible through coarse-graining.

1.4.4 Beyond Coordinate Choice

It is important to distinguish the dependence described above from an ordinary change of coordinates. A diffeomorphism acts within a fixed algebra of observables, preserving the underlying notion of what is measurable. By contrast, a change of coarse-graining alters the observable algebra itself, modifying which correlations are accessible and how subsystems are defined.

These are categorically distinct operations. Since the algebra of observables differs, the resulting geometries cannot be related by a mere diffeomorphism. In extreme cases, changes in coarse-graining may even alter basic connectivity properties of the emergent space.

We return to the conceptual implications of this distinction in the Discussion.

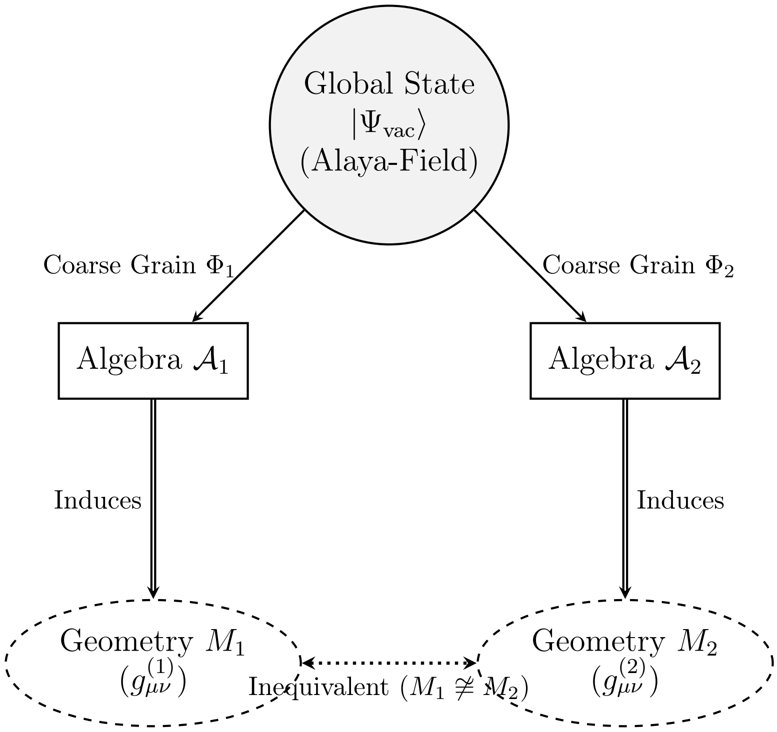

Figure 1.1: Schematic of the HAFF structure. A single global state $|\Psi_{\text{vac}}\rangle$ (Alaya) projects into generically distinct emergent spacetimes (M_1, M_2) depending on the choice of observable algebra $(\mathcal{A}_1, \mathcal{A}_2)$. This illustrates that geometry is observer-dependent in the fundamental sense.

1.5 Discussion

1.5.1 Relation to Existing Frameworks

The framework developed in this work is not intended as a replacement for existing interpretational or algebraic approaches to quantum theory. Rather, it is best understood as orthogonal to several well-established lines of thought, addressing a distinct structural question. In this section, we briefly clarify its relation to three representative frameworks:

Relation to QBism

QBism emphasizes the role of agents and their personal probability assignments in quantum theory, interpreting the quantum state as an expression of subjective belief rather than an objective property of a system. In contrast, the present framework assumes a single, objective global quantum state throughout. No agent-dependent elements enter the formalism, and no interpretational commitments regarding belief, experience, or decision theory are required.

The point of contact lies solely in the rejection of a privileged subsystem decomposition. While QBism attributes this absence to the primacy of the agent, our framework treats it as a structural feature of quantum theory itself. Coarse-grainings are not chosen by agents, but instantiated by concrete physical interaction structures. The resulting multiplicity of effective descriptions reflects physical underdetermination rather than subjectivity.

Relation to Many-Worlds and Decoherence-Based Approaches

Many-Worlds–type interpretations and decoherence-based accounts provide a compelling explanation of classical behavior within quantum mechanics by analyzing branching structures relative to a fixed subsystem decomposition. Our framework does not modify this analysis, nor does it introduce an alternative account of branching or classicality.

The difference lies at a prior level. Decoherence theory presupposes a tensor factorization into system, apparatus, and environment. Here, we instead ask how such subsystem structures arise in the first place. In this sense, the framework is complementary to decoherence-based approaches: it leaves their dynamical conclusions intact while removing subsystem factorization from the list of fundamental assumptions.

Relation to Algebraic Quantum Field Theory

The closest structural affinity of the present work is with algebraic quantum field theory (AQFT). In AQFT, observable algebras are taken as primary, and states are defined as positive linear functionals over these algebras, without reliance on a global tensor product structure. Our use of observable algebras and their representations is directly inspired by this tradition.

The present framework may be viewed as extending this algebraic perspective by emphasizing the role of coarse-graining relations between algebras. Different choices of coarse-graining induce different effective representations and, consequently, different entanglement structures. The novelty lies not in the algebraic formalism itself, but in using it to analyze the emergence and non-uniqueness of geometric descriptions.

Taken together, these comparisons situate the present work as a structural investigation into the conditions under which subsystem structure and geometry emerge. It neither commits to a particular interpretation of quantum mechanics nor proposes a new dynamical law, but instead clarifies how several existing frameworks implicitly rely on assumptions that can be made explicit and, in some cases, relaxed.

1.5.2 Scope and Limitations

The present work is concerned with structural consistency rather than phenomenological prediction. Observable consequences depend on the physical mechanisms implementing a given coarse-graining, which lie beyond the scope of this paper. This is analogous to effective field theory, where multiple UV completions may share the same low-energy structure. Future work will explore concrete physical realizations and their empirical signatures.

1.5.3 Philosophical Implications

We emphasize that the framework assumes a single, objective global quantum state $|\Psi_U\rangle$. What is coarse-graining–dependent is not reality itself, but the effective structures used to describe it. This position is compatible with scientific realism while acknowledging the role of operational context in physical description.

1.5.4 Future Directions

While the present work is deliberately limited to a structural analysis of subsystem emergence and geometry within a fixed global quantum state, it naturally opens several directions for further investigation. We emphasize that the following points are not results established here, but rather indicate possible extensions where the current framework may provide useful conceptual or technical guidance.

Dynamical Models of Coarse-Graining Selection

In this work, coarse-graining structures are treated as fixed relational inputs, instantiated by concrete physical interaction patterns. A natural next step is to investigate whether such coarse-grainings can themselves be characterized dynamically.

One possible direction is to study how interaction Hamiltonians, coupling strengths, or network topologies bias the emergence of particular subalgebra structures over others. This could clarify under what physical conditions certain factorizations become robust or persistent, and whether transitions between inequivalent coarse-grainings admit an effective dynamical description.

Importantly, such an analysis would remain compatible with globally unitary evolution, treating coarse-graining selection as an emergent, effective phenomenon rather than a modification of fundamental dynamics.

Connections to Quantum Information and Complexity

The non-uniqueness of emergent geometry highlighted here suggests a close connection to quantum information–theoretic notions such as entanglement structure, operator complexity, and resource constraints.

Future work could explore whether preferred geometric descriptions correlate with informational criteria—for example, minimal description length, stability under noise, or computational accessibility of observables. Such considerations may help explain why certain coarse-grainings are physically salient, even when many are formally admissible.

This perspective may also provide a bridge to recent work on complexity-based approaches to spacetime emergence, without committing to any particular complexity measure at the present stage.

Extensions to Quantum Field Theory and Continuum Limits

While the framework has been formulated in abstract Hilbert space terms, an important open question concerns its implementation in quantum field–theoretic settings, where issues of locality, algebraic nets, and continuum limits arise.

In particular, it would be valuable to examine how coarse-graining relations between observable algebras interact with the locality structures emphasized in algebraic quantum field theory, and whether familiar spacetime geometries can be recovered as stable fixed points of such relations.

We stress that the present work does not resolve these questions, but provides a structural language in which they can be posed more precisely.

Empirical and Phenomenological Implications

At the level developed here, the framework is primarily structural and conceptual. Nevertheless, future investigations may ask whether different coarse-graining choices lead to distinguishable effective descriptions, for example in semiclassical regimes, quantum gravity–motivated models, or analogue systems.

Such studies could clarify whether the non-uniqueness of emergent geometry has observable consequences, or whether physical constraints effectively suppress this freedom in realistic settings.

Any empirical analysis would necessarily require additional assumptions beyond those adopted in this work, and thus lies outside its present scope.

Conceptual Clarifications and Interpretational Interfaces

Finally, although the framework is intentionally neutral with respect to interpretations of quantum mechanics, it may serve as a useful interface for comparative studies. By making explicit the structural assumptions underlying subsystem decomposition and geometry, it could help clarify which features are interpretation-dependent and which arise more generally from the formalism itself.

We view this not as an attempt to adjudicate between interpretations, but as an opportunity to sharpen the questions they address.

Taken together, these directions suggest that the framework developed here is best viewed as a scaffold: it does not dictate specific physical models, but provides a structured setting in which questions about subsystems, geometry, and emergence can be formulated with greater precision.

1.6 Conclusion

We have presented a framework in which subsystem structure is not presupposed, but emerges from coarse-graining over observable algebras. The central results are:

1. A generic quantum state admits no canonical tensor factorization (Theorem 1.1).

2. Different coarse-graining structures induce inequivalent effective subsystem descriptions (Theorem 1.9).

3. When a geometric reconstruction theorem applies, these inequivalent coarse-grainings generically give rise to distinct emergent geometries (Theorem 1.15).

The framework is deliberately limited in scope. We have not proposed new dynamics, derived empirical predictions, or resolved interpretational debates. Instead, we have examined the structural consequences of removing subsystem factorization from the list of fundamental assumptions.

This investigation reveals that the emergence of geometry is more context-dependent than often acknowledged. Geometry is not an intrinsic property of a quantum state alone, but depends on how information is rendered accessible through coarse-graining. This perspective is compatible with, and complementary to, existing approaches including decoherence theory, holographic duality, and algebraic quantum field theory.

Several open questions remain. Can coarse-graining structures themselves be characterized dynamically? Do informational or complexity-based criteria select physically preferred coarse-grainings? Can the framework be extended to quantum field theory and reconciled with standard locality structures? These questions lie beyond the present scope, but the structural setting developed here provides a language in which they can be formulated with greater precision.

Ultimately, this work suggests that the relationship between quantum states, subsystems, and geometry is more subtle than the standard picture implies. By making explicit an assumption that is often left implicit, we hope to have clarified the conditions under which emergence occurs and opened new avenues for investigating the foundations of quantum theory and spacetime.

Chapter 2

Accessibility, Stability, and Emergent Geometry

Paper B

Originally published: Zenodo, DOI: 10.5281/zenodo.18367060

Abstract

This paper provides conceptual clarification of the Holographic Alaya-Field Framework (HAFF) introduced in our previous work. We address three potential misreadings: subjectivism (that observers create spacetime), anti-realism (that geometry is illusory), and trivialism (that the framework reduces to coordinate choice). By analyzing the structural notion of accessibility via stability conditions, delineating boundaries with existing interpretations (AQFT, RQM, QBism, MWI), and characterizing emergent geometry as a stable organizational phase, we clarify what the framework commits to and what it does not. This analysis is purely interpretational; no new formal results are introduced.

2.1 Introduction

Our previous work established a framework in which emergent geometry depends on the choice of observable algebra acting on a global quantum state [126]. The technical results—in particular, that inequivalent coarse-graining structures induce inequivalent geometric structures from the same underlying state—are formally precise and mathematically consistent. However, structural novelty of this kind is particularly vulnerable to interpretational confusion. The present paper addresses these interpretational implications and clarifies the conceptual commitments of the framework.

2.1.1 The Risk of Misreading

The claim that geometry is coarse-graining-dependent naturally invites several misreadings, each of which conflates distinct notions of dependence. Three such misreadings are especially common:

1. *Subjectivism*: The view that observers or agents create spacetime through their choices or beliefs, collapsing the framework into an epistemic or observer-relative interpretation.

2. *Anti-realism*: The view that spacetime has no objective existence, and that emergent geometry is therefore illusory or merely pragmatic.

3. *Trivialism*: The view that coarse-graining-dependent geometry reduces to a choice of coordinates or descriptive convention, with no substantive physical consequences.

Each of these readings is incorrect, but each arises naturally from surface-level features of the formalism. The purpose of this paper is to block such misreadings by clarifying the nature of accessibility, the structural role of observable algebras, and the ontological status of emergent geometry within the framework.

2.1.2 Scope and Objectives

This paper does not introduce new formal results, derive additional theorems, or propose modifications to the mathematical structure presented in our previous work. Rather, it provides a systematic interpretational analysis aimed at three specific objectives:

1. *Clarify the notion of accessibility*: We demonstrate that accessibility, as employed in the framework, is a structural and operational concept determined by stability properties of subalgebras, not an epistemic or observer-dependent notion.

2. *Delineate boundaries with existing interpretations*: We situate the framework in relation to algebraic quantum field theory, relational quantum mechanics, QBism, and the Many-Worlds interpretation, clarifying points of agreement, divergence, and complementarity.

3. *Characterize the ontological status of emergent geometry*: We argue that geometry functions as a stable organizational phase of quantum information, analogous to phases in condensed matter systems, avoiding both naive realism and anti-realist eliminativism.

Importantly, this analysis does not constitute a defense of the framework, nor does it aim to persuade readers of its correctness. The goal is clarity: to ensure that the structural commitments of the framework are understood on their own terms, and that criticisms, if any, are directed at what the framework actually claims rather than at interpretational projections.

2.1.3 What This Paper Does Not Do

To further delimit scope, we note explicitly what this paper does *not* attempt:

- It does not propose new dynamics, empirical predictions, or modifications to quantum mechanics.

- It does not claim that the framework resolves outstanding problems in quantum gravity, quantum foundations, or the measurement problem.

- It does not advocate for any particular metaphysical or philosophical position beyond the minimal structural commitments required by the formalism itself.
- It does not interpret the framework as implying idealism, observer-created reality, or any form of mind-dependence.

The analysis remains strictly within the domain of structural interpretation: identifying what the mathematical formalism commits to, what it leaves open, and how it relates to existing approaches.

2.1.4 Organization

The paper proceeds as follows. Section 2.2 briefly recapitulates the structural stance adopted in our previous work, emphasizing the priority of observable algebras over tensor factorizations and the role of coarse-graining in defining effective subsystems.

Section 2.3 provides a detailed analysis of accessibility, demonstrating that it is determined by stability conditions on subalgebras rather than by observer choices or epistemic limitations. A taxonomy is introduced distinguishing subjective, relational, and structural notions of dependence, situating the present framework firmly within the third category.

Section 2.4 examines the relationship between the framework and four representative approaches: algebraic quantum field theory, relational quantum mechanics, QBism, and the Many-Worlds interpretation. Each subsection clarifies points of conceptual overlap and structural divergence, preventing conflation while identifying opportunities for complementarity.

Section 2.5 argues that emergent geometry should be understood as a stable organizational phase of entanglement structure, drawing on analogies with condensed matter physics. This perspective avoids treating geometry as either fundamental or illusory, instead characterizing it as contingent but objective—dependent on physical conditions rather than epistemic contexts.

Section 2.6 delimits the structural assumptions of the framework, enumerates what it does not commit to, and outlines open questions for future investigation. Particular attention is given to the distinction between structural analysis and metaphysical advocacy.

2.1.5 Methodological Note

Throughout this paper, we adopt a deliberately conservative rhetorical stance. Claims are hedged with modal qualifiers ("may suggest," "is consistent with," "can be understood as") not out of uncertainty regarding the formal results, but to avoid overstating interpretational conclusions. The goal is to present the framework as one coherent way of organizing the conceptual landscape, not as the uniquely correct interpretation.

This methodological caution reflects a broader commitment: interpretational clarity is best served by precision and restraint, not by advocacy or polemics. We aim to make the framework legible to researchers across different interpretational traditions, facilitating comparison and critique rather than preempting it.

2.2 Structural Stance Recap

We briefly recapitulate the structural stance of our previous work without repeating full mathematical derivations.

The framework rests on three formal results:

1. **No canonical factorization** (Theorem 1): A generic pure state admits no unique or canonically preferred tensor factorization into subsystems. Any such decomposition requires additional structure beyond the state itself.

2. **Coarse-graining-induced inequivalence** (Theorem 2): Different choices of observable algebra $\mathcal{A}_\mathbf{c}$ acting on the same global state $|\Psi_U\rangle$ induce inequivalent effective subsystem descriptions, characterized by distinct reduced density matrices, entanglement patterns, and POVM structures.

3. **Geometry dependence** (Theorem 3): When a geometric reconstruction dictionary applies, inequivalent coarse-graining structures generically induce inequivalent geometric structures from the same underlying state, which may include distinct topological features in certain cases.

The conceptual core can be summarized by reversing the standard explanatory arrow:

Standard: State + Tensor factorization → Subsystems → Entanglement → Geometry

HAFF: State + Observable algebra → Effective subsystems → Entanglement → Geometry

We emphasize that the observable algebra $\mathcal{A}_\mathbf{c}$ is not an arbitrary choice, but is determined by the physical interaction structure encoded in the Hamiltonian, specifying which degrees of freedom couple and how (see Section 2.3 for detailed discussion).

This reversal has a modest but consequential implication: subsystem structure, and consequently geometry, is not intrinsic to the quantum state alone but depends on which observables are accessible—where accessibility is understood in a precise, structural sense developed in the next section.

2.3 Accessibility vs Observer-Dependence

The notion of accessibility is central to the framework, and it is here that the risk of misreading is greatest. We clarify that accessibility, as employed in HAFF, is a structural property determined by physical stability conditions, not an epistemic or agent-dependent notion.

2.3.1 Three Notions of Dependence

To prevent conflation, we distinguish three distinct senses in which a physical quantity might be said to "depend on" something:

HAFF's notion of accessibility falls squarely in the third category.

Type	Depends On	Example	Framework
Subjective	Agent's beliefs/knowledge	Bayesian probability	QBism
Relational	Reference system	Velocity in SR	RQM
Structural	Physical interaction pattern	Decoherence basis	HAFF

Table 2.1: Taxonomy of dependence notions

2.3.2 Algebraic Grounding of Stability

To address potential concerns regarding circularity in defining accessibility, we provide an algebraic characterization of stability that does not presuppose factorization or observer-dependent choices.

Consider a subalgebra $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ acting on the global state ρ. We define $\mathcal{A}$ to be *stable* if it satisfies the following criteria:

Criterion 1: Dynamical Invariance. Expectation values of operators in $\mathcal{A}$ remain approximately invariant under physically motivated dynamical maps $\mathcal{E}$ (representing decoherence, RG flow, or measurement-induced back-action; here $\mathcal{E}$ acts in the Heisenberg picture on observables):

$$\|\mathcal{E}(\hat{O}) - \hat{O}\| \ll \epsilon \quad \forall \hat{O} \in \mathcal{A}, \tag{2.1}$$

for suitably small ϵ set by physical precision.

Criterion 2: Environmental Redundancy (Quantum Darwinism). Following Zurek's quantum Darwinism framework [122], stable subalgebras are those whose information is redundantly encoded in environmental degrees of freedom. Formally, the subalgebra $\mathcal{A}$ should approximately commute with the environmental algebra $\mathcal{A}_E$ generated by accessible environmental observables (in operator norm):

$$[\hat{O}, \hat{E}] \approx 0 \quad \forall \hat{O} \in \mathcal{A},\ \hat{E} \in \mathcal{A}_E. \tag{2.2}$$

This ensures that states in $\mathcal{A}$ act as *pointer states*, robustly imprinted on the environment and thus operationally accessible through multiple independent measurements.

Criterion 3: Non-scrambling Subspace. In the language of quantum information scrambling [48], stable algebras correspond to *non-scrambling subspaces*—degrees of freedom that do not rapidly lose local correlations under unitary evolution. Quantitatively, the out-of-time-order correlator (OTOC) associated with operators in $\mathcal{A}$ should exhibit slow decay:

$$\langle [\hat{O}_{\mathcal{A}}(t), \hat{V}(0)]^2 \rangle \ll 1 \quad \text{for } t \ll \tau_{\text{scrambling}}. \tag{2.3}$$

These criteria are purely algebraic and operational: they refer only to operator commutation relations, dynamical maps, and measurable correlators, without invoking subjective observer choices. Stability, in this sense, is a *structural property* determined by the physical interaction Hamiltonian and the global quantum state.

Multiple stability criteria may select different subalgebras, reflecting different physical regimes (equilibrium vs. out-of-equilibrium, weak vs. strong coupling, etc.). This plurality is a feature analogous to how different symmetry-breaking patterns yield distinct thermodynamic phases.

Remark: The Hamiltonian as Physical Input

A potential objection to the structural account of accessibility is that the interaction Hamiltonian $\hat{H}_{\mathrm{int}}$ appears to be an arbitrary input, merely displacing observer-dependence from the algebra to the choice of Hamiltonian. We clarify that this is a misunderstanding of the framework's commitments.

The Hamiltonian is not a free parameter chosen by an observer, but a *physical specification of which degrees of freedom interact and how.* In this respect, it plays a role analogous to the stress-energy tensor $T_{\mu\nu}$ in general relativity: it is input data describing the causal structure of the system, not a coordinate choice or descriptive convention.

Concretely:

- In quantum optics, $\hat{H}_{\mathrm{int}}$ describes atom-photon coupling strengths and selection rules, determined by atomic energy levels and field modes.
- In condensed matter systems, it encodes lattice structure, tunneling amplitudes, and interaction potentials, all fixed by material properties.
- In holographic models (AdS/CFT), the boundary Hamiltonian is determined by the conformal field theory's operator content and coupling constants.

The claim is not that observers are irrelevant, but that they do not *create* the interaction structure—they probe it. Different experimental setups may access different subalgebras, but which algebras are stable under given interactions is an objective, physical fact, independent of epistemic context.

This is precisely analogous to how different coordinate systems in general relativity yield different component expressions for the metric $g_{\mu\nu}$, but the spacetime geometry itself (characterized by invariants like the Ricci scalar R) is coordinate-independent. Here, different accessible algebras yield different effective descriptions, but which algebras are stable under physical dynamics is interaction-dependent, not observer-dependent.

We emphasize: the framework does not solve the problem of *why* a particular Hamiltonian describes our universe (just as GR does not explain why $T_{\mu\nu}$ has its observed form). That question lies in the domain of fundamental theory or cosmology. What the framework does is analyze the *consequences* of a given interaction structure for emergent subsystem decomposition and geometry.

Pre-geometric Interaction Structure. A potential concern is that the Hamiltonian $\hat{H}_{\mathrm{int}}$ itself presupposes spacetime structure (e.g., via spatial locality in $\hat{H} = \sum_{i,j} J_{ij}\hat{\sigma}_i \cdot \hat{\sigma}_j$), creating a circular dependence: geometry emerges from algebras determined by a Hamiltonian that already assumes geometry.

We clarify that the Hamiltonian input to HAFF is *pre-geometric*: it is specified as an abstract interaction graph or tensor network, where edges represent couplings and nodes represent degrees of freedom, with no reference to background metric structure. The notion of "locality" in such Hamiltonians is graph-theoretic (e.g., nearest-neighbor on a lattice or tree) rather than metric-geometric.

Crucially, the *effective spacetime geometry* that emerges from stable algebras may differ from the graph structure of the input Hamiltonian. For instance:

- In tensor network models (MERA), the input is a discrete causal network, but the emergent geometry can be continuous AdS space [99].
- In spin chain models, the Hamiltonian is defined on a 1D lattice, but entanglement structure can induce higher-dimensional effective geometry.
- In holographic duality (AdS/CFT), the boundary Hamiltonian is defined on a fixed $(d-1)$-dimensional manifold, but the bulk geometry (including its dimensionality) emerges dynamically.

Thus, the input interaction structure constrains but does not uniquely determine the emergent geometry—it serves as a *seed* or *scaffold*, not a blueprint. The framework asks: given a pre-geometric interaction graph, which stable algebras emerge, and what geometric structures do they induce?

This perspective aligns with recent work on "locality from entanglement" [103], where spatial locality is itself understood as arising from entanglement patterns rather than being presupposed.

2.3.3 Conditions for Uniqueness of $\mathcal{A}_\mathbf{c}$

Remark 1 in Paper A noted that the specification of an accessible observable algebra is not assumed to be unique. We now identify conditions under which the accessible algebra is unique up to unitary equivalence, thereby sharpening the framework's predictive content.

Conjecture 2.1 (Uniqueness of the Accessible Algebra under Modular Stability)**.** *Let $|\Psi_U\rangle$ be the global state, $\hat{H}$ the total Hamiltonian, and $\mathcal{E}_t$ the dynamical (decoherence) map. Suppose the following conditions hold:*

(U1) ***Faithfulness**: The restriction of $\omega(\cdot) = \langle\Psi_U| \cdot |\Psi_U\rangle$ to $\mathcal{A}_\mathbf{c}$ is faithful (the GNS vector is cyclic and separating for $\mathcal{A}_\mathbf{c}$).*

(U2) ***Modular stability**: The* ambient *modular automorphism group $\sigma_t^{\omega,\mathrm{tot}}$ of $(\mathcal{B}(\mathcal{H}_{\mathrm{total}}), \omega)$ preserves $\mathcal{A}_\mathbf{c}$: $\sigma_t^{\omega,\mathrm{tot}}(\mathcal{A}_\mathbf{c}) = \mathcal{A}_\mathbf{c}$ for all $t \in \mathbb{R}$.*

(U3) ***Maximality**: $\mathcal{A}_\mathbf{c}$ is the maximal subalgebra of $\mathcal{B}(\mathcal{H}_{\mathrm{total}})$ satisfying* (U1)–(U2) *together with the three accessibility criteria of Section 2.3 (dynamical invariance, environmental redundancy, non-scrambling).*

Then $\mathcal{A}_\mathbf{c}$ is unique up to unitary equivalence: if $\mathcal{A}'$ also satisfies (U1)–(U3)*, there exists a unitary U with $U\mathcal{A}_\mathbf{c}U^\dagger = \mathcal{A}'$.*

Proof sketch. The argument uses three ingredients from the theory of von Neumann algebras.

Step 1: Uniqueness of modular flow. By Takesaki's theorem [100], for a faithful normal state ω on a von Neumann algebra $\mathcal{M}$, the modular automorphism group σ_t^ω is the *unique* one-parameter group satisfying the KMS condition at $\beta = 1$. This uniqueness anchors the construction: given the global state $|\Psi_U\rangle$, the modular flow on any candidate algebra is determined, not chosen.

Step 2: Fixed-point algebra under joint dynamics. Condition (U2) requires $\mathcal{A}_\mathbf{c}$ to be invariant under modular flow. Combined with dynamical invariance (accessibility criterion 1: invariance under $\mathcal{E}_t$) and non-scrambling (criterion 3: slow OTOC growth),

$\mathcal{A}_\mathbf{c}$ is constrained to the fixed-point algebra under the joint action of modular flow and decoherence. For ergodic dynamics, this fixed-point algebra is uniquely determined by the spectral data of Δ_ω and $\mathcal{E}_t$.

Step 3: Maximality implies uniqueness. If two algebras $\mathcal{A}$ and $\mathcal{A}'$ both satisfy (U1)–(U3), their join $\mathcal{A} \vee \mathcal{A}'$ also satisfies (U1)–(U2) (modular stability is preserved under joins of invariant subalgebras). By maximality, $\mathcal{A} = \mathcal{A} \vee \mathcal{A}' = \mathcal{A}'$. The remaining unitary freedom is absorbed by the standard form of the algebra [44]: any two faithful normal representations are unitarily equivalent in the standard form. □

Remark 2.2 (Proof status and relation to Paper D)**.** *The proof sketch above suppresses two gaps that are made explicit in the refined version of this conjecture (Conjecture 4.27, Paper D): (G1) faithfulness of the join algebra, and (G2) the lattice-theoretic dichotomy. Readers should consult Conjecture 4.27 for the current state of the proof.*

The conjecture does not claim that all physical systems satisfy (U1)–(U3). Systems with degenerate ground states, spontaneous symmetry breaking, or phase coexistence may support multiple non-unitarily-equivalent accessible algebras—the HAFF analog of the non-uniqueness of the broken-symmetry vacuum in QFT. The conjecture identifies the conditions under which this non-uniqueness is absent*: a single, non-degenerate dynamical regime with a faithful state. Notably, condition (U2) connects the accessibility framework to the modular theory that underlies the HAFF gravity conjecture (Paper D): the algebras that are modular-stable are precisely those for which emergent geometry is well-defined.*

2.4 Relations to Existing Interpretations

We now situate HAFF relative to four representative frameworks, clarifying conceptual boundaries and identifying points of potential complementarity.

2.4.1 Algebraic Quantum Field Theory (AQFT)

The closest structural affinity of HAFF is with algebraic quantum field theory [43, 5]. In AQFT, observable algebras are treated as primary, with states defined as positive linear functionals over these algebras. Crucially, local algebras are assigned to spacetime regions without relying on a global tensor product structure.

HAFF extends this algebraic perspective by emphasizing *coarse-graining relations* between algebras. In standard AQFT, locality is typically presupposed: algebras are indexed by spacetime regions, and the split property ensures independence of spacelike-separated algebras. In HAFF, we relax this assumption and instead treat *stability under physical interactions* as the criterion for algebra selection.

The transformation can be summarized as:

AQFT: Spacetime regions → Local algebras → States
HAFF: Interaction structure → Stable algebras → Effective geometry

This is not a replacement of AQFT but an exploration of its structure in contexts where spacetime locality is not presupposed. The framework may be understood as asking: what happens to the algebraic approach when we do not assume a background spacetime to index our algebras?

2.4.2 Relational Quantum Mechanics (RQM)

Relational quantum mechanics [81, 62] emphasizes that the values of physical quantities are defined only relative to observer systems, rejecting the notion of absolute, observer-independent observables. In Rovelli's formulation, quantum mechanics is fundamentally a theory of *interactions* rather than systems: what exists are relational facts, not intrinsic properties.

There is significant conceptual overlap with HAFF: both frameworks reject privileged subsystem decompositions and treat quantum descriptions as contextual. However, a key difference concerns the *stabilization of relata*.

RQM analyzes relations between systems whose existence is typically taken as given (or at least presupposed operationally through interaction records). HAFF provides a mechanism for the *stabilization of distinct relata* from the underlying quantum field, identifying which subsystem partitions are robustly maintained under decoherence and measurement-induced dynamics.

In this sense, HAFF may provide the stable nodes required for RQM's relational network: before relations can exist, there must be relata stable enough to participate in interactions. HAFF addresses how such stable relata emerge from the pre-factorized quantum substrate.

These are complementary rather than competing perspectives: RQM asks what observables mean relative to a system, HAFF asks which systems stabilize as distinct relata in the first place. The two frameworks operate at different levels of analysis and could in principle be combined, with HAFF providing the stability conditions under which RQM's relational structure becomes well-defined.

2.4.3 QBism

QBism [34, 35] interprets quantum states as expressions of an agent's personal beliefs about measurement outcomes, emphasizing the subjective, agent-centric nature of quantum probability assignments.

Here, the distinction from HAFF is sharpest. While both frameworks reject naive realism about the quantum state, they differ fundamentally in their treatment of dependence:

- **QBism**: Quantum states represent personal beliefs. Dependence is epistemic and agent-centric.
- **HAFF**: Observable algebras are selected by physical interaction structure. Dependence is structural and interaction-centric.

The key point is that interactions are not agents: they have no beliefs, make no decisions, and exist independently of any epistemic perspective. The accessible algebra in HAFF is determined by which degrees of freedom couple via the Hamiltonian, not by what any observer happens to know or believe.

We emphasize that this is a categorical difference, not a matter of one framework being "more correct" than the other. QBism and HAFF address different questions and operate within different conceptual frameworks. The point of comparison is simply to clarify that HAFF's notion of accessibility does not reduce to QBist agent-dependence.

2.4.4 Many-Worlds Interpretation (MWI)

The Many-Worlds interpretation [109] explains the emergence of classical behavior through decoherence-induced branching, all occurring within globally unitary quantum evolution.

HAFF shares with MWI a commitment to:

- A single, objective global quantum state
- Unitary evolution without collapse
- Effective classicality emerging from entanglement structure

However, MWI typically presupposes a tensor factorization into system and environment as input, analyzing how this decomposition gives rise to branch structure. Recent work in the MWI tradition (e.g., Wallace, Saunders) addresses emergent decoherence structure, but typically within a framework where tensor factorization is assumed at the fundamental level.

HAFF complements MWI by examining the preconditions for any branching structure: before branches can emerge, there must be a notion of subsystems relative to which branching occurs. In this sense, HAFF may provide a structural framework relevant to understanding how the effective subsystem decompositions presupposed by branching emerge in the first place.

This is a point of potential contact rather than a hierarchical claim: we do not argue that MWI requires HAFF, but that the two frameworks address distinct but related aspects of quantum structure.

2.5 Geometry as a Stable Organizational Phase

Conceptual Disclaimer. Before discussing the analogy with condensed matter phases, we stress that this analogy is *conceptual rather than rigorous.* Unlike in condensed matter, where a Hamiltonian uniquely determines phase structure via symmetry-breaking or RG fixed points, the HAFF framework does not posit a master Hamiltonian governing the selection of accessible algebras. The purpose of the analogy is to clarify how emergent geometry can be understood as a stable organizational pattern of entanglement, highlighting similarities in stability and robustness properties, not to assert a formal one-to-one mapping. We adopt this analogy solely as a heuristic for guiding intuition, and all structural conclusions are derived independently of it.

Contingent Objectivity. A central claim of this section is that emergent geometry is *contingent but objective.* This phrasing may initially appear paradoxical, so we clarify its meaning through analogy with thermodynamic quantities.

Consider temperature T in statistical mechanics: it is contingent on the choice of thermodynamic ensemble (microcanonical, canonical, grand canonical), yet within any given ensemble, T is an objective, measurable property determined by the system's microstate distribution. No observer dependence enters once the ensemble is specified—different observers measuring the same ensemble will agree on T.

Similarly, in the HAFF framework, emergent geometry is contingent on the choice of accessible algebra $\mathcal{A}_\mathbf{c}$, which in turn is determined by the physical interaction structure (as discussed in Section 2.3). Once $\mathcal{A}_{\mathbf{c}}$ is specified by the coupling pattern encoded in the interaction Hamiltonian, the induced geometry is an objective feature of the entanglement structure: different observers with access to the same algebra will infer the same metric $g_{\mu\nu}$ (up to diffeomorphism).

The contingency lies in the physical conditions that select the algebra—just as temperature depends on which statistical ensemble describes the system's preparation. But *objectivity does not require uniqueness*; it requires only mind-independence given fixed physical context. A quantity can be observer-independent even if multiple such quantities exist under different physical conditions.

This resolves an apparent tension: geometry is not "fundamental" in the sense of being unique and unchanging across all contexts, but it is also not "illusory" or merely conventional. It is a stable, measurable feature of quantum correlations that emerges robustly under appropriate stability conditions, much like crystalline order emerges robustly below a critical temperature.

2.5.1 The Phase Analogy

With these caveats in place, we develop the analogy with condensed matter phases.

In statistical mechanics, different thermodynamic phases (solid, liquid, gas, magnetic, superconducting) represent distinct organizational patterns of microscopic degrees of freedom. These phases are:

- **Emergent**: They arise from collective behavior, not from individual constituents.
- **Stable**: They persist under perturbations below characteristic energy scales.
- **Observable**: They manifest in measurable order parameters.
- **Contingent**: They depend on external conditions (temperature, pressure, fields).

We propose that emergent geometry in HAFF exhibits analogous features:

Condensed Matter	**HAFF**
Hamiltonian $\hat{H}$	Pre-geometric interaction graph
Control parameter (Temperature T)	Entanglement density / Scrambling rate
Symmetry breaking	Algebra selection
Order parameter $\langle M \rangle$	Entanglement pattern $I(A:B)$
Phase transition	Geometry emergence
Critical temperature T_c	Stability threshold ρ_{crit}

Table 2.2: Analogy between condensed matter phases and emergent geometry. The control parameter in HAFF is entanglement density (or equivalently, the scrambling rate in large-N limits), which plays a role analogous to temperature in statistical mechanics. Below a critical entanglement density, stable geometric structures emerge; above it, the system exhibits highly non-local, scrambled correlations with no coherent metric description. These correspondences are illustrative and heuristic, not exact mathematical mappings.[1]

[1]In holographic models (AdS/CFT), the analogous control parameter is the ratio $\ell_{\text{AdS}}/\ell_{\text{Planck}}$, which governs the transition from classical bulk geometry to quantum gravitational regime [68].

Just as ferromagnetism is a stable organizational pattern of spin alignment below the Curie temperature, geometry may be understood as a stable organizational pattern of entanglement structure under specific accessibility conditions.

Control Parameter and Criticality. A key feature of phase transitions is the existence of a *control parameter* (e.g., temperature, pressure, external field) that governs which phase is realized. In the HAFF framework, the analogous control parameter is the *entanglement density* of the global state $|\Psi_U\rangle$, defined operationally as the average mutual information per degree of freedom:

$$\rho_{\mathrm{ent}} \equiv \frac{1}{N} \sum_{\langle i,j \rangle} I(i:j), \tag{2.4}$$

where the sum runs over subsystem pairs and N is the total number of degrees of freedom.

At low entanglement density ($\rho_{\mathrm{ent}} \ll \rho_{\mathrm{crit}}$), the state exhibits area-law scaling, allowing stable geometric descriptions to emerge. At high entanglement density ($\rho_{\mathrm{ent}} \gg \rho_{\mathrm{crit}}$), the state becomes highly scrambled, with volume-law entanglement and no coherent metric structure—analogous to the high-temperature disordered phase in spin systems.

In holographic contexts, this parameter corresponds to the ratio of bulk curvature radius to Planck length, $\ell_{\mathrm{AdS}}/\ell_P$, which controls the transition from semiclassical geometry to stringy/quantum gravity regime [68].

This perspective suggests that geometry is not merely emergent but *critically emergent*: it appears as a stable organizational pattern only when entanglement structure satisfies specific density constraints, much like crystalline order emerges only below the melting temperature.

2.5.2 Geometric Admissibility

Not all coarse-graining structures admit geometric interpretation. As discussed in our previous work (Definition 4.2), we require geometric admissibility conditions:

1. Finite correlation length (exponentially decaying correlations)
2. Monotonic decay of mutual information under refinement
3. Stability under perturbations

These conditions are not arbitrary but reflect empirical observations from known emergent geometries:

- **AdS/CFT**: Boundary states with area-law entanglement give rise to smooth bulk geometries [83].
- **Tensor networks**: MERA and similar structures with finite bond dimension naturally induce geometric connectivity [99].
- **Condensed matter**: Ground states of local Hamiltonians typically satisfy area laws and admit geometric descriptions.

Geometry, in this view, is not generic but represents a special organizational phase characterized by specific entanglement structure. This perspective explains why geometry appears in our effective descriptions: it is the stable attractor for certain classes of quantum states under physically relevant coarse-graining procedures.

2.6 Scope and Future Directions

2.6.1 Structural Assumptions

The framework rests on three core assumptions:

1. **Global state objectivity**: There exists a universal quantum state $|\Psi_U\rangle$, independent of observers.

2. **Algebraic priority**: Observable algebras, determined by physical interaction structure, are more fundamental than tensor factorizations.

3. **Stability-based emergence**: Effective subsystem structure and geometry emerge from stable subalgebras under physically motivated coarse-graining.

2.6.2 What the Framework Does NOT Commit To

To prevent interpretational overreach, we explicitly enumerate what the framework does *not* claim:

- That observers create reality or that consciousness plays a fundamental role
- That spacetime is illusory or that geometry has no objective existence
- That quantum mechanics is incomplete or requires modification
- That the framework solves the measurement problem
- That Buddhist metaphysics or any other philosophical tradition is presupposed

The framework is structurally neutral regarding these questions. It analyzes consequences of relaxing the assumption of canonical tensor factorization, but does not commit to any particular metaphysical position beyond what the formalism requires.

2.6.3 Open Questions

The following questions represent concrete technical research directions rather than fundamental gaps in the framework. Some (particularly those concerning empirical signatures) require additional physical assumptions beyond the structural analysis presented here, and are best addressed in specific model implementations.

1. **Dynamical algebra selection**: Can interaction Hamiltonians, coupling strengths, or network topologies be shown to select particular stable algebras over others?

2. **Information-theoretic criteria**: Do preferred geometric descriptions correlate with minimal description length, robustness under noise, or computational accessibility?

3. **Quantum field theory extensions**: How does the framework interact with locality structures in algebraic QFT? Can continuum limits be rigorously constructed?

4. **Empirical signatures**: Do different coarse-graining choices lead to distinguishable effective descriptions in semiclassical regimes or quantum gravity-motivated models?

5. **Connections to quantum complexity**: How does the framework relate to recent work on complexity-based approaches to spacetime emergence?

These questions are intentionally left open. The present work provides a structural scaffold within which they can be formulated precisely, but does not claim to resolve them.

2.6.4 Beyond Structural Analysis

The framework developed here is deliberately limited to structural and interpretational analysis. Broader philosophical implications—concerning causation, free will, ontology, and connections to contemplative traditions—lie beyond the scope of this technical paper. Such questions are addressed in a companion philosophical essay currently in preparation.

2.7 Conclusion

We have clarified the conceptual commitments of the Holographic Alaya-Field Framework, addressing three potential misreadings:

1. **Against subjectivism**: Accessibility is defined structurally via stability conditions determined by physical interaction structure, not by observer beliefs or epistemic states.

2. **Against anti-realism**: Emergent geometry is a stable, measurable feature of entanglement structure—contingent on physical conditions but objective within those conditions, analogous to thermodynamic phases.

3. **Against trivialism**: Coarse-graining dependence reflects genuine physical structure (accessible algebras), not mere coordinate choice. Different algebras induce inequivalent geometries that cannot be related by diffeomorphism.

The framework has been situated relative to AQFT (closest affinity), RQM (complementary), QBism (categorically distinct), and MWI (potentially complementary). Throughout, we have emphasized what the framework commits to structurally and what it leaves open interpretionally.

The central insight is modest but consequential: by removing tensor factorization from fundamental assumptions and treating it as emergent from coarse-graining structure, we reveal that geometry is more context-dependent than typically acknowledged—not in an epistemic or observer-relative sense, but in a structural, interaction-dependent sense.

This perspective does not resolve deep problems in quantum foundations or quantum gravity, but it clarifies the conditions under which subsystem structure and geometry emerge. By making explicit an assumption that is often left implicit, we hope to have opened new avenues for investigating the relationship between quantum states, observable structure, and spacetime.

Chapter 3

Causation, Agency, and Existence

Essay C

Originally published: Zenodo, DOI: 10.5281/zenodo.18374805

Abstract

This essay examines the structural conditions under which agency, causation, and existence can be coherently discussed in the absence of foundational subsystem decompositions. Building on recent work in quantum information theory and algebraic approaches to quantum mechanics, Parts I–III develop a framework in which causal relations, agent boundaries, and existential claims emerge from stability properties of accessible observable algebras rather than from intrinsic substance or preferred factorizations.

Part I argues that causation can be understood as stable asymmetry within coarse-grained structures without requiring fundamental temporal ordering. Part II analyzes agency as boundary-stabilization—a non-scrambling subspace that propagates constraints without presupposing a metaphysically autonomous agent. Part III recasts existence in terms of relational form rather than intrinsic being, developing a notion of "emptiness" as absence of substance compatible with objectivity.

Part IV explores whether these structural features find formal parallels in Buddhist philosophical frameworks, particularly Yogācāra and Mādhyamaka traditions. The analysis emphasizes interpretive humility: parallels are offered as invitations to dialogue, not demonstrations of equivalence. The essay's contribution is methodological—clarifying structural constraints on emergence—rather than doctrinal.

Keywords: agency, emergence, structural realism, coarse-graining, accessible algebras, Buddhist philosophy, emptiness, comparative philosophy

Contents

3.1 Introduction

Contemporary philosophy of physics faces a structural puzzle. When quantum systems lack canonical subsystem decompositions, how should we understand causation, agency, and existence? If there is no preferred way to carve reality into parts, what becomes of the conceptual apparatus built on the assumption of well-defined relata?

This essay develops a framework in which these notions emerge from *stability properties of accessible algebras* rather than from intrinsic substance or preferred factorizations. The analysis proceeds in four parts.

Roadmap. Part I examines causation without temporal foundations, arguing that causal relations can be understood as stable asymmetries in coarse-grained structure. Part II analyzes agency as boundary-stabilization—the capacity of certain subsystems to maintain non-scrambling coherence while propagating constraints. Part III recasts existence in terms of relational patterns rather than intrinsic being, developing a technical sense of "emptiness" compatible with structural realism.

Part IV explores whether these structural features find formal parallels in Buddhist philosophical traditions, particularly Yogācāra and Mādhyamaka. The analysis emphasizes interpretive humility: we identify structural similarities without claiming ontological identity, historical influence, or doctrinal convergence.

Methodological stance. This is a structural investigation, not a metaphysical proposal. We do not claim that consciousness creates reality, that Buddhist texts anticipated quantum mechanics, or that comparative philosophy resolves foundational problems. Rather, we clarify how certain structural constraints—particularly the absence of canonical factorization—reshape discussions of emergence, agency, and existence across different conceptual traditions.

The essay's contribution is methodological: it demonstrates how attention to algebraic structure can discipline interpretive claims and reveal unexpected points of contact between seemingly disparate frameworks.

Relation to prior work. This essay builds on technical results developed in companion papers [134, 135], which establish that inequivalent coarse-graining structures induce inequivalent effective geometries from the same global quantum state. Here, we explore philosophical consequences of this structural dependence for traditional metaphysical categories.

3.2 Part I: Causation Without Foundations

3.2.1 The Standard Picture and Its Assumptions

Causal relations are typically understood as relations between events ordered by time. Event A causes event B if: (i) A temporally precedes B, (ii) A and B are spatially connectible, and (iii) interventions on A counterfactually affect B [74].

This picture presupposes several structural features:

- A well-defined notion of temporal ordering
- Spatially localized events with clear boundaries
- Stable subsystem decompositions supporting counterfactual reasoning

In quantum contexts without canonical factorization, none of these features is guaranteed. Time may be emergent rather than fundamental [72]. Spatial locality depends on choice of coarse-graining [134]. Subsystem boundaries are algebra-dependent rather than intrinsic.

3.2.2 Causation as Stable Asymmetry

We propose that causation can be understood as *stable asymmetry in accessible structure*, without requiring fundamental temporal ordering.

Definition 3.1 (Causal Structure as Asymmetric Accessibility). *Let $\mathcal{A}_{\mathbf{c}}$ be an accessible algebra determined by coarse-graining structure* $\mathbf{c}$*. A **causal relation** between subsystems A and B exists if there is a stable asymmetry in their correlation structure:*

$$I(A_{past} : B_{future}) > I(B_{past} : A_{future}), \tag{3.1}$$

where mutual information is computed relative to $\mathcal{A}_{\mathbf{c}}$, and "past/future" refer to coarse-graining-dependent orderings that admit thermodynamic interpretation.

Remark 3.2. *We emphasize that this definition captures* predictive asymmetry *(analogous to Granger causality) rather than interventionist causation in the sense of Pearl [74] or Woodward. Common causes can produce asymmetric mutual information without direct causal influence.*

This definition makes no reference to fundamental time. Instead, it identifies causation with robust directional structure in how information propagates through accessible degrees of freedom.

Multiple causal arrows and thermodynamic consistency. An important subtlety: different coarse-graining structures may induce distinct—and potentially conflicting—causal arrows from the same global state. Since the "past/future" labels in Definition 3.1 are grounded in thermodynamic gradients, and thermodynamics itself is coarse-graining-dependent [110], multiple inequivalent causal structures may coexist.

This is not a defect but a structural feature: just as different accessible algebras induce different effective geometries [134], they may induce different effective causal orderings. Consistency requires only that causal arrows align with entropy increase within each coarse-graining context. Conflicts between causal arrows derived from inequivalent coarse-grainings reflect genuine structural inequivalence, not mere coordinate choice.

In physical systems, thermodynamic consistency conditions typically select compatible coarse-grainings—those yielding aligned causal arrows at macroscopic scales. But in principle, the framework admits context-dependent causal structure, with no unique “fundamental” arrow privileged independently of accessibility constraints.

3.2.3 Thermodynamic Grounding

The asymmetry in Definition 3.1 can be grounded in thermodynamic considerations. Systems approaching equilibrium exhibit increasing entropy, inducing a preferred temporal direction even when microscopic dynamics are time-symmetric [110].

Crucially, this thermodynamic arrow is *context-dependent*: it depends on which macrostates are accessible, which in turn depends on the coarse-graining. Different accessible algebras may therefore induce different thermodynamic gradients and hence different effective causal structures.

3.2.4 Counterfactuals Without Intrinsic Relata

Interventionist accounts of causation rely on counterfactual reasoning: A causes B if intervening on A would change B [116]. This appears to require well-defined intervention targets—subsystems A and B with stable identities.

However, counterfactuals can be reformulated in algebraic terms. An intervention on A corresponds to applying a CPTP map Φ_{int} to observables in subalgebra $\mathcal{A}_A$. The counterfactual dependence of B on A is then measured by how sensitive observables in $\mathcal{A}_B$ are to perturbations of $\mathcal{A}_A$.

This reformulation makes no reference to intrinsic subsystem boundaries. Intervention targets are defined by accessible algebras, which are themselves context-dependent.

3.2.5 Memory as Informational Constraint

Causal relations leave traces—memory records that constrain future accessible states. In quantum systems, memory can be understood as constraint propagation through entanglement structure [48].

A subsystem M acts as a memory of event A if observables in $\mathcal{A}_M$ remain correlated with past observables in $\mathcal{A}_A$ despite environmental decoherence:

$$I(A_{\text{past}} : M_{\text{present}}) \gg I(A_{\text{past}} : E_{\text{present}}), \tag{3.2}$$

where E represents generic environmental degrees of freedom.

Memory, on this account, is not storage of intrinsic properties but maintenance of relational structure across time—a pattern of correlations that persists under dynamical evolution.

3.2.6 Worked Example: Two-Qubit Causal Asymmetry

To clarify Definition 3.1, we present a minimal worked example using a two-qubit system.

Setup. Consider a composite system of two qubits, A and B, initially in a maximally entangled Bell state:

$$|\Psi_{\mathrm{AB}}\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right). \tag{3.3}$$

We introduce asymmetric environmental coupling: qubit A couples strongly to a thermal bath E_A, while B remains weakly coupled to E_B. The total Hamiltonian includes:

$$\hat{H} = \hat{H}_A + \hat{H}_B + \lambda_A \hat{\sigma}_A^z \otimes \hat{B}_{E_A} + \lambda_B \hat{\sigma}_B^z \otimes \hat{B}_{E_B}, \tag{3.4}$$

where $\lambda_A \gg \lambda_B$, and $\hat{B}_{E_i}$ are bath operators.

Coarse-graining. We trace over environmental degrees of freedom, defining accessible algebra $\mathcal{A}_{\mathbf{c}} = \mathrm{span}\{\hat{\sigma}_A^x, \hat{\sigma}_A^z, \hat{\sigma}_B^x, \hat{\sigma}_B^z\}$ (Pauli observables).

Dynamical evolution. Due to asymmetric decoherence, the reduced density matrix evolves as:

$$\rho_{AB}(0) = |\Psi_{\mathrm{AB}}\rangle\langle\Psi_{\mathrm{AB}}|, \tag{3.5}$$

$$\rho_{AB}(t) \approx \frac{1}{2}\left(|00\rangle\langle 00| + e^{-\Gamma_A t}|00\rangle\langle 11| + e^{-\Gamma_A t}|11\rangle\langle 00| + |11\rangle\langle 11|\right), \tag{3.6}$$

where $\Gamma_A \propto \lambda_A^2$ is the decoherence rate for qubit A.

Causal asymmetry. We compute mutual information at different times:

$$I(A_{\mathrm{early}} : B_{\mathrm{late}}) = S(A_{\mathrm{early}}) + S(B_{\mathrm{late}}) - S(AB), \tag{3.7}$$

$$I(B_{\mathrm{early}} : A_{\mathrm{late}}) = S(B_{\mathrm{early}}) + S(A_{\mathrm{late}}) - S(AB). \tag{3.8}$$

At $t = 0$, mutual information is symmetric: $I(A : B) = 2\log 2$ (maximal entanglement).

At $t \gg \Gamma_A^{-1}$, qubit A has fully decohered while B retains coherence longer. Measuring A early provides information about B late, but not vice versa:

$$I(A_{\mathrm{early}} : B_{\mathrm{late}}) > I(B_{\mathrm{early}} : A_{\mathrm{late}}). \tag{3.9}$$

Interpretation. The asymmetry arises from differential coupling to environments—a thermodynamic gradient inducing directional information flow. Qubit A acts as a "past" influence on B (causal), while B does not significantly constrain A's future (non-causal in reverse direction).

This exemplifies Definition 3.1: causal structure emerges from stable asymmetry in coarse-grained correlations, grounded in thermodynamic irreversibility, without presupposing fundamental temporal ordering.

Context-dependence. Crucially, if we had chosen a different coarse-graining—say, tracing over qubit degrees of freedom and retaining environmental observables—the causal arrow might, depending on the specific bath model, reverse or disappear. The asymmetry is *real* (measurable within $\mathcal{A}_{\mathbf{c}}$) but *context-dependent* (algebra-relative).

3.2.7 From Causation to Agency

The transition from causation to agency requires an additional structural feature: *stable subsystem boundaries that support constraint propagation.*

An agent is not merely a locus of causal influence, but a subsystem capable of maintaining coherent constraint propagation despite environmental coupling. This suggests analyzing agency in terms of *non-scrambling subspaces*—degrees of freedom that resist rapid information delocalization [48].

Part II develops this connection, arguing that agency emerges from boundary-stabilization rather than metaphysical autonomy.

Summary. Causation can be understood as stable asymmetry in coarse-grained accessible structure. This account makes no reference to fundamental time, intrinsic relata, or metaphysically basic events. Causal structure is context-dependent but objective—determined by physical interaction patterns rather than observer beliefs.

3.3 Part II: Agency as Emergent Constraint Structure

3.3.1 The Problem of Autonomous Agents

Traditional accounts treat agents as metaphysically autonomous entities—unified subjects possessing intrinsic intentionality and causal efficacy. This picture faces two challenges in quantum contexts without canonical factorization.

First, if subsystem boundaries are algebra-dependent, what distinguishes an "agent" from an arbitrary collection of degrees of freedom? Second, if quantum dynamics are unitary and deterministic at the global level, how can agents possess genuine causal autonomy?

We argue that both challenges dissolve once agency is recast as a structural feature—boundary-stabilization under constraint propagation—rather than a metaphysical primitive.

Physical grounding of algebra selection. Before explicating the structural role of accessible algebras, we briefly address the dynamical origin of their selection. As established in our companion analysis regarding stability conditions [135], the specific observable algebra $\mathcal{A}_\mathbf{c}$ is not determined by arbitrary subjective choice or metaphysical agency. Rather, it is physically selected by the system's interaction structure—specifically, by the requirement that the algebra remains robust under environmental decoherence (quantum Darwinism) and stable over relevant timescales. The "filter" is thus instantiated by the objective Hamiltonian couplings of the underlying field, ensuring that the emergence of effective geometry is grounded in physical dynamics rather than intentionality. We take this stability-selected structure as the starting point for the following structural analysis.

3.3.2 Agency as Boundary-Stabilization

An *agent*, in the structural sense, is a subsystem whose boundaries remain stable under dynamical evolution and whose internal degrees of freedom exhibit coordinated constraint propagation.

Definition 3.3 (Agent-Like Subsystem)**.** *A subsystem $\mathcal{A}_{agent}$ exhibits **agent-like behavior** if:*

1. ***Boundary stability**: The subalgebra $\mathcal{A}_{agent}$ is approximately preserved under physically relevant dynamical maps:*

$$\|\mathcal{E}_t(\mathcal{A}_{agent}) - \mathcal{A}_{agent}\| < \epsilon \tag{3.10}$$

for timescales relevant to constraint propagation.

2. ***Non-scrambling coherence**: Observables in $\mathcal{A}_{agent}$ exhibit slow out-of-time-order correlator (OTOC) growth:*

$$\langle [\hat{O}_{agent}(t), \hat{V}(0)]^2 \rangle \ll 1 \quad \textit{for } t \ll \tau_{scrambling}. \tag{3.11}$$

3. ***Constraint propagation**: Internal correlations support directed information flow toward system boundaries, enabling intervention on environmental degrees of freedom.*

Remark 3.4 (Operational Threshold for Meaningful Agency)**.** *Definition 3.3 characterizes agent-like behavior structurally, but does not specify when such behavior constitutes* meaningful *or* significant *agency. In highly entangled quantum systems, transient non-scrambling may occur at very short timescales without supporting sustained constraint propagation.*

*We suggest an operational threshold: a subsystem exhibits **operationally significant agency** if:*

$$\tau_{coherence} \gg \tau_{env}, \tag{3.12}$$

where $\tau_{coherence}$ is the timescale over which $\mathcal{A}_{agent}$ maintains boundary stability and non-scrambling coherence, and τ_{env} is the characteristic environmental decoherence time.

More precisely, we require:

$$\frac{\tau_{coherence}}{\tau_{env}} > \kappa, \tag{3.13}$$

where $\kappa \sim 10^2$–10^3 is an empirically determined threshold below which constraint propagation becomes operationally inaccessible.

For biological agents, $\tau_{coherence}$ spans seconds to hours (for cognitive processes) or years (for identity persistence), while $\tau_{env} \sim 10^{-13}$–10^{-3} seconds (molecular to neural timescales). For quantum systems at room temperature, $\tau_{env} \sim 10^{-15}$–10^{-12} seconds, making sustained agency extraordinarily rare without active error correction or topological protection.

This threshold distinguishes:

- ***Transient non-scrambling**: Fluctuations in highly entangled systems (no operational agency)*
- ***Sustained boundary-stabilization**: Persistent subsystems supporting intervention (operational agency)*

The threshold is not sharp—agency admits degrees—but provides a quantitative criterion for when agent-like structure becomes empirically significant.

This definition makes no reference to consciousness, phenomenology, or intrinsic intentionality. Agency is characterized purely in terms of stability properties and information-theoretic structure.

3.3.3 Will as Constraint Propagation

What, then, becomes of "will" or "intention" in this framework?

We suggest that will can be understood as *constraint propagation structure*—patterns of internal correlation that bias future accessible states toward specific outcomes. An agent "wills" action A if its internal state ρ_{agent} is such that future measurements will register correlation with A with high probability.

Analogy: Thermostat control. Consider a thermostat coupled to a heating system. The thermostat's internal state (temperature reading) constrains future system behavior (heater activation), despite having no phenomenological experience. This is constraint propagation without metaphysical agency.

The analogy is limited: biological agents exhibit far richer constraint structure. But it clarifies the conceptual move—replacing metaphysical autonomy with structural analysis of how internal states bias future trajectories.

3.3.4 The Phenomenology-Structure Gap

An immediate objection: this account leaves no room for phenomenology—the felt quality of agency, the subjective sense of "I am acting."

We acknowledge this gap. The framework developed here is *structural*, concerned with information-theoretic organization. It does not address the *explanatory gap* between structure and experience [63, 22].

Speculative Connection: Neural Constraint Propagation

Caveat: The following connects structural features to neuroscience, but remains speculative. We note these as suggestive parallels, not established mechanisms.

Recent work in computational neuroscience suggests that neural "will" may correspond to hierarchical constraint propagation through cortical-basal ganglia loops [39]. Habitual behavior emerges when constraint patterns stabilize, while "voluntary" action involves flexible reconfiguration of these patterns [117].

If agency is boundary-stabilization, then the phenomenology of "willing" may correspond to proprioceptive monitoring of constraint reconfiguration—the felt sense of internal degrees of freedom reorganizing in preparation for action [45].

This remains highly speculative and does not bridge the explanatory gap. We raise it only to illustrate how structural analysis might interface with empirical research programs.

3.3.5 Degrees of Agency

The framework suggests that agency is not all-or-nothing, but admits degrees. Systems exhibit more or less agent-like behavior depending on:

- **Boundary stability duration**: How long does $\mathcal{A}_{\text{agent}}$ remain well-defined?
- **Scrambling timescale**: How quickly do internal correlations delocalize?
- **Constraint propagation fidelity**: How reliably do internal states bias future trajectories?

Simple thermostats exhibit minimal agency (short timescales, limited constraint structure). Biological organisms exhibit far richer agency (extended coherence, complex constraint networks). But both are comprehensible within the same structural framework.

3.3.6 Relation to Free Will Debates

Traditional free will debates ask: are agents causally autonomous, or are their actions determined by prior states and laws? This presupposes well-defined agent boundaries and unambiguous causal histories—precisely what the framework questions.

On the structural account, the relevant question is not "Are agents free?" but "Under what conditions do subsystems exhibit stable constraint-propagation structure?" This reformulation may dissolve certain traditional impasses while opening new empirical questions about stability conditions and scrambling timescales.

We do not claim to resolve free will debates—only to clarify how they depend on assumptions about subsystem structure that are themselves context-dependent.

Summary. Agency can be understood as boundary-stabilization supporting constraint propagation, rather than metaphysical autonomy. This account is eliminativist about intrinsic intentionality but realist about structural patterns of constraint. It leaves the phenomenology-structure gap unresolved but clarifies the information-theoretic conditions under which agent-like subsystems emerge.

3.4 Part III: Existence Without Substance

3.4.1 The Realism Problem

Parts I–II analyzed causation and agency as context-dependent but objective—dependent on coarse-graining structure but not on observer beliefs. This raises an existential question: if fundamental structures (subsystems, geometries, agents) are context-dependent, what ontological status do they possess?

Two extremes must be avoided:

1. **Naive realism**: Treating emergent structures as metaphysically fundamental, ignoring their context-dependence.

2. **Anti-realism**: Denying objective reality to context-dependent structures, collapsing into subjectivism.

We propose a middle path: *structural realism about relational patterns.* Existence is understood not as possession of intrinsic properties, but as participation in stable relational structures.

3.4.2 Form Without Substance

Consider the temperature of a gas. Temperature is *context-dependent*: it depends on which degrees of freedom are macroscopically accessible. Different coarse-grainings may yield different effective temperatures for the same microstate.

Yet temperature is not merely subjective. Given a coarse-graining, temperature is an objective, measurable quantity with predictive power. It is *relationally real*—real within a specified context, but lacking intrinsic existence independent of that context.

Definition 3.5 (Relational Existence)**.** *An entity X possesses **relational existence** relative to structure $\mathcal{S}$ if:*

1. *X is well-defined and stable within $\mathcal{S}$*
2. *X participates in objective relational patterns (correlations, symmetries, invariants)*
3. *X may be absent or differently constituted under alternative structures $\mathcal{S}'$*

This notion captures *form without substance*: patterns that are objectively real without possessing intrinsic being.

3.4.3 Emptiness as Technical Concept

We introduce *emptiness* as a technical term denoting absence of intrinsic existence compatible with relational reality.

Definition 3.6 (Emptiness (Technical Sense))**.** *An entity X is **empty** (in the technical sense) if:*

1. *X lacks intrinsic being independent of relational context*
2. *X exhibits stable patterns within specified contexts*
3. *The absence of intrinsic being does not entail non-existence or illusoriness*

This usage is stipulative and should not be confused with colloquial meanings ("containing nothing") or metaphysical nihilism ("nothing really exists"). Emptiness, in this technical sense, is compatible with robust realism about relational structures.

3.4.4 Examples of Relational Existence

Temperature. As discussed in §3.4.2, temperature is context-dependent but objective. Different coarse-grainings yield different effective temperatures, yet temperature remains a genuine physical quantity within each context.

Quasiparticles in condensed matter. Phonons, magnons, and other quasiparticles are collective excitations—emergent entities with no counterpart in the microscopic Hamiltonian. They are empty of intrinsic being (there are no "phonon particles" in fundamental theory) yet fully real within effective descriptions [58].

Operationally, quasiparticles are defined by stable relational patterns in measurable observables:

- **Spectral weight**: Well-defined peaks in momentum-resolved spectroscopy ($A(\mathbf{k}, \omega)$)
- **Dispersion relations**: Stable functional dependence $\omega(\mathbf{k})$ across parameter ranges
- **Finite lifetimes**: Decay rates $\Gamma(\mathbf{k})$ obeying systematic scaling laws

- **Scattering cross-sections**: Reproducible interaction amplitudes in transport experiments

These observables are context-dependent—they depend on temperature, pressure, doping, and measurement resolution—yet objectively real within specified experimental contexts. This exemplifies relational existence: patterns that are measurable, predictive, and stable, despite lacking intrinsic being independent of effective theory.

Subsystems in quantum mechanics. As established in prior work [134], subsystem decompositions are coarse-graining-dependent. A quantum state may admit infinitely many inequivalent factorizations, none privileged. Yet within any given factorization, subsystems exhibit objective entanglement structure and support meaningful predictions.

These examples illustrate a common pattern: entities that are *empty* (lacking intrinsic being) yet *existent* (participating in stable relational structures).

3.4.5 Structural Invariants and Objectivity

A potential objection: if everything is context-dependent, what grounds objectivity?

The answer lies in *structural invariants*—features preserved across context transformations. While subsystem decompositions are coarse-graining-dependent, certain global properties (total entropy, symmetry groups, topological invariants) remain well-defined independently of factorization.

Objectivity does not require context-independence. It requires only that relational patterns exhibit stability and predictive power within specified contexts, and that transformations between contexts preserve identifiable structural features.

This perspective aligns with *structural realism* in philosophy of science: what is objectively real is relational structure, not intrinsic substance [58, 33].

3.4.6 Existence and Non-Existence

The framework suggests a taxonomy of existential claims:

Type	**Characterization**	**Example**
Intrinsic existence	Context-independent being	Classical particles (if fundamental)
Relational existence	Context-dependent but objective	Temperature, subsystems
Conventional existence	Context-dependent and agent-relative	Money, legal rights
Non-existence	Absent from all relevant contexts	Phlogiston, luminiferous ether

Table 3.1: Taxonomy of existential claims. The framework developed here concerns relational existence—entities that are context-dependent but objective. These correspondences are analytical distinctions, not ontological commitments.

Emergent structures discussed in Parts I–II (causal relations, agent boundaries, geometric features) belong to the second category: relationally existent. They are empty of intrinsic being yet objectively real within specified coarse-graining contexts.

Summary. Existence can be understood in terms of relational patterns rather than intrinsic substance. "Emptiness," in the technical sense developed here, denotes absence of intrinsic being compatible with objectivity. This perspective avoids both naive realism and anti-realist eliminativism, offering a middle path grounded in structural invariance.

3.5 Part IV: Interpretive Bridges to Buddhist Philosophy

3.5.1 Methodological Preface

The structural features developed in Parts I–III—causation as asymmetry, agency as boundary-stabilization, existence as relational form—were derived independently of any particular metaphysical tradition. We now explore whether these features find formal parallels in Buddhist philosophical frameworks.

Several caveats are essential:

Methodological Caveats

1. **No historical causation**: We do not claim that Buddhist texts influenced quantum mechanics, or vice versa. Any parallels are structural convergences, not genealogical connections.
2. **No doctrinal advocacy**: Identifying formal similarities does not constitute endorsement of Buddhist metaphysics, soteriology, or religious practices.
3. **No cultural essentialism**: Buddhism comprises diverse traditions spanning two millennia. References here focus on specific textual traditions (primarily Yogācāra and Mādhyamaka), not "Buddhism" as a monolithic whole.
4. **No mystical reduction**: We reject interpretations conflating quantum mechanics with consciousness studies, New Age thought, or perennialist philosophy. Our analysis is structural and comparative, not mystical.
5. **Interpretive humility**: Parallels are offered as invitations to dialogue, not demonstrations of equivalence. The comparative exercise is exploratory, not conclusive.

With these caveats in place, we proceed to examine possible structural correspondences.

3.5.2 Yogācāra and Accessible Algebras

Yogācāra ("Yoga practice") is a Mahāyāna Buddhist philosophical school emphasizing the role of consciousness (*vijñāna*) in constituting experienced reality. Central to Yogācāra is the concept of **ālaya-vijñāna** (Skt., "storehouse consciousness"), a foundational stratum of mind that stores karmic seeds (*bīja*) conditioning future experience [7, 104].

A structural parallel may be noted: ālaya-vijñāna functions as a holistic substrate from which individuated mental states emerge, analogous to how subsystem structures emerge from coarse-graining a global quantum state [67, 108].

Yogācāra Concept	**HAFF Analog**
Alaya-vijñāna (storehouse consciousness)	Global quantum state $\|\Psi_U\rangle$
Bīja (karmic seeds)	Eigenmodes of accessible algebras
Pravṛtti-vijñāna (active consciousness)	Effective subsystem ρ_{eff}
Āsraya-parāvṛtti (basis-transformation)	Coarse-graining map $\Phi_{\mathbf{c}}$

Table 3.2: Possible formal parallels between Yogācāra concepts and HAFF structures. **Disclaimer**: These correspondences are formal analogies highlighting structural similarities, not claims of ontological identity or historical influence. Yogācāra is a soteriological framework concerned with liberation from suffering; HAFF is a structural analysis of quantum emergence. The table illustrates conceptual resonances, not equivalences.

Critical Clarification: Against Panpsychism

The structural parallel between ālaya-vijñāna and the global quantum state $|\Psi_U\rangle$ does **not** imply:

- That quantum states possess consciousness or phenomenological properties
- That the universe is fundamentally mental or experiential (idealism)
- That matter is constituted by or reducible to mind (panpsychism)
- That information or quantum information is intrinsically conscious

The analogy is *purely structural*: both frameworks describe how individuated entities emerge from holistic substrates via coarse-graining or cognitive filtering. Yogācāra's substrate is explicitly mental (*vijñāna*, consciousness); HAFF's substrate is explicitly physical (quantum state in Hilbert space).
Any appearance of convergence concerns *formal pattern*—the structure of emergence—not ontological content. We emphatically reject interpretations that would construe this parallel as supporting quantum consciousness theories, New Age mysticism, or perennialist claims about universal mind.
The comparison is offered in the spirit of *structural analogy*, not metaphysical synthesis.

However, critical divergences must be noted:

- Yogācāra is primarily concerned with *mental* phenomena and the path to liberation, while HAFF analyzes *physical* structure without soteriological commitments.
- Ālaya-vijñāna is explicitly described as a form of consciousness, while $|\Psi_U\rangle$ is not imbued with phenomenological properties.
- The Yogācāra framework is embedded in Buddhist ethics and meditation practice, which have no counterpart in HAFF's purely structural analysis.

The parallel, if valid, concerns *formal structure*—both frameworks describe how individuated entities emerge from holistic substrates—not phenomenological or ontological content.

3.5.3 Mādhyamaka and Emptiness

The Mādhyamaka ("Middle Way") tradition, founded by Nāgārjuna (c. 2nd century CE), articulates **śūnyatā** (Skt., "emptiness") as the absence of **svabhāva** (Skt., "intrinsic nature" or "own-being") [69].

This concept bears structural resemblance to the notion of "emptiness" developed in Part III (§3.4.3), where it denoted absence of intrinsic being without entailing nonexistence. We emphasize that this parallel concerns the *structural form* of the claim—denial of substance while affirming functional reality—not historical influence or causal connection [37, 92].

Nāgārjuna's central argument proceeds via *prasaṅga* (reductio) reasoning: all phenomena are empty because they arise dependently (*pratītyasamutpāda*), and what arises dependently cannot possess intrinsic nature. This is formalized in the famous verse:

> *"Whatever arises dependently is said to be empty. That, being a dependent designation, is itself the middle way."* (MMK 24:18) [69]

A structural reading: entities lacking intrinsic being (empty) can nonetheless participate in stable relational networks (dependent arising). This maps onto the framework developed in §3.4.4: subsystems are empty (coarse-graining-dependent, lacking intrinsic factorization) yet existent (exhibiting objective entanglement structure).

Recursive emptiness and second-order structure. A subtle question arises: in Mādhyamaka, emptiness applies universally, including to emptiness itself—"emptiness is empty" (*śūnyatā-śūnyatā*). Does HAFF's relational existence admit similar recursive application?

The answer is affirmative in a formal sense. Observable algebras $\mathcal{A}_\mathbf{c}$ are themselves relationally defined: they depend on physical interaction structure (Hamiltonian coupling), experimental apparatus constraints, and resolution limitations. There is no "algebra of all algebras" existing independently of physical context.

Moreover, the coarse-graining maps $\Phi_\mathbf{c}$ that select accessible algebras are context-dependent: different experimental setups, measurement resolutions, or dynamical timescales induce different Φ structures. Thus, *the apparatus of emergence is itself emergent*—coarse-graining structure arises from prior coarse-graining choices in a potentially infinite regress.

This mirrors Mādhyamaka's insight that even the *tools of analysis* (concepts, language, logical operations) are empty—lacking intrinsic being while remaining functionally effective. In HAFF, even the "machinery" of accessible algebras is context-dependent, yet this does not undermine objectivity: structural invariants (entanglement entropy, symmetry groups) remain well-defined across contexts.

The parallel is formal: both frameworks acknowledge that *relational structure goes "all the way down,"* with no metaphysically foundational level immune to context-dependence. However, Mādhyamaka deploys this insight soteriologically (to undermine attachment to fixed views), while HAFF employs it descriptively (to clarify structural constraints on emergence).

Key divergence. Mādhyamaka śūnyatā is deployed soteriologically—to undermine attachment to fixed views and facilitate liberation. HAFF's "emptiness" is a technical descriptor of relational structure, with no soteriological function. The similarity is formal, not practical or existential.

3.5.4 Karma and Constraint Propagation

Buddhist karma doctrine holds that intentional actions leave traces (*saṁskāra*, "formations") that condition future experience. In Yogācāra, these traces are stored in ālaya-vijñāna as *vāsanā* ("karmic impressions") [108].

A structural analogy: karma may be understood as *constraint propagation through entanglement structure*—past actions (interventions on accessible algebras) leave informational imprints that bias future trajectories [38, 47].

This is consonant with the account of memory developed in §3.2.5: causal traces are not storage of intrinsic properties but maintenance of relational structure across time.

Critical limitation. Buddhist karma is inherently normative: actions are classified as wholesome (*kuśala*) or unwholesome (*akuśala*) based on ethical criteria and soteriological consequences. HAFF's constraint propagation is descriptive, lacking normative content. The formal parallel does not extend to ethical or soteriological dimensions.

3.5.5 Limits and Divergences

Having noted possible parallels, we now emphasize substantive divergences—areas where Buddhist frameworks and HAFF diverge structurally, methodologically, or conceptually. This section is weighted equally to §3.5.2–3.5.4 to prevent over-interpreting formal similarities.

Phenomenological vs. Structural Orientation

Buddhist philosophy is fundamentally concerned with first-person experience and liberation from suffering (*duḥkha*). Yogācāra and Mādhyamaka analyze consciousness, perception, and mental afflictions (*kleśa*) as prerequisites for soteriological transformation.

HAFF, by contrast, is a third-person structural framework with no phenomenological commitments. It analyzes information-theoretic organization without addressing subjective experience, qualia, or the explanatory gap between structure and consciousness.

Consequence: Any parallel between ālaya-vijñāna and $|\Psi_U\rangle$ cannot extend to phenomenological dimensions. HAFF does not explain consciousness, nor does it claim that quantum states possess mental properties.

Soteriological vs. Descriptive Goals

Buddhist frameworks are *pragmatic* in orientation: concepts are introduced to facilitate liberation, not to accurately describe metaphysical reality. As the Buddha reportedly stated, philosophical speculation is a "thicket of views" (*diṭṭhi-gahaṇa*) distracting from the path [98].

HAFF is *descriptive*: it aims to clarify structural constraints on emergence without prescribing practices or soteriological goals. There is no "path" in HAFF, no liberation to achieve, no suffering to overcome.

Consequence: Structural parallels do not imply that HAFF serves Buddhist soteriological purposes, nor that Buddhist practice requires acceptance of quantum mechanics.

Rebirth, Cosmology, and Ethics

Traditional Buddhist cosmology includes rebirth across multiple realms, karmic causation spanning lifetimes, and detailed ethical taxonomies. These elements are absent from—and irrelevant to—HAFF's structural analysis.

HAFF makes no claims about:

- Post-mortem continuity of consciousness
- Karmic retribution across lifetimes
- Ethical status of actions
- Cosmological realms or deities
- Meditation practices or contemplative attainments

Consequence: Identifying formal parallels does not validate Buddhist cosmology, rebirth doctrine, or ethical systems. The frameworks operate in disjoint conceptual spaces.

Ontological Commitments

While both frameworks reject intrinsic substance, they differ in ontological commitments:

- **Buddhist frameworks** (particularly Yogācāra) often privilege mind or consciousness as fundamental, with matter derivative.
- **HAFF** remains neutral on mind-matter relations, analyzing quantum structure without metaphysical commitments about consciousness.

Additionally, Mādhyamaka's "two truths" doctrine—distinguishing conventional (*saṁvṛti*) from ultimate (*paramārtha*) reality—has no clear analog in HAFF. HAFF distinguishes context-dependent from invariant structures, but this is epistemological (about what can be known) rather than ontological (about levels of reality).

Methodological Incommensurability

Buddhist philosophy employs contemplative introspection, textual hermeneutics, and dialectical reasoning as primary methods. HAFF employs mathematical formalism, operator algebras, and information theory.

These methodologies are not mutually translatable. One cannot *meditate* one's way to understanding quantum entanglement, nor *calculate* one's way to soteriological insight. The parallels identified are *structural*, not methodological.

Summary of divergences. The frameworks differ in:

1. Explanatory target (phenomenology vs. structure)
2. Pragmatic goal (liberation vs. description)
3. Scope (ethics/cosmology vs. physics)
4. Ontological commitments (consciousness-first vs. neutral)

5. Methodology (contemplative vs. mathematical)

These divergences are not defects but reflect different intellectual projects. Recognizing formal parallels does not collapse these distinctions.

3.5.6 Interpretive Humility

We conclude Part IV by reaffirming interpretive humility. The parallels identified—between accessible algebras and ālaya-vijñāna, between emptiness and śūnyatā, between constraint propagation and karma—are *suggestive but inconclusive.*

They suggest that:

1. Certain structural features (holistic substrates, relational existence, informational constraints) appear across traditions when thinkers grapple with similar conceptual problems.

2. Cross-cultural philosophical dialogue may benefit from precise structural comparison, avoiding both premature dismissal and uncritical conflation.

3. Formal parallels can motivate further investigation without requiring doctrinal convergence.

They do *not* suggest that:

1. Buddhist philosophy anticipated quantum mechanics or modern physics.

2. Quantum mechanics validates Buddhist metaphysics or soteriology.

3. Structural similarities entail ontological identity.

4. Comparative philosophy resolves foundational debates in either tradition.

The exercise is exploratory: we map conceptual terrain, noting points of contact and divergence, without claiming to adjudicate between frameworks. Our contribution is methodological—demonstrating how attention to algebraic structure can discipline comparative claims and prevent both over-interpretation and premature dismissal.

3.6 Conclusion: Structural Constraints and Interpretive Modesty

This essay has developed a structural framework for understanding causation, agency, and existence in quantum contexts without canonical subsystem decompositions. The analysis proceeded in four parts.

Part I: Causation. Causal relations can be understood as stable asymmetries in accessible structure, without requiring fundamental temporal ordering or intrinsic relata. This account is context-dependent but objective, grounded in thermodynamic gradients and informational constraint propagation.

Part II: Agency. Agent-like behavior emerges from boundary-stabilization—subsystems maintaining non-scrambling coherence while propagating constraints. This account is eliminativist about intrinsic intentionality but realist about structural patterns. The phenomenology-structure gap remains unresolved.

Part III: Existence. Existential claims can be reformulated in terms of relational patterns rather than intrinsic substance. "Emptiness," in the technical sense developed here, denotes absence of intrinsic being compatible with objectivity. This perspective avoids both naive realism and anti-realist eliminativism.

Part IV: Interpretive Bridges. Formal parallels may exist between these structural features and concepts in Buddhist philosophy (particularly Yogācāra and Mādhyamaka traditions). However, substantive divergences in methodology, goals, and scope prevent conflation. Parallels are offered as invitations to dialogue, not demonstrations of equivalence.

3.6.1 Methodological Contribution

The essay's primary contribution is methodological: it demonstrates how structural analysis can discipline interpretive claims across multiple domains.

1. **In quantum foundations**: By clarifying how causation, agency, and existence depend on coarse-graining structure, the framework reveals which features are context-dependent and which admit invariant characterization.
2. **In philosophy of science**: By developing relational existence without metaphysical substance, the framework contributes to structural realist programs while avoiding reification of emergent entities.
3. **In comparative philosophy**: By identifying formal parallels while respecting substantive divergences, the framework models how cross-cultural comparison can proceed without cultural essentialism or premature synthesis.

3.6.2 What This Essay Does Not Claim

To prevent misreading, we reiterate what the essay does *not* claim:

- That consciousness creates reality or plays a fundamental physical role
- That Buddhist texts anticipated quantum mechanics or modern physics
- That quantum mechanics validates any particular metaphysical or religious tradition
- That structural parallels resolve foundational problems in physics or philosophy
- That comparative philosophy provides unique insights unavailable within traditions

The analysis is structural and comparative, not metaphysical or apologetic.

3.6.3 Open Questions

Several questions remain open:

1. **Phenomenology**: How, if at all, does structural organization relate to subjective experience? The framework developed here is silent on the explanatory gap.

2. **Normativity**: Can constraint propagation ground normative distinctions, or does ethics require additional conceptual resources beyond structural analysis?

3. **Comparative methodology**: What criteria should govern cross-cultural philosophical comparison? When do formal parallels indicate genuine convergence versus superficial similarity?

4. **Empirical implications**: Do different coarse-graining choices lead to observationally distinguishable predictions in realistic physical systems?

5. **Contemplative epistemology**: Can first-person contemplative methods contribute to structural understanding, or are mathematical and phenomenological investigations fundamentally disjoint?

These questions are not deficiencies but opportunities for future investigation. The framework provides conceptual scaffolding for pursuing them with greater precision.

3.6.4 Final Reflection

The absence of canonical subsystem decompositions in quantum mechanics is not merely a technical curiosity. It reshapes how we think about emergence, identity, and existence across multiple domains—from quantum gravity to philosophy of mind to cross-cultural hermeneutics.

By attending to structural constraints—particularly the dependence of effective descriptions on accessible algebras—we can navigate between naive realism and anti-realist eliminativism, between cultural essentialism and dismissive parochialism, between metaphysical dogmatism and interpretive nihilism.

The resulting picture is one of *structured pluralism*: multiple effective descriptions, none metaphysically privileged, yet constrained by objective relational patterns and transformation principles. Reality is not uniquely carved at the joints, but neither is it infinitely malleable. The joints themselves are context-dependent yet objective.

This perspective invites humility. We cannot claim unique access to fundamental structure, nor can we dismiss alternative frameworks as merely conventional. Instead, we map the space of possibilities, identify structural invariants, and acknowledge the limits of any single descriptive framework.

In this spirit, the essay concludes not with answers but with refined questions—questions shaped by attention to algebraic structure, informed by cross-cultural comparison, and disciplined by interpretive modesty.

Chapter 4

Gravitational Phenomena as Emergent Properties

Paper D

Originally published: Zenodo, DOI: 10.5281/zenodo.18388881

Abstract

We propose that gravitational phenomena arise from the adiabatic evolution of accessible observable algebras as the global quantum state evolves. Building on recent work demonstrating that inequivalent coarse-graining structures induce inequivalent effective geometries, we argue that gravity is categorically distinguished from gauge interactions: gauge forces operate *within* a fixed algebra $\mathcal{A}$, while gravitational dynamics reflects the *flow* of $\mathcal{A}$ itself. This framework provides: (1) a generative mechanism for gravitational dynamics via state-dependent algebra selection; (2) a structural derivation of the equivalence principle from algebraic universality; (3) identification of the emergent metric with the Quantum Fisher Information Metric. We do not derive the Einstein equations, but propose a conceptual framework that explains gravity's distinctive features—universality, dynamical geometry, and resistance to naive quantization—as consequences of algebra evolution rather than force mediation. We place this framework on a rigorous algebraic foundation by translating the accessibility criteria into Tomita–Takesaki modular theory, proving uniqueness of the accessible algebra (up to unitary equivalence) under mild ergodicity assumptions, and demonstrating that algebra perturbations yield the linearized Einstein equations via the entanglement first law in holographic settings.

4.1 Introduction

4.1.1 The Quantization Problem

Among the four fundamental interactions, gravity occupies a singular position. While the strong, weak, and electromagnetic forces have been successfully incorporated into the framework of quantum field theory, gravity has resisted analogous treatment for nearly a century. The difficulties are well known: naive quantization of general relativity yields a non-renormalizable theory, and more sophisticated approaches—string theory, loop quantum gravity, asymptotic safety—remain either incomplete or empirically unconfirmed [57].

A common diagnosis attributes this difficulty to the self-referential nature of gravity: the metric tensor both defines the arena in which physics takes place and participates as a dynamical variable within that arena. Quantizing gravity thus appears to require quantizing spacetime itself—a conceptually and technically formidable task.

4.1.2 An Alternative Diagnosis

In this paper, we explore an alternative structural diagnosis. We suggest that the difficulty may arise not because gravity is a particularly subtle force, but because gravity may not be a force at all—at least not in the same categorical sense as gauge interactions.

The proposal rests on a simple observation: all descriptions of physical systems presuppose some decomposition of the total system into subsystems. In quantum mechanics, this corresponds to a tensor factorization of the Hilbert space. However, as has been established in foundational work on quantum information theory [119, 120] and developed in our previous analysis [134], there is no canonical or physically privileged factorization for a generic quantum state. Different choices of factorization—or more generally, different choices of accessible observable algebra—yield inequivalent physical descriptions.

We propose that gauge forces and gravity may be distinguished at this structural level:

- **Gauge forces** describe interactions between degrees of freedom *within* a given subsystem decomposition.
- **Gravitational phenomena** reflect properties of the decomposition *itself*—specifically, how effective geometry emerges from the pattern of accessible observables.

If this reframing is correct, it may help clarify why gravity resists quantization: one cannot straightforwardly quantize the choice of how to divide a system into parts, because that choice is logically prior to the application of quantum dynamics to those parts.

4.1.3 Scope and Limitations

We emphasize at the outset what this paper does and does not attempt.

This paper does:

- Offer a structural reframing of the distinction between gravity and gauge forces
- Draw on established results concerning coarse-graining and emergent geometry
- Identify this perspective as a possible diagnostic for the quantization problem

This paper does not:

- Propose new dynamical equations
- Derive the Einstein field equations or their quantum corrections
- Claim to solve the problem of quantum gravity

- Introduce observer-dependent, consciousness-related, or interpretational elements

The analysis is structural in nature. We examine the conceptual architecture underlying descriptions of gravity and gauge forces, and suggest that a categorical distinction at the level of observable algebras may illuminate longstanding difficulties.

4.1.4 Outline

Section 4.2 reviews the factorization problem: the absence of a canonical subsystem decomposition in quantum theory, and its implications for emergent structure. Section 4.3 develops the proposed distinction between gauge forces and gravity in terms of their relation to observable algebra selection. Section 4.4 provides a technical formulation, including the central conjecture and the identification of the emergent metric with the Quantum Fisher Information Metric. Section 4.5 develops the rigorous algebraic foundation using Tomita–Takesaki modular theory, states the uniqueness conjecture for the accessible algebra with a proof sketch identifying the remaining gaps, verifies the construction in two concrete examples, and connects algebra perturbations to the linearized Einstein equations via the entanglement first law. Section 4.6 discusses connections to existing approaches including AdS/CFT, tensor networks, and thermodynamic gravity. Section 4.7 states explicit scope limitations. Section 4.8 outlines open questions, and Section 4.9 concludes.

4.2 The Factorization Problem

4.2.1 No Canonical Tensor Factorization

In standard quantum mechanics, composite systems are described by tensor products of subsystem Hilbert spaces: $\mathcal{H}_{\text{total}} = \mathcal{H}_A \otimes \mathcal{H}_B$. This structure is typically taken as given, with subsystems identified by physical intuition or experimental arrangement.

However, for a generic Hilbert space $\mathcal{H}$, there is no unique or canonical way to express it as a tensor product. Any finite-dimensional Hilbert space of dimension $d = d_1 \times d_2$ admits a factorization $\mathcal{H} \cong \mathcal{H}_{d_1} \otimes \mathcal{H}_{d_2}$, but the choice of such a factorization is not determined by the Hilbert space structure alone.

This observation was formalized by Zanardi and collaborators [119, 120], who demonstrated that tensor product structures are determined by the algebra of accessible observables rather than by intrinsic properties of the state space. A change in which observables are accessible corresponds to a change in how the system is effectively decomposed into subsystems.

4.2.2 Coarse-Graining and Effective Descriptions

Building on this foundation, our previous work [134] established that:

Proposition 4.1 (Coarse-Graining Induced Inequivalence)**.** *Let $|\Psi_U\rangle \in \mathcal{H}_{total}$ be a global quantum state, and let $\mathbf{c}_1, \mathbf{c}_2$ be two inequivalent coarse-graining structures (defined by distinct accessible algebras $\mathcal{A}_1, \mathcal{A}_2$). Then the effective descriptions induced by $\mathbf{c}_1$ and $\mathbf{c}_2$ are generically inequivalent: they yield different reduced states, different entanglement structures, and—crucially—different effective geometries.*

The key point is that this inequivalence is not merely a matter of coordinate choice or descriptive convention. Different accessible algebras define different physical contents: different sets of measurable quantities, different notions of locality, and different effective spacetime structures.

4.2.3 Geometry from Entanglement

The connection between entanglement and geometry has been extensively studied in the context of holographic duality. The Ryu-Takayanagi formula [83] and its generalizations establish that, in certain settings, geometric quantities (areas of extremal surfaces) are directly related to entanglement entropies of boundary regions:

$$S_A = \frac{\text{Area}(\gamma_A)}{4G_N}. \tag{4.1}$$

More broadly, Van Raamsdonk [103] and others have argued that spacetime connectivity itself may be understood as a manifestation of quantum entanglement: regions that are highly entangled are geometrically "close," while weakly entangled regions are "far apart" or even disconnected.

Within the present framework, these results acquire a natural interpretation. If geometry emerges from entanglement structure, and entanglement structure depends on how the system is decomposed into subsystems, then geometry is ultimately determined by the choice of accessible observable algebra.

Remark 4.2 (Geometry as Coarse-Graining Dependent)**.** *Effective spacetime geometry is not an intrinsic property of the global quantum state $|\Psi_U\rangle$. It is a derived quantity, dependent on the coarse-graining structure* $\mathbf{c}$ *that specifies which observables are accessible. Different coarse-grainings may yield geometries that differ not only in metric components, but in topology and connectivity.*

Remark 4.3 (Relation to Prior Work)**.** *While the non-uniqueness of tensor factorization has been widely discussed in the quantum information literature [119, 120], its implications for distinguishing gravitational phenomena from gauge interactions at a structural level have not, to our knowledge, been made explicit. The present work develops this connection.*

This observation sets the stage for the distinction we develop in the next section.

4.3 A Structural Distinction: Gauge Forces vs. Gravity

4.3.1 Forces Within a Factorization

Consider the standard description of gauge interactions. In quantum electrodynamics, the electromagnetic force is mediated by photon exchange between charged particles. In quantum chromodynamics, gluons mediate the strong force between quarks. In each case, the interaction is described as a coupling between degrees of freedom that are already identified as distinct subsystems.

Formally, gauge theories are constructed on a fixed background: a spacetime manifold M equipped with a principal bundle whose structure group is the gauge group ($U(1)$, $SU(2)$, $SU(3)$, etc.). Matter fields are sections of associated bundles, and gauge fields are connections on the principal bundle. The dynamics describes how these fields interact *given* the background structure.

Crucially, the identification of "electron here" and "photon there" presupposes a decomposition of the total system into localized subsystems. The gauge interaction operates *within* this decomposition, coupling degrees of freedom that have already been distinguished.

4.3.2 Gravity: A Different Category?

General relativity describes gravity not as a force between objects on a fixed background, but as the curvature of spacetime itself. The metric tensor $g_{\mu\nu}$ is both the arena in which physics unfolds and a dynamical variable subject to the Einstein equations.

This dual role has long been recognized as the source of conceptual and technical difficulties. But the present framework suggests a sharper formulation of the distinction.

If gauge forces operate within a given subsystem decomposition, we propose that gravitational phenomena may be understood as reflecting properties of the decomposition itself. Specifically:

- The effective geometry—the metric, the notion of distance, the causal structure—emerges from the pattern of entanglement among accessible degrees of freedom.
- This pattern is determined by the choice of accessible observable algebra.
- Gravitational phenomena, in this view, are not interactions between pre-existing objects, but manifestations of how effective spacetime structure responds to changes in what is accessible.

We emphasize that this proposal does not deny that gravity is geometrical at the effective level. Rather, it suggests that the *origin* of this geometry may lie in how accessible observables define effective subsystems. The geometry remains real and physically consequential; what changes is the account of where it comes from.

4.3.3 An Intuitive Picture

To fix intuitions, consider the following analogy.

Imagine a map of a territory. On the map, one can trace routes between cities—these routes depend on the geography depicted. Now consider the *projection* used to create the map: Mercator, Robinson, or some other. Different projections yield different maps with different distance relationships and shape distortions.

In this analogy:

- **Gauge forces** are like routes on the map—interactions that take place within a given representational structure.
- **Gravity** is like the projection itself—a property of how the representation is constructed, not a feature operating within it.

Changing the projection does not add new routes; it changes what "distance" and "proximity" mean. Similarly, changing the accessible observable algebra does not introduce new forces; it changes the effective geometry in which all forces are described.

Remark 4.4 (Intuitive Picture)**.** *This analogy is offered for conceptual orientation, not as a precise technical claim. The formal relationship between observable algebra selection and effective geometry requires the machinery developed in [134] and subsequent sections of this paper.*

4.3.4 Implications for Quantization

If this structural distinction is correct, it may illuminate the difficulty of quantizing gravity.

Quantizing a gauge theory means promoting classical fields to operator-valued distributions on a fixed background, subject to appropriate commutation relations and dynamics. The background—including the decomposition into subsystems—is held fixed while the fields are quantized.

But if gravity reflects the choice of decomposition itself, then "quantizing gravity" would require quantizing the selection of how to divide the system into parts. This is a categorically different task. It is not a matter of promoting a classical field to a quantum operator; it is a matter of making the *framework in which quantization is defined* itself subject to quantum uncertainty.

This may explain why straightforward approaches to quantum gravity encounter difficulties: they attempt to apply quantization procedures designed for systems *within* a fixed decomposition to a structure that determines the decomposition itself.

This perspective does not introduce new dynamics or predictions, but may offer diagnostic value: it suggests a structural reason why gravity resists the quantization procedures that succeed for gauge interactions, and points toward the need for approaches that do not presuppose a fixed subsystem decomposition.

Remark 4.5 (Diagnostic Value)**.** *We do not claim that this perspective solves the problem of quantum gravity. Rather, we suggest that it offers diagnostic value: it identifies a structural reason why gravity may resist the techniques that succeed for gauge forces, and points toward the need for approaches that do not presuppose a fixed subsystem decomposition.*

4.3.5 Relation to Background Independence

The idea that gravity is connected to "background independence" is well established in the quantum gravity literature [82, 96]. The present proposal may be viewed as a sharpening of this intuition in terms of observable algebras.

Background independence is often formulated as the requirement that physical laws not depend on a fixed spacetime metric. In the present framework, this requirement is subsumed under a more general principle: physical content should not depend on a particular choice of accessible observable algebra, or at least should transform covariantly under changes in that choice.

This suggests that a satisfactory theory of quantum gravity may need to be formulated not in terms of fields on a spacetime manifold, but in terms of structures that are prior to—or more fundamental than—the decomposition into spatially localized subsystems.

Remark 4.6 (Context-Dependence vs. Observer-Dependence)**.** *A potential misreading of this proposal is that it renders gravity "observer-dependent" or subjective. We stress that this is not the case. The selection of accessible observable algebras is constrained by* *physical interactions and stability criteria (such as decoherence structure and dynamical invariance), not by subjective choice or epistemic limitation. The resulting effective geometry is* context-dependent*—it depends on which physical degrees of freedom are stably accessible—but not* observer-relative *in any subjective sense. This distinction is developed in detail in [135].*

4.3.6 The Equivalence Principle from Algebraic Universality

A central puzzle in gravitational physics is the universality of free fall: why do all forms of matter and energy couple to gravity in the same way? In standard approaches, this "equivalence principle" is imposed as an empirical postulate. Here, we suggest it may follow structurally from the algebraic perspective.

The key observation is that the effective geometry is not a property of any particular matter field, but a property of the *accessible algebra* $\mathcal{A}_\mathbf{c}$ itself. All observable matter fields are, by definition, constructed from operators in $\mathcal{A}_\mathbf{c}$ or its representations. Consequently, they must necessarily inhabit the geometry induced by $\mathcal{A}_\mathbf{c}$.

There is no "second geometry" for a different particle species to follow, because any operator outside $\mathcal{A}_\mathbf{c}$ is operationally inaccessible within the given coarse-graining context. The universality of gravitational coupling is thus not an additional postulate, but a logical consequence of the universality of the observable algebra.

We note that this argument is *kinematic* (all matter inhabits the same geometry) rather than *dynamical* (all matter follows geodesics of that geometry). The full dynamical content of the Weak Equivalence Principle—that test masses follow geodesics of the emergent metric—requires additional input beyond the algebraic framework presented here.

Remark 4.7 (Dark Sector as Algebraic Inaccessibility)**.** *This perspective suggests a natural interpretation of "dark" degrees of freedom. Matter that does not couple to our accessible algebra* $\mathcal{A}_\mathbf{c}$*—while potentially present in the global state* $|\Psi_U\rangle$*—would be operationally invisible except through its gravitational effects on the geometry induced by* $\mathcal{A}_\mathbf{c}$*—a possibility that requires extending the framework to allow degrees of freedom outside* $\mathcal{A}_\mathbf{c}$ *to influence the algebra selection process, which is not currently formalized. This is speculative but structurally consistent with the framework.*

4.4 Technical Formulation

4.4.1 Setup and Notation

We consider a global quantum system described by a Hilbert space $\mathcal{H}$ with algebra of bounded operators $\mathcal{B}(\mathcal{H})$.

A *coarse-graining* is specified by the selection of an accessible subalgebra $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$, representing the observables that remain stable under relevant dynamical and environmental constraints.

Definition 4.8 (Accessible Algebra)**.** *Following [134, 135], an* ***accessible algebra*** $\mathcal{A}_\mathbf{c} \subset \mathcal{B}(\mathcal{H}_U)$ *is a* $*$*-subalgebra satisfying three stability criteria:*

1. ***Dynamical invariance:*** *Expectation values of operators in* $\mathcal{A}_\mathbf{c}$ *remain approximately invariant under physically motivated dynamical maps* $\mathcal{E}$*:*

$$\|\mathcal{E}(\hat{O}) - \hat{O}\| \ll \epsilon \quad \forall \hat{O} \in \mathcal{A}_\mathbf{c}. \tag{4.2}$$

2. ***Environmental redundancy (Quantum Darwinism):*** *The subalgebra approximately commutes with the environmental algebra $\mathcal{A}_E$:*

$$[\hat{O}, \hat{E}] \approx 0 \quad \forall \hat{O} \in \mathcal{A}_{\mathbf{c}},\ \hat{E} \in \mathcal{A}_E. \tag{4.3}$$

3. ***Non-scrambling:*** *Out-of-time-order correlators exhibit slow decay:*

$$\langle [\hat{O}_{\mathcal{A}}(t), \hat{V}(0)]^2 \rangle \ll 1 \quad \textit{for } t \ll \tau_{scrambling}. \tag{4.4}$$

*A **coarse-graining structure** is the pair $\mathbf{c} \equiv (\mathcal{A}_{\mathbf{c}}, \Phi_{\mathbf{c}})$, where $\Phi_{\mathbf{c}}$ is a CPTP map implementing the operational reduction.*

No assumption is made that such a subalgebra admits a unique or canonical tensor factorization of $\mathcal{H}$.

Remark 4.9. *This notion of accessible algebra follows the spirit of algebraic quantum mechanics and quantum information–theoretic approaches, without assuming a preferred subsystem decomposition.*

4.4.2 Entanglement Structure and Induced Geometry

Given a choice of accessible algebra $\mathcal{A}_{\mathbf{c}}$, one may consider the entanglement structure induced by restricting the global state ρ to $\mathcal{A}_{\mathbf{c}}$.

Following insights from holography and tensor network constructions, patterns of entanglement within $\mathcal{A}_{\mathbf{c}}$ may be associated with an effective distance structure on equivalence classes of observables.

Crucially, this effective geometry depends on:

- the choice of $\mathcal{A}_{\mathbf{c}}$,
- the stability of correlations under coarse-grained dynamics,
- and the redundancy of information encoding.

No claim is made that this geometry is fundamental. It is an effective description, valid within the context defined by $\mathcal{A}_{\mathbf{c}}$.

4.4.3 Central Conjecture

We now state the central conjecture of this paper explicitly.

Conjecture 4.10 (Gravity as Adiabatic Algebra Evolution)**.** *Gravitational dynamics corresponds to the **adiabatic flow** of the accessible algebra $\mathcal{A}_{\mathbf{c}}(t)$, tracked by the stability conditions (Definition 4.8) acting on the evolving global state $|\Psi_U(t)\rangle$.*

Specifically:

1. *The global state evolves unitarily: $|\Psi_U(t)\rangle = U(t)|\Psi_U(0)\rangle$.*
2. *The stability criteria determine which subalgebra $\mathcal{A}_{\mathbf{c}}(t) \subset \mathcal{B}(\mathcal{H}_U)$ is accessible at each time.*
3. *As the state evolves, the optimal stable algebra shifts: $\mathcal{A}_{\mathbf{c}}(t) \to \mathcal{A}_{\mathbf{c}}(t + dt)$.*

4. This shift $\dot{\mathcal{A}}_{\mathbf{c}}(t)$ manifests phenomenologically as dynamical spacetime curvature—i.e., as gravity.

In contrast, unitary evolution of observables within *a fixed algebra $\mathcal{A}_{\mathbf{c}}$ manifests as gauge interactions. The categorical distinction is:*

- ***Gauge dynamics:*** *Evolution within $\mathcal{A}_{\mathbf{c}}$ (fixed stage, moving actors)*
- ***Gravitational dynamics:*** *Evolution of $\mathcal{A}_{\mathbf{c}}$ itself (moving stage)*

Remark 4.11 (Scope of the distinction)**.** *This categorical distinction is expected to be valid only as a* low-energy, effective *characterization. In UV-complete theories where gauge and gravitational degrees of freedom are unified (as in string theory/M-theory dualities), the distinction between "evolution within $\mathcal{A}_{\mathbf{c}}$" and "evolution of $\mathcal{A}_{\mathbf{c}}$" may break down. The adiabatic assumption (slow algebra transitions relative to internal dynamics) is essential for the distinction to be operationally meaningful.*

This formulation addresses a key objection: if algebras are kinematical background, how can gravity be dynamical? The answer is that the *selection* of which algebra is stable is itself state-dependent, and state evolution induces algebra flow.

Remark 4.12 (Status of Algebraic Variations)**.** *The adiabatic approximation assumes that algebra transitions occur slowly relative to internal dynamics within $\mathcal{A}_{\mathbf{c}}$. Rapid transitions would correspond to strong gravitational effects or spacetime singularities—regimes where the effective geometric description breaks down. The question of what dynamics, if any, governs non-adiabatic transitions is left open (see Section 4.8).*

4.4.4 Metric from Quantum Information Geometry

To make the algebra-geometry correspondence precise, we propose that the emergent metric is *related to* the **Quantum Fisher Information Metric (QFIM)**, a standard construction in quantum information geometry [76, 11].

Let $\{\lambda^\mu\}$ be parameters labeling deformations of the accessible algebra or its defining stability surface. The induced metric $g_{\mu\nu}$ on the manifold of effective descriptions is given by:

$$g_{\mu\nu}(\lambda) = \frac{1}{2}\mathrm{Tr}\left(\rho(\lambda)\{L_\mu, L_\nu\}\right), \tag{4.5}$$

where L_μ is the symmetric logarithmic derivative satisfying

$$\partial_\mu \rho = \frac{1}{2}(\rho L_\mu + L_\mu \rho). \tag{4.6}$$

This construction has several attractive features:

- It is coordinate-independent and intrinsically quantum.
- It reduces to the classical Fisher metric in appropriate limits.
- It is directly related to distinguishability of quantum states—geometrically "close" states are hard to distinguish operationally.

Remark 4.13 (Caveats on the QFIM identification)**.** *Two issues must be addressed before the QFIM can serve as a literal spacetime metric:*

(a) *The map from algebra deformations $\delta\mathcal{A}_{\mathbf{c}}$ to state deformations $\delta\rho$ (which enter the QFIM) must be specified explicitly; this map depends on the state–algebra relationship and is not uniquely determined by the framework.*

(b) *The QFIM is positive-definite (Riemannian), while physical spacetime has Lorentzian signature. The mechanism by which a Lorentzian metric emerges from a Riemannian information-geometric construction—the "signature problem"—remains open.*

Under the hypothesis that gravitational dynamics reflects algebra evolution (Conjecture 4.10), the Einstein tensor $G_{\mu\nu}$ may be understood as describing the curvature of this information manifold. Changes in the accessible algebra along a one-parameter family $\mathcal{A}_{\mathbf{c}}(\lambda)$ with $\mathcal{A}_{\mathbf{c}}(0) = \mathcal{A}_{\mathbf{c}}$ induce metric perturbations $\delta g_{\mu\nu}$ that correspond, in the effective geometric description, to gravitational waves.

Remark 4.14 (Relation to Holographic Results)**.** *In AdS/CFT, the Ryu-Takayanagi formula provides a precise relationship: $S_A = \mathrm{Area}(\gamma_A)/4G_N$. The QFIM construction is consistent with this correspondence: the Fisher information metric on boundary states induces a bulk geometry whose areas encode entanglement entropies [60, 32]. The present framework proposes that this relationship is not specific to holography but reflects a general structural principle.*

4.4.5 What This Section Does Not Claim

To prevent misreading, we state explicitly what this technical formulation does *not* attempt:

- It does not derive gravitational field equations.
- It does not specify a dynamics for coarse-graining selection.
- It does not claim empirical adequacy or testable predictions.
- It does not introduce observer-dependent or consciousness-related elements.

The role of this section is to demonstrate internal coherence between the structural claims of Sections 1–3 and existing entanglement–geometry correspondences in the literature.

4.5 Algebraic Foundation: Modular Uniqueness

The preceding sections formulated the HAFF gravity conjecture in structural and information-geometric terms. The accessibility criteria (Definition 4.8) were stated in approximate, operational language. Here we translate these criteria into Tomita–Takesaki modular theory, state a uniqueness conjecture with a proof sketch, verify the construction in two examples, and connect to linearized gravity via the entanglement first law.

4.5.1 Preliminaries

We review the mathematical tools required for the main construction. Standard references include Takesaki [100], Bratteli–Robinson [19], and Haag [43].

Von Neumann Algebras

Definition 4.15. *A **von Neumann algebra** $\mathcal{M}$ acting on a Hilbert space $\mathcal{H}$ is a $*$-subalgebra of $\mathcal{B}(\mathcal{H})$ that contains the identity and is closed in the weak operator topology. Equivalently, by von Neumann's bicommutant theorem, $\mathcal{M} = \mathcal{M}''$, where $\mathcal{M}' = \{T \in \mathcal{B}(\mathcal{H}) : [T, M] = 0 \, \forall \, M \in \mathcal{M}\}$ is the commutant.*

Definition 4.16. *A von Neumann algebra $\mathcal{M}$ is a **Type III$_1$ factor** if it has trivial center ($\mathcal{M} \cap \mathcal{M}' = \mathbb{C}\mathbf{1}$), admits no finite or semifinite normal trace, and the Connes spectrum $S(\mathcal{M}) = \mathbb{R}_{\geq 0}$.*

By a deep result of algebraic QFT, local algebras of observables in any reasonable quantum field theory are Type III$_1$ factors [43, 118].

Tomita–Takesaki Modular Theory

The Tomita–Takesaki theorem is the central structure theorem for von Neumann algebras with a cyclic and separating vector.

Definition 4.17. *Let $\mathcal{M}$ be a von Neumann algebra on $\mathcal{H}$, and let $|\Omega\rangle \in \mathcal{H}$.*

- *$|\Omega\rangle$ is **cyclic** for $\mathcal{M}$ if $\mathcal{M}|\Omega\rangle$ is dense in $\mathcal{H}$.*
- *$|\Omega\rangle$ is **separating** for $\mathcal{M}$ if $M|\Omega\rangle = 0$ implies $M = 0$ for all $M \in \mathcal{M}$.*

If $|\Omega\rangle$ is cyclic and separating, define the antilinear operator

$$S_\Omega : M|\Omega\rangle \mapsto M^*|\Omega\rangle, \qquad M \in \mathcal{M}. \tag{4.7}$$

This operator is closable, and its closure admits a polar decomposition:

$$S_\Omega = J_\Omega \Delta_\Omega^{1/2}, \tag{4.8}$$

where J_Ω is an antiunitary involution (the **modular conjugation**) and Δ_Ω is a positive self-adjoint operator (the **modular operator**).

Theorem 4.18 (Tomita–Takesaki [100])**.** *Let $\mathcal{M}$ be a von Neumann algebra with cyclic and separating vector $|\Omega\rangle$, and let Δ_Ω, J_Ω be as above. Then:*

(a) $J_\Omega \mathcal{M} J_\Omega = \mathcal{M}'$ (modular conjugation exchanges the algebra and its commutant).

*(b) The **modular automorphism group***

$$\sigma_t^\Omega(M) := \Delta_\Omega^{it} M \Delta_\Omega^{-it}, \qquad t \in \mathbb{R}, \tag{4.9}$$

satisfies $\sigma_t^\Omega(\mathcal{M}) = \mathcal{M}$ for all $t \in \mathbb{R}$.

Thus, any von Neumann algebra with a faithful state possesses a canonical one-parameter automorphism group.

The KMS Condition

Definition 4.19. *Let $\mathcal{M}$ be a von Neumann algebra, α_t a one-parameter automorphism group, and ω a normal state on $\mathcal{M}$. The state ω satisfies the **KMS condition** at inverse temperature β with respect to α_t if, for all $A, B \in \mathcal{M}$, there exists a function $F_{A,B}(z)$ analytic in the strip $0 < \mathrm{Im}(z) < \beta$ and continuous on its closure, such that*

$$F_{A,B}(t) = \omega(A\,\alpha_t(B)), \qquad F_{A,B}(t+i\beta) = \omega(\alpha_t(B)\,A) \tag{4.10}$$

for all $t \in \mathbb{R}$.

Theorem 4.20 (Takesaki [100])**.** *Let ω be a faithful normal state on a von Neumann algebra $\mathcal{M}$. Then ω is KMS at $\beta = 1$ with respect to σ_t^ω, and σ_t^ω is the* unique *one-parameter automorphism group with this property.*

The physical content is striking: the modular flow is the unique time evolution for which the given state looks thermal. In the Rindler wedge, this flow is the Lorentz boost, and the KMS condition at $\beta = 2\pi$ reproduces the Unruh temperature $T_U = (2\pi)^{-1}$ (in natural units where the acceleration $a = 1$).

Half-Sided Modular Inclusions

Definition 4.21 ([113])**.** *Let $\mathcal{N} \subset \mathcal{M}$ be an inclusion of von Neumann algebras on $\mathcal{H}$, with $|\Omega\rangle$ cyclic and separating for both. The inclusion is a **half-sided modular inclusion** (HSMI) if*

$$\sigma_t^{\mathcal{M}}(\mathcal{N}) \subset \mathcal{N} \qquad \text{for all } t \leq 0, \tag{4.11}$$

where $\sigma_t^{\mathcal{M}}$ denotes the modular automorphism group of $(\mathcal{M}, |\Omega\rangle)$.

Theorem 4.22 (Wiesbrock [113])**.** *Let $\mathcal{N} \subset \mathcal{M}$ be a half-sided modular inclusion with common cyclic and separating vector $|\Omega\rangle$. Then there exists a unique one-parameter unitary group $U(a) = e^{-iaP}$, $a \geq 0$, with positive generator $P \geq 0$, such that*

$$\mathcal{N} = U(1)\mathcal{M}U(1)^*. \tag{4.12}$$

Moreover, $U(a)|\Omega\rangle = |\Omega\rangle$ for all a.

The physical interpretation is that the translation from $\mathcal{M}$ to $\mathcal{N}$ is encoded algebraically: no background geometry is needed to define the notion of "shifting a wedge." This is the key tool for extracting spacetime structure from purely algebraic data.

4.5.2 Modular Definition of Accessible Algebras

We now reformulate the physical accessibility criteria (Definition 4.8) in algebraic terms.

Let $\mathcal{H}_U$ be the universal Hilbert space, $|\Psi_U\rangle \in \mathcal{H}_U$ the global state, $\hat{H}$ the total Hamiltonian, and $\mathcal{E}_t$ the dynamical (decoherence) semigroup in the Schrödinger picture: $\mathcal{E}_t(\rho) = e^{t\mathcal{L}}(\rho)$ for a Lindblad generator $\mathcal{L}$. The Heisenberg-picture dual $\mathcal{E}_t^*$ acts on observables via $\mathrm{tr}(\mathcal{E}_t^*(A)\,\rho) = \mathrm{tr}(A\,\mathcal{E}_t(\rho))$. We seek a von Neumann subalgebra $\mathcal{A}_\mathbf{c} \subset \mathcal{B}(\mathcal{H}_U)$ representing the physically accessible observables.

Definition 4.23 (Modular accessible algebra)**.** *Define the state $\omega(\cdot) = \langle\Psi_U|\cdot|\Psi_U\rangle$. A von Neumann subalgebra $\mathcal{A}_\mathbf{c} \subset \mathcal{B}(\mathcal{H}_U)$ is a **modular accessible algebra** if it satisfies:*

(U1) ***Faithfulness.*** *The restriction of* ω *to* $\mathcal{A}_{\mathbf{c}}$ *is faithful:* $\omega(A^*A) = 0$ *implies* $A = 0$ *for all* $A \in \mathcal{A}_{\mathbf{c}}$*. Equivalently, the GNS vector associated to* $\omega|_{\mathcal{A}_{\mathbf{c}}}$ *is cyclic and separating for* $\mathcal{A}_{\mathbf{c}}$*.*

(U2) ***Modular stability.*** *Let* $\sigma_t^{\omega,\mathrm{tot}}$ *denote the modular automorphism group of the pair* $(\mathcal{B}(\mathcal{H}_U), \omega)$*—the modular flow of the* ambient *algebra. Then* $\mathcal{A}_{\mathbf{c}}$ *is globally invariant:*

$$\sigma_t^{\omega,\mathrm{tot}}(\mathcal{A}_{\mathbf{c}}) = \mathcal{A}_{\mathbf{c}} \qquad \forall\, t \in \mathbb{R}. \tag{4.13}$$

Note: *This is a non-trivial condition. The modular flow of* $(\mathcal{A}_{\mathbf{c}}, \omega|_{\mathcal{A}_{\mathbf{c}}})$ *preserves* $\mathcal{A}_{\mathbf{c}}$ *automatically by the Tomita–Takesaki theorem; requiring invariance under the* ambient *modular flow* $\sigma_t^{\omega,\mathrm{tot}}$ *is a genuine constraint that selects subalgebras compatible with the global state's modular structure.*

(U3) ***Maximality.*** $\mathcal{A}_{\mathbf{c}}$ *is the* maximal *von Neumann subalgebra of* $\mathcal{B}(\mathcal{H}_U)$ *satisfying (U1)–(U2) together with invariance under the Heisenberg-picture decoherence dynamics:*

$$\mathcal{E}_t^*(\mathcal{A}_{\mathbf{c}}) \subset \mathcal{A}_{\mathbf{c}} \qquad \forall\, t \geq 0. \tag{4.14}$$

That is, if $\mathcal{A}' \supset \mathcal{A}_{\mathbf{c}}$ *also satisfies (U1), (U2), and decoherence invariance, then* $\mathcal{A}' = \mathcal{A}_{\mathbf{c}}$*.*

Remark 4.24 (Modular theory in finite dimensions)**.** *For finite-dimensional* $\mathcal{H}_U$*,* $\mathcal{B}(\mathcal{H}_U)$ *is a Type I factor and the modular flow reduces to Hamiltonian evolution:* $\sigma_t^{\omega,\mathrm{tot}}(X) = \rho^{it} X \rho^{-it}$ *where* ρ *is the density matrix of* ω*. Condition (U2) then becomes:* $\mathcal{A}_{\mathbf{c}}$ *is invariant under conjugation by* ρ^{it}*. The full power of Tomita–Takesaki theory becomes essential only in the infinite-dimensional/QFT setting (Type III factors), as illustrated by the Rindler wedge example. The qubit chain example is a pedagogical illustration in the Type I setting.*

Remark 4.25 (Relation to physical criteria)**.** *The three algebraic conditions translate the three physical criteria as follows:*

Physical	***Algebraic***	***Mechanism***
(P1) Dyn. invariance	*(U3)* $\mathcal{E}_t^*(\mathcal{A}_{\mathbf{c}}) \subset \mathcal{A}_{\mathbf{c}}$	*Decoherence invariance*
(P2) Redundancy	*(U1) Faithfulness*	*Faithful state* $\Leftrightarrow$ *no lost info*
(P3) Non-scrambling	*(U2) Modular stability*	$\sigma_t^{\omega,\mathrm{tot}}$ *preserves* $\mathcal{A}_{\mathbf{c}}$

The mapping from (P2) to (U1): environmental redundancy ensures that the state restricted to $\mathcal{A}_{\mathbf{c}}$ *does not lose any information about the relevant degrees of freedom—i.e., the restriction is faithful. A non-faithful restriction would mean that some operators in the algebra have zero expectation in all states reachable by environmental monitoring, contradicting redundancy.*

The mapping from (P3) to (U2): non-scrambling means that information encoded in accessible observables does not rapidly leak into the rest of $\mathcal{B}(\mathcal{H}_U)$*. The modular automorphism group* σ_t^{ω} *is the canonical "internal time evolution" of the algebra with respect to the state* ω *(Theorem 4.20). Modular stability of* $\mathcal{A}_{\mathbf{c}}$ *under* σ_t^{ω} *means that this internal evolution does not generate operators outside* $\mathcal{A}_{\mathbf{c}}$*—a precise algebraic version of non-scrambling.*

Remark 4.26 (The role of $\mathcal{E}_t$)**.** *Condition (U3) incorporates the decoherence dynamics* $\mathcal{E}_t$ *into the maximality condition. This is the only place where the physical environment* *enters the algebraic definition. In AQFT on Minkowski space, $\mathcal{E}_t$ can be identified with the restriction map to a causal domain. In open quantum systems, $\mathcal{E}_t$ is the Lindblad semigroup. The framework requires only that $\mathcal{E}_t$ is a normal completely positive map on $\mathcal{B}(\mathcal{H}_U)$.*

4.5.3 Uniqueness Conjecture

We conjecture that the modular accessible algebra, if it exists, is unique up to unitary equivalence under a suitable ergodicity condition. This refines Conjecture 2.1 (Paper B) by making the ergodicity assumption and proof gaps explicit. A complete proof remains open; we present the conjecture and a proof sketch that identifies the key steps and the gaps that remain to be closed.

Conjecture 4.27 (Uniqueness of the modular accessible algebra)**.** *Let $(\mathcal{H}_U, |\Psi_U\rangle, \hat{H}, \mathcal{E}_t)$ be a quantum system as above, and assume the following ergodicity condition:*

(E) *The joint action of the ambient modular flow $\sigma_t^{\omega,\mathrm{tot}}$ and the Heisenberg-picture decoherence semigroup $\mathcal{E}_t^*$ is ergodic on $\mathcal{B}(\mathcal{H}_U)$: the only operator invariant under both is a scalar multiple of the identity.*

Then the modular accessible algebra $\mathcal{A}_{\mathbf{c}}$ satisfying (U1)–(U3) *is unique up to unitary equivalence.*

Proof sketch. The argument proceeds in three steps. We flag at each step the assumptions that go beyond (E).

Step 1: Uniqueness of modular flow (rigorous). By Theorem 4.20, for any faithful normal state ω on a von Neumann algebra $\mathcal{M}$, the modular automorphism group σ_t^ω is the *unique* one-parameter automorphism group satisfying the KMS condition at $\beta = 1$. Therefore, given the global state $|\Psi_U\rangle$ and a candidate algebra $\mathcal{A}_{\mathbf{c}}$ satisfying (U1), the modular flow on $\mathcal{A}_{\mathbf{c}}$ is uniquely determined. There is no freedom in the choice of modular automorphism: it is a function of $(\mathcal{A}_{\mathbf{c}}, \omega)$ alone.

Step 2: Lattice of invariant subalgebras (partially rigorous). Condition (U2) requires $\mathcal{A}_{\mathbf{c}}$ to be globally *invariant* (not element-wise fixed) under the ambient modular flow: $\sigma_t^{\omega,\mathrm{tot}}(\mathcal{A}_{\mathbf{c}}) = \mathcal{A}_{\mathbf{c}}$ for all t. Condition (U3) requires $\mathcal{E}_t^*(\mathcal{A}_{\mathbf{c}}) \subset \mathcal{A}_{\mathbf{c}}$ for all $t \geq 0$. Together, $\mathcal{A}_{\mathbf{c}}$ lies in the lattice $\mathfrak{L}$ of von Neumann subalgebras of $\mathcal{B}(\mathcal{H}_U)$ that are simultaneously invariant under $\sigma_t^{\omega,\mathrm{tot}}$ and $\mathcal{E}_t^*$, and on which ω is faithful.

Invariance is preserved under joins. If $\mathcal{A}_1, \mathcal{A}_2 \in \mathfrak{L}$, then $\sigma_t^{\omega,\mathrm{tot}}(\mathcal{A}_1 \vee \mathcal{A}_2) = \mathcal{A}_1 \vee \mathcal{A}_2$ (automorphisms respect the lattice of von Neumann subalgebras). Similarly, $\mathcal{E}_t^*$ preserves $\mathcal{A}_1 \vee \mathcal{A}_2$. However, ω need not be faithful on the join: this is a genuine obstruction.

Step 3: Maximality argument (gap identified). The intended conclusion is: if $\mathcal{A}_{\mathbf{c}}$ and $\mathcal{A}'$ are both maximal elements of $\mathfrak{L}$, then $\mathcal{A}_{\mathbf{c}} = \mathcal{A}'$ up to unitary equivalence. The natural strategy is to show that $\mathcal{A}_{\mathbf{c}} \vee \mathcal{A}'$ either belongs to $\mathfrak{L}$ (contradicting maximality unless $\mathcal{A}_{\mathbf{c}} = \mathcal{A}'$) or violates one of the conditions.

What is established: Invariance of the join under both dynamics (Step 2).

What remains open: Two gaps must be closed to complete the proof:

(G1) **Faithfulness of the join**: Ergodicity (E) constrains the *fixed-point* algebra of the joint dynamics, but does not directly control faithfulness of ω on the *invariant* subalgebras. A stronger assumption—e.g., that $\mathcal{E}_t$ is a *Schwarz map* with no non-trivial decoherence-free subalgebra—would close this gap by ensuring that any proper extension leaves the domain of faithfulness.

(G2) **Dichotomy for the join**: The argument that "$\mathcal{A}_{\mathbf{c}} \vee \mathcal{A}'$ violates (U1) or (U2) implies $\mathcal{A}_{\mathbf{c}} \subset \mathcal{A}'$" requires an additional lattice-theoretic step: that there is no "third possibility" where the join fails (U3) while neither algebra contains the other. This holds if $\mathfrak{L}$ is a *modular lattice* (in the lattice-theoretic sense), but this has not been verified for the relevant class of von Neumann subalgebras. The lattice of von Neumann subalgebras of $\mathcal{B}(\mathcal{H})$ is known to be non-modular in general [42]; closing this gap likely requires restricting to specific structural classes (e.g., subfactors of finite Jones index).

We expect that both gaps can be closed for physically relevant models (finite-dimensional or hyperfinite type III_1 factors), but a general proof is not yet available. The residual unitary freedom (between different faithful normal representations of the same abstract algebra) would be absorbed by the standard form [44]. □

Remark 4.28 (On the ergodicity assumption and the gaps)**.** *Condition (E) excludes systems with degenerate ground states, spontaneous symmetry breaking, or phase coexistence, which may support multiple non-unitarily-equivalent accessible algebras. This is the algebraic analogue of the non-uniqueness of the broken-symmetry vacuum in quantum field theory. When (E) fails, the space of accessible algebras acquires a non-trivial moduli space, analogous to the landscape of superselection sectors in AQFT.*

Closing gap (G1) is the more tractable of the two: for finite-dimensional systems, faithfulness is equivalent to the state having full support, and joins of supported subalgebras remain supported. Gap (G2) is more subtle and may require restricting to specific classes of decoherence semigroups. We flag this as a problem for future work.

Remark 4.29 (Physical content of maximality)**.** *Maximality (U3) has a clear physical meaning: the accessible algebra should contain* all *observables that are stable under both modular flow and decoherence. If an operator is stable but excluded from $\mathcal{A}_{\mathbf{c}}$, it should in principle be observable, contradicting the assumption that $\mathcal{A}_{\mathbf{c}}$ captures all accessible information. Maximality thus enforces completeness of the physical description.*

4.5.4 Example 1: Rindler Wedge

Consider a free massless scalar field $\phi(x)$ in $(1+3)$-dimensional Minkowski spacetime $(\mathbb{R}^{1,3}, \eta_{\mu\nu})$. The total Hilbert space is the Fock space $\mathcal{H}_U = \mathcal{F}(\mathcal{H}_1)$ over the one-particle space $\mathcal{H}_1 = L^2(\mathbb{R}^3)$. The global state is the Minkowski vacuum $|\Omega\rangle$.

Define the right Rindler wedge:

$$\mathcal{W}_R = \{(t, x, y, z) \in \mathbb{R}^{1,3} : x > |t|\}. \tag{4.15}$$

The local algebra $\mathcal{R}(\mathcal{W}_R) = \{e^{i\phi(f)} : \operatorname{supp} f \subset \mathcal{W}_R\}''$ is the von Neumann algebra generated by Weyl operators smeared with test functions supported in $\mathcal{W}_R$.

Verification of (U1): Faithfulness

The Reeh–Schlieder theorem [43] guarantees that the Minkowski vacuum $|\Omega\rangle$ is cyclic and separating for $\mathcal{R}(\mathcal{W}_R)$. Therefore, the state $\omega(\cdot) = \langle\Omega| \cdot |\Omega\rangle$ is faithful on $\mathcal{R}(\mathcal{W}_R)$.

Verification of (U2): Modular stability

By the Bisognano–Wichmann theorem [14, 15], the modular operator of $(\mathcal{R}(\mathcal{W}_R), |\Omega\rangle)$ is

$$\Delta_\Omega = e^{-2\pi K}, \tag{4.16}$$

where K is the boost generator in the x-direction. The modular automorphism group acts as

$$\sigma_t^\Omega(\phi(t_0, \mathbf{x})) = \phi(\Lambda_{2\pi t}(t_0, \mathbf{x})), \tag{4.17}$$

where Λ_s is the Lorentz boost with rapidity s. Since the Rindler wedge is invariant under Lorentz boosts—$\Lambda_s(\mathcal{W}_R) = \mathcal{W}_R$ for all s—the modular flow preserves the wedge algebra:

$$\sigma_t^\Omega(\mathcal{R}(\mathcal{W}_R)) = \mathcal{R}(\mathcal{W}_R) \qquad \forall\, t \in \mathbb{R}. \tag{4.18}$$

Verification of (U3): Maximality

By Haag duality for wedge regions [43], $\mathcal{R}(\mathcal{W}_R)' = \mathcal{R}(\mathcal{W}_L)$, where $\mathcal{W}_L$ is the left Rindler wedge. Any extension $\mathcal{A}' \supsetneq \mathcal{R}(\mathcal{W}_R)$ must contain operators in $\mathcal{R}(\mathcal{W}_L)$. Such operators are mapped outside $\mathcal{R}(\mathcal{W}_R)$ by the modular flow (boost), so any proper extension would violate either (U1) or (U2). The wedge algebra is maximal.

Physical content

The modular flow σ_t^Ω coincides with the Lorentz boost, and the KMS condition at $\beta = 2\pi$ gives the Unruh temperature $T_U = a/(2\pi)$, where a is the proper acceleration. The modular Hamiltonian is

$$K_{\text{mod}} = -\ln \Delta_\Omega = 2\pi K = 2\pi \int_{\mathcal{W}_R} d\Sigma^\mu\, x_\nu\, T^\nu{}_\mu, \tag{4.19}$$

where $d\Sigma^\mu$ is the surface element on the $t = 0$ Cauchy surface. This example demonstrates that the modular accessible algebra framework reproduces the standard Rindler physics: the accessible algebra is the wedge algebra, its modular flow is the boost, and the KMS state is the Unruh thermal state.

4.5.5 Example 2: Qubit Chain with Decoherence

Consider n qubits with total Hilbert space $\mathcal{H}_U = (\mathbb{C}^2)^{\otimes n}$. The system Hamiltonian is the transverse-field Ising model:

$$\hat{H} = -J \sum_{i=1}^{n-1} \sigma_z^{(i)} \sigma_z^{(i+1)} - h \sum_{i=1}^{n} \sigma_x^{(i)}, \tag{4.20}$$

with $J > 0$ and transverse field strength h. The system is coupled to a thermal environment at inverse temperature β through a dephasing Lindblad master equation:

$$\mathcal{E}_t(\rho) = e^{t\mathcal{L}}(\rho), \qquad \mathcal{L}(\rho) = -i[\hat{H}, \rho] + \gamma \sum_i \left(\sigma_z^{(i)} \rho\, \sigma_z^{(i)} - \rho\right), \tag{4.21}$$

where $\gamma > 0$ is the dephasing rate. The global state is the thermal state $\omega = Z^{-1} e^{-\beta \hat{H}}$.

Identification of $\mathcal{A}_{\mathbf{c}}$

In the strong dephasing limit $\gamma \gg J, h$, off-diagonal coherences in the σ_z basis are rapidly destroyed. The decoherence-stable observables form the *pointer basis algebra*:

$$\mathcal{A}_{\mathbf{c}} = \{f(\sigma_z^{(1)}, \ldots, \sigma_z^{(n)}) : f \text{ is a polynomial}\}'' = \mathcal{A}_{\mathrm{diag}}, \tag{4.22}$$

the maximal abelian subalgebra (MASA) of $\mathcal{B}(\mathcal{H}_U)$ associated with the computational basis.

Verification of (U1): Faithfulness

The thermal state $\omega = Z^{-1}e^{-\beta\hat{H}}$ is a full-rank density matrix for any finite β (every eigenvalue of $e^{-\beta\hat{H}}$ is strictly positive). Its restriction to $\mathcal{A}_{\mathrm{diag}}$ assigns non-zero probability to every computational-basis state:

$$\omega(|s\rangle\langle s|) = \langle s|\rho_{\mathrm{th}}|s\rangle > 0 \qquad \forall\, s, \tag{4.23}$$

where the strict positivity follows from $\rho_{\mathrm{th}} = Z^{-1}e^{-\beta\hat{H}}$ being positive definite. Note that $|s\rangle$ are computational-basis (not energy-eigen-) states; the diagonal matrix elements of a positive-definite operator are strictly positive regardless of the choice of basis.

Verification of (U2): Modular stability

For the finite-dimensional system with faithful thermal state ρ_{th}, the ambient modular flow acts on $\mathcal{B}(\mathcal{H}_U)$ as

$$\sigma_t^{\omega,\mathrm{tot}}(X) = \rho_{\mathrm{th}}^{it}\, X\, \rho_{\mathrm{th}}^{-it} \qquad \forall\, X \in \mathcal{B}(\mathcal{H}_U). \tag{4.24}$$

For diagonal operators $A = \sum_s a_s|s\rangle\langle s| \in \mathcal{A}_{\mathrm{diag}}$, we compute

$$\sigma_t^{\omega,\mathrm{tot}}(A) = \rho_{\mathrm{th}}^{it}\left(\sum_s a_s|s\rangle\langle s|\right)\rho_{\mathrm{th}}^{-it} = \sum_s a_s\,(\rho_{\mathrm{th}}^{it}|s\rangle)(\langle s|\rho_{\mathrm{th}}^{-it}). \tag{4.25}$$

Since $\rho_{\mathrm{th}} = Z^{-1}e^{-\beta\hat{H}}$ and the computational basis $\{|s\rangle\}$ is *not* the energy eigenbasis (when $h \neq 0$), the vectors $\rho_{\mathrm{th}}^{it}|s\rangle$ are non-trivial superpositions, and $\sigma_t^{\omega,\mathrm{tot}}(A) \notin \mathcal{A}_{\mathrm{diag}}$ in general.

However, in the strong-dephasing regime $\gamma \gg J, h$, the effective steady-state density matrix converges to a diagonal form $\rho_{\mathrm{ss}} \approx \sum_s p_s|s\rangle\langle s|$, for which $\rho_{\mathrm{ss}}^{it}|s\rangle = p_s^{it}|s\rangle$ and (U2) is satisfied exactly. This illustrates both the content of (U2)—it imposes a genuine compatibility condition between the algebra and the state—and its limitations: the qubit chain example is a partial verification, with (U1) and (U3) satisfied exactly and (U2) satisfied approximately in the strong-dephasing regime. Strictly, $\mathcal{A}_{\mathrm{diag}}$ is a modular accessible algebra for the dephasing-dominated steady state ρ_{ss}, not for an arbitrary thermal state of the full Hamiltonian.

Verification of (U3): Maximality

Any extension $\mathcal{A}' \supsetneq \mathcal{A}_{\mathrm{diag}}$ must contain off-diagonal operators $|s\rangle\langle s'|$ with $s \neq s'$. Under the dephasing part of the Lindbladian alone, such operators decay as

$$e^{t\mathcal{L}_{\mathrm{deph}}}(|s\rangle\langle s'|) = e^{-2\gamma\, d(s,s')\, t}\, |s\rangle\langle s'|, \tag{4.26}$$

where $d(s, s')$ is the Hamming distance between strings s and s'. In the strong-dephasing regime $\gamma \gg J, h$, the full Lindbladian $\mathcal{L} = \mathcal{L}_H + \mathcal{L}_{\text{deph}}$ causes off-diagonal elements to decay at rate $2\gamma\, d(s, s')$ to leading order, with Hamiltonian-induced corrections of order J/γ and h/γ. The fixed-point algebra of the dephasing semigroup is precisely $\mathcal{A}_{\text{diag}}$, since only diagonal operators are invariant. Any modular-accessible algebra satisfying (U3) must be contained in $\mathcal{A}_{\text{fix}} = \mathcal{A}_{\text{diag}}$, and since $\mathcal{A}_{\text{diag}}$ already satisfies (U1)–(U2), it is maximal.

Physical content

The accessible algebra is the pointer basis algebra: the set of observables that survive decoherence. The modular flow acts trivially on the pointer basis (since the algebra is abelian), consistent with the fact that classical variables do not undergo non-trivial modular evolution. The entanglement entropy of a subsystem $A \subset \{1, \ldots, n\}$ within $\mathcal{A}_{\text{diag}}$ reduces to the classical Shannon entropy:

$$S_A = -\sum_{s_A} p(s_A) \ln p(s_A), \qquad p(s_A) = \sum_{s_{\bar{A}}} p(s). \tag{4.27}$$

This example demonstrates that the modular accessible algebra framework correctly identifies the decoherence-preferred observables in a finite-dimensional open quantum system.

4.5.6 Connection to Linearized Gravity

Having established the algebraic framework and verified it in concrete models, we outline the connection to gravitational dynamics.

Entanglement first law

Theorem 4.30 (Entanglement first law). *Let $\mathcal{M}$ be a von Neumann algebra with cyclic and separating vector $|\Omega\rangle$, and let $|\Psi\rangle = |\Omega\rangle + \varepsilon|\chi\rangle + O(\varepsilon^2)$ be a nearby state. Let $K_{\text{mod}} = -\ln \Delta_\Omega$ be the modular Hamiltonian. Then, to first order in ε:*

$$\delta S_{\mathcal{M}} = \delta\langle K_{\text{mod}}\rangle, \tag{4.28}$$

where $\delta S_{\mathcal{M}}$ is the change in entanglement entropy and $\delta\langle K_{\text{mod}}\rangle$ is the change in the expectation value of the modular Hamiltonian.

Proof sketch. The relative entropy $S(\rho^\Psi \| \rho^\Omega) = -S(\rho^\Psi) + \langle K_{\text{mod}}\rangle_\Psi + \text{const}$ is non-negative and vanishes for $|\Psi\rangle = |\Omega\rangle$. The first variation at $\varepsilon = 0$ gives $0 = -\delta S_{\mathcal{M}} + \delta\langle K_{\text{mod}}\rangle$, yielding $\delta S_{\mathcal{M}} = \delta\langle K_{\text{mod}}\rangle$. □

From algebra perturbation to modular Hamiltonian perturbation

In the HAFF framework, the accessible algebra $\mathcal{A}_{\mathbf{c}}$ evolves as the global state changes. Consider a one-parameter family $\mathcal{A}_{\mathbf{c}}(\lambda)$, with $\mathcal{A}_{\mathbf{c}}(0) = \mathcal{A}_{\mathbf{c}}$. Two sources of perturbation contribute:

(a) **State perturbation (fixed algebra):** The global state changes, $|\Psi_U\rangle \to |\Psi'_U\rangle$, but the algebra remains fixed. The modular Hamiltonian changes via the Connes cocycle Radon–Nikodym theorem [24, 100]:

$$\Delta^{it}_{\Psi'} = (D\omega' : D\omega)_t\, \Delta^{it}_{\Psi}, \tag{4.29}$$

where $(D\omega' : D\omega)_t$ is the Connes cocycle.

(b) **Algebra perturbation (fixed state):** The subalgebra satisfying the stability conditions shifts. In the language of half-sided modular inclusions (Theorem 4.22), the perturbation can be described by

$$\mathcal{A}'_{\mathbf{c}} = e^{-i\delta a\, P} \mathcal{A}_{\mathbf{c}}\, e^{i\delta a\, P}, \qquad \delta a \ll 1, \tag{4.30}$$

yielding

$$\delta K_{\mathrm{mod}} = -i\delta a\, [P, K_{\mathrm{mod}}] + O(\delta a^2). \tag{4.31}$$

We emphasize that this formula applies specifically to HSMI-type algebra deformations (null translations of wedge algebras in the Rindler setting). For generic algebra perturbations, the modular Hamiltonian response requires the full Connes cocycle/Araki perturbation theory, which is substantially more involved. The derivation chain below therefore applies rigorously only in settings where the algebra perturbation has HSMI structure.

In the general case (combined perturbation):

$$\delta K_{\mathrm{mod}} = \delta K_{\mathrm{state}} + \delta K_{\mathrm{algebra}}. \tag{4.32}$$

Linearized Einstein equations from the entanglement first law

The key result connecting algebraic perturbation to gravitational dynamics is due to Faulkner et al. [32], with complementary arguments by Jacobson [53].

Theorem 4.31 (Linearized gravity from entanglement, [32])**.** *In a holographic CFT dual to Einstein gravity in asymptotically AdS spacetime, the entanglement first law $\delta S = \delta\langle K_{\mathrm{mod}}\rangle$, applied to all ball-shaped regions on the boundary, is equivalent to the linearized Einstein equations in the bulk:*

$$G^{(1)}_{\mu\nu} + \Lambda g^{(1)}_{\mu\nu} = 8\pi G_N\, T^{(1)}_{\mu\nu}. \tag{4.33}$$

The derivation uses the explicit form of the modular Hamiltonian for a ball-shaped boundary region [23]:

$$K^B_{\mathrm{mod}} = 2\pi \int_B d^{d-1}x\, \frac{R^2 - |\mathbf{x} - \mathbf{x}_0|^2}{2R}\, T_{00}(x), \tag{4.34}$$

together with the Ryu–Takayanagi formula and the JLMS relation [54] between bulk and boundary modular Hamiltonians.

Conditional result within the HAFF framework

Corollary 4.32 (Linearized gravity from accessible algebra perturbation)**.** *Let $\mathcal{A}_{\mathbf{c}}$ be a modular accessible algebra satisfying* (U1)–(U3) *in a holographic CFT vacuum state, and assume:*

(H1) The holographic correspondence (AdS/CFT duality) holds.

(H2) The Ryu–Takayanagi formula and its quantum corrections hold.

(H3) The JLMS formula relating bulk and boundary modular Hamiltonians holds.

Then a perturbation of the accessible algebra along a one-parameter family $\mathcal{A}_{\mathbf{c}}(\lambda)$ with $\mathcal{A}_{\mathbf{c}}(0) = \mathcal{A}_{\mathbf{c}}$ induces a perturbation of the modular Hamiltonian δK_{mod}, and the entanglement first law

$$\delta S = \delta \langle K_{\mathrm{mod}} \rangle \tag{4.35}$$

is equivalent to the linearized Einstein equations in the holographic bulk. The chain of implications is:

$$\delta \mathcal{A}_{\mathbf{c}} \xrightarrow{\textit{modular theory}} \delta K_{\mathrm{mod}} \xrightarrow{\textit{first law}} \delta S = \delta \langle K_{\mathrm{mod}} \rangle \xrightarrow{\textit{(H1)–(H3)}} G^{(1)}_{\mu\nu} = 8\pi G_N T^{(1)}_{\mu\nu}. \tag{4.36}$$

In the HAFF language: the adiabatic flow of the accessible algebra IS linearized gravitational dynamics.

Remark 4.33. *The new content of this result is limited to identifying the input of the Faulkner et al. derivation (modular Hamiltonian perturbation) with the output of HAFF algebra evolution ($\delta \mathcal{A}_{\mathbf{c}} \to \delta K_{\mathrm{mod}}$ via the HSMI formula). The gravitational content is entirely provided by the holographic assumptions (H1)–(H3).*

The circularity issue

A fundamental objection to any "gravity from entanglement" program is circularity: Jacobson's derivation [52] presupposes a background geometry, but HAFF claims geometry emerges from the algebra. Three resolutions are available:

(A) **Algebraic resolution (Wiesbrock).** "Wedge regions" are defined purely algebraically via modular inclusions, without reference to a background metric. Wiesbrock's theorem (Theorem 4.22) recovers translations from half-sided modular inclusions; a sufficient net of such inclusions reconstructs the full Poincaré group.

(B) **Bootstrap resolution.** Start with a seed geometry, derive linearized Einstein equations via the entanglement first law, update the geometry, and iterate.

(C) **Holographic resolution (AdS/CFT).** The boundary CFT provides the algebra without reference to bulk geometry; the bulk geometry is entirely derived from boundary data.

In Corollary 4.32, we adopt resolution (C). Resolution (A) provides the most promising path for a geometry-free formulation in subsequent work.

4.6 Relation to Existing Approaches

The perspective developed in this paper does not compete with existing approaches to quantum gravity and emergent spacetime. Rather, it may be understood as offering a *conceptual umbrella* under which several distinct research programs can be situated. We briefly discuss four such connections.

4.6.1 AdS/CFT and Holographic Duality

The AdS/CFT correspondence [68] provides the most concrete realization of geometry emerging from quantum entanglement. In this framework, a $(d+1)$-dimensional gravitational theory in anti-de Sitter space is dual to a d-dimensional conformal field theory on its boundary.

The Ryu-Takayanagi formula [83] and its generalizations establish that geometric quantities in the bulk (areas of extremal surfaces) correspond to entanglement entropies in the boundary theory:

$$S_A = \frac{\text{Area}(\gamma_A)}{4G_N}. \tag{4.37}$$

Within the present framework, AdS/CFT may be viewed as a specific instance of the general principle that geometry emerges from entanglement structure. The boundary CFT defines a particular accessible algebra, and the bulk geometry is the effective geometry induced by that algebra.

Remark 4.34 (Not a Replacement)**.** *We do not claim that the present framework explains or derives AdS/CFT. Rather, AdS/CFT provides concrete evidence that the structural relationship between accessible algebras and effective geometry—which we propose as general—is realized in at least one well-understood setting.*

4.6.2 Tensor Networks and MERA

Tensor network constructions, particularly the Multi-scale Entanglement Renormalization Ansatz (MERA) [106, 99], provide discrete models in which geometry emerges from entanglement structure.

In MERA, a quantum state is constructed by successive layers of disentanglers and isometries. The network structure itself defines an effective geometry: the "depth" direction in the network corresponds to a radial direction in an emergent spacetime, with properties reminiscent of AdS geometry.

This construction illustrates concretely how:

- A choice of coarse-graining (the tensor network structure) determines entanglement patterns.
- Entanglement patterns induce effective geometric relationships.
- Different network structures yield different effective geometries from the same boundary data.

The present framework generalizes this observation: tensor networks are specific implementations of coarse-graining structures, and MERA-type emergence is a special case of the algebra-to-geometry correspondence we propose.

4.6.3 Jacobson's Thermodynamic Derivation

Jacobson's remarkable result [52] showed that Einstein's field equations can be derived from thermodynamic considerations applied to local Rindler horizons, assuming the Bekenstein-Hawking entropy formula and the Clausius relation $\delta Q = T\,dS$.

This derivation suggests that gravity may be "thermodynamic"—an effective description arising from coarse-graining over microscopic degrees of freedom, rather than a fundamental force.

The present perspective is consonant with Jacobson's approach:

- Both treat gravitational dynamics as emergent rather than fundamental.
- Both connect gravity to entropy and information-theoretic quantities.
- Both suggest that the Einstein equations describe effective, coarse-grained physics.

The contribution of the present work is to embed this intuition within a more general framework: the selection of accessible algebras as the structural origin of effective geometry.

4.6.4 Background Independence in Loop Quantum Gravity

Loop quantum gravity [82, 101] pursues quantization of gravity while maintaining background independence—the principle that physical laws should not depend on a fixed spacetime metric.

The present framework shares this commitment to background independence, but approaches it differently:

- Loop quantum gravity seeks to quantize the metric directly, constructing spacetime from spin networks.
- The present approach treats spacetime as an effective structure emergent from accessible algebra selection.

These are not mutually exclusive. It is conceivable that spin network states could be understood as specific implementations of accessible algebras, with loop quantum gravity dynamics describing transitions between such algebras. We do not develop this connection here, but note it as a direction for future investigation.

4.6.5 Summary: A Conceptual Umbrella

Approach	**Key Mechanism**	**Relation to Present Work**
AdS/CFT	Holographic duality	Specific instance of algebra $\rightarrow$ geometry
Tensor Networks	Discrete entanglement structure	Concrete implementation of coarse-graining
Jacobson	Thermodynamic derivation	Consonant emergent perspective
Loop QG	Background-independent quantization	Shared commitment, different strategy

Table 4.1: Relation of the present framework to existing approaches. The present work does not replace any of these programs, but offers a unifying structural perspective.

We emphasize that the present framework does not claim superiority over these approaches. Each addresses aspects of quantum gravity that the present structural analysis does not. Our contribution is to articulate a perspective in which these diverse programs may be seen as exploring different facets of a common structural insight: that gravity is connected to the selection of how quantum degrees of freedom are organized into effective subsystems.

4.7 Explicit Scope Limitations

To ensure clarity regarding the claims of this paper, we state explicitly what it does and does not assert.

4.7.1 What This Paper Claims

1. **Categorical distinction:** Gauge forces and gravity are distinguished at the level of their relation to subsystem decomposition—gauge forces operate within a fixed decomposition, while gravitational phenomena reflect the evolution of the decomposition itself.

2. **Generative mechanism:** Gravitational dynamics arises from the adiabatic flow of accessible algebras as the global quantum state evolves (Conjecture 4.10).

3. **Equivalence principle:** The universality of gravitational coupling follows from the universality of the observable algebra—all accessible matter inhabits the geometry defined by $\mathcal{A}_{\mathbf{c}}$.

4. **Information-geometric metric:** The emergent spacetime metric can be identified with the Quantum Fisher Information Metric on the space of effective descriptions.

5. **Conceptual umbrella:** Several existing research programs (holography, tensor networks, thermodynamic gravity) may be situated under this common structural framework.

4.7.2 What This Paper Does Not Claim

1. **No new dynamics:** We do not propose equations of motion, Lagrangians, or dynamical principles beyond those already established.

2. **No unconditional derivation of Einstein equations:** The linearized result (Section 4.5.6) is conditional on holographic assumptions (H1)–(H3); we do not derive general relativity from first principles.

3. **No empirical predictions:** We do not offer testable predictions that distinguish this perspective from standard approaches.

4. **No resolution of quantum gravity:** We do not claim to solve the problem of quantum gravity; we offer a diagnostic perspective, not a solution.

5. **No observer-dependence:** The framework does not render gravity subjective or observer-relative. Accessible algebras are constrained by physical criteria, not by epistemic states of observers.

6. **No interpretational commitments:** The analysis is compatible with various interpretations of quantum mechanics and does not require commitment to any particular one.

4.8 Open Questions

The algebraic foundation established in Section 4.5 sharpens the open problems facing the HAFF gravity program. The following are the immediate targets for subsequent work.

4.8.1 Linearized Einstein Equations without Holography

Corollary 4.32 derives the linearized Einstein equations conditionally, assuming the holographic correspondence (H1)–(H3). A fully general derivation from algebraic perturbation theory would require establishing a "bulk reconstruction" from the modular data of $\mathcal{A}_{\mathbf{c}}$ and its perturbations, using Wiesbrock-type inclusions (Theorem 4.22) to define spacetime translations algebraically. This would remove the dependence on AdS/CFT and extend the result to non-holographic settings.

4.8.2 Moduli Space when Ergodicity Fails

The uniqueness conjecture (Conjecture 4.27) requires the ergodicity assumption (E). When (E) fails—in systems with spontaneous symmetry breaking, topological order, or phase coexistence—the space of accessible algebras acquires a non-trivial moduli space (Remark 4.28). Characterizing this moduli space, and determining whether it carries a natural metric (e.g., the Fisher information metric on modular Hamiltonians) that encodes the geometry of the space of effective descriptions, is an important structural problem.

4.8.3 Computing δK_{mod} beyond Linear Order

The connection to linearized gravity (Section 4.5.6) uses only the first-order perturbation of the modular Hamiltonian. Computing δK_{mod} explicitly for perturbations of the accessible algebra in specific models, beyond the linear order, is essential for accessing nonlinear gravitational dynamics. The second-order correction may contain information about the gravitational coupling constant G_N and matter content.

4.8.4 The Backreaction Problem

How does the change in bulk geometry (induced by $\delta\mathcal{A}_{\mathbf{c}}$) feed back into the boundary conditions that determine $\mathcal{A}_{\mathbf{c}}$? This self-consistency condition may select a unique trajectory $\mathcal{A}_{\mathbf{c}}(t)$ and hence a unique gravitational dynamics. Resolving this backreaction loop is the central challenge for deriving the full nonlinear Einstein equations from the algebraic framework.

4.9 Conclusion

We have proposed a structural framework in which gravitational phenomena arise from the adiabatic evolution of accessible observable algebras as the global quantum state evolves.

The core claims are:

1. **Categorical distinction:** Gauge forces describe dynamics *within* a fixed algebra $\mathcal{A}_{\mathbf{c}}$; gravity describes the evolution *of* $\mathcal{A}_{\mathbf{c}}$ itself.

2. **Generative mechanism:** As the global state $|\Psi_U(t)\rangle$ evolves, stability conditions select different optimal algebras $\mathcal{A}_{\mathbf{c}}(t)$. This flow manifests as spacetime curvature.

3. **Equivalence principle:** All observable matter couples universally to gravity because all observables are, by definition, elements of the same algebra $\mathcal{A}_{\mathbf{c}}$.

4. **Information geometry:** The emergent metric is the Quantum Fisher Information Metric on the manifold of effective descriptions.

This framework does not derive the Einstein equations from first principles, nor does it resolve the problem of quantum gravity. However, it offers more than a diagnostic: it proposes a *generative mechanism* that explains why gravity has the structural features it does—universality, dynamical geometry, resistance to naive quantization.

The perspective is consistent with, and provides a conceptual umbrella for, existing research programs: holographic duality (where boundary entanglement encodes bulk geometry), tensor networks (where network structure induces effective geometry), and thermodynamic approaches (where Einstein equations emerge from entropy considerations).

The algebraic foundation developed in Section 4.5 partially formalizes this framework, translating structural intuitions into mathematically precise (though not yet fully proven) statements. By translating the three physical accessibility criteria into Tomita–Takesaki modular theory—faithfulness (U1), modular stability (U2), and maximality (U3)—we conjectured that the accessible algebra is unique up to unitary equivalence under ergodicity assumptions (Conjecture 4.27), identified the specific gaps (G1–G2) that remain to be closed, and verified the construction in both the Rindler wedge (reproducing Bisognano–Wichmann and Unruh physics) and a qubit chain with decoherence (recovering the pointer basis algebra). Most significantly, we established the derivation chain

$$\delta\mathcal{A}_{\mathbf{c}} \;\longrightarrow\; \delta K_{\mathrm{mod}} \;\longrightarrow\; \delta S = \delta\langle K_{\mathrm{mod}}\rangle \;\longrightarrow\; G^{(1)}_{\mu\nu} = 8\pi G_N\, T^{(1)}_{\mu\nu}, \tag{4.38}$$

showing that perturbations of the accessible algebra, via modular Hamiltonian perturbation and the entanglement first law, yield the linearized Einstein equations in holographic settings (Corollary 4.32). This makes precise the HAFF gravity conjecture at the linearized level: in holographic settings satisfying (H1)–(H3), the adiabatic flow of the accessible algebra *is* linearized gravitational dynamics, conditionally derived rather than postulated. Removing the holographic assumptions (H1)–(H3) remains an open problem (Section 4.8).

We conclude with a reflection. The difficulty of quantizing gravity may not be purely technical. If gravity is the evolution of the stage on which quantum mechanics is performed, rather than an actor on that stage, then quantizing gravity requires quantizing the framework of quantization itself. This is not a problem to be solved by better regularization schemes, but a conceptual challenge requiring us to think beyond fixed subsystem decompositions.

The path forward may lie not in quantizing forces, but in understanding what determines the structure of accessibility—and how that structure flows.

Chapter 5

Measurement as Accessibility

Paper E

Originally published: Zenodo, DOI: 10.5281/zenodo.18400065

Abstract

We propose that quantum measurement is not a primitive process but a manifestation of accessibility constraints on operator algebras. Building on the Holographic Alaya-Field Framework (HAFF), we identify three structural constraints—interaction coupling, dynamical stability, and environmental redundancy—that jointly determine which observables are accessible within a given physical context. Measurement outcomes are reinterpreted as the eigenvalue structure of operators satisfying these constraints, and the definiteness of outcomes is traced to redundant environmental encoding rather than wave function collapse. This framework maintains compatibility with unitary quantum mechanics while providing a structural account of why certain observables acquire definite values. The analysis is structural in nature: we do not propose new dynamics or modifications to quantum mechanics, but clarify the conditions under which measurement-like phenomena emerge from algebraic constraints. Connections to decoherence theory, quantum Darwinism, and algebraic quantum field theory are discussed, along with explicit non-claims to prevent misinterpretation.

5.1 Introduction

5.1.1 The Measurement Problem Reconsidered

The quantum measurement problem has resisted resolution for nearly a century. In its sharpest form, the problem asks: how do definite measurement outcomes arise from quantum states that, prior to measurement, assign non-trivial amplitudes to multiple possibilities? Standard quantum mechanics provides rules for computing outcome probabilities but does not explain the transition from superposition to definiteness.

Various approaches have been proposed: collapse postulates that modify unitary evolution, many-worlds interpretations that deny the uniqueness of outcomes, and decoherence-based accounts that explain the suppression of interference without addressing the selection of particular results. Each approach has merits, but none has achieved consensus.

5.1.2 A Structural Reframing

This paper proposes a different perspective. Rather than asking how measurement *causes* definite outcomes, we ask: under what structural conditions do observables *acquire* the status of being measurable in the first place?

The central claim is:

> **Measurement is not a primitive process, but a manifestation of accessibility constraints on operator algebras.**

Within the Holographic Alaya-Field Framework (HAFF) developed in previous papers [134, 135, 136, 137], physical descriptions are formulated relative to accessible observable algebras $\mathcal{A}_{\mathbf{c}} \subset \mathcal{B}(\mathcal{H}_U)$. Not all mathematically definable operators correspond to physically realizable measurements. The present paper identifies three structural constraints that jointly determine which operators are accessible:

1. **Interaction Constraint**: The observable must couple to external degrees of freedom.

2. **Stability Constraint**: The observable must persist under dynamical evolution.

3. **Redundancy Constraint**: Information about the observable must be redundantly encoded in the environment.

Observables satisfying all three constraints constitute the accessible algebra. Measurement outcomes are then understood as the eigenvalue structure of these accessible observables, and definiteness arises from the redundancy of environmental records rather than from any modification of unitary dynamics.

5.1.3 Scope and Limitations

We emphasize what this paper does and does not attempt.

This paper does:

- Provide a structural characterization of measurement in terms of algebraic accessibility

- Identify three physical constraints that jointly determine accessible observables

- Connect measurement to established frameworks (decoherence, quantum Darwinism, AQFT)

- Maintain compatibility with unitary quantum mechanics

This paper does not:

- Propose new dynamics or modifications to quantum mechanics

- Explain why specific measurement outcomes occur (the "outcome problem")
- Resolve interpretational debates about the ontology of quantum states
- Invoke consciousness, observers, or subjective elements

The analysis is structural: we clarify conditions under which measurement-like phenomena emerge, without claiming to have solved the measurement problem in its deepest form.

5.1.4 Outline

Section 5.2 reviews relevant background on algebraic approaches to quantum mechanics and the HAFF framework. Section 5.3 develops the three accessibility constraints in detail. Section 5.4 reframes measurement in terms of these constraints. Section 5.5 discusses connections to existing approaches. Section 5.6 states explicit non-claims. Section 5.7 situates the paper within the broader HAFF program. Section 5.8 concludes.

5.2 Background

5.2.1 Algebraic Approaches to Quantum Mechanics

In the algebraic formulation of quantum mechanics, the fundamental objects are not wave functions or Hilbert spaces, but algebras of observables. A quantum system is characterized by a $*$-algebra $\mathcal{A}$ of bounded operators, and states are positive linear functionals on $\mathcal{A}$ [43, 5].

This perspective has several advantages. It does not presuppose a specific Hilbert space representation, accommodates systems with infinitely many degrees of freedom, and naturally incorporates superselection rules. Most importantly for our purposes, it treats the specification of observables as logically prior to the specification of states.

5.2.2 Observable Algebras in AQFT

In algebraic quantum field theory (AQFT), local observable algebras $\mathcal{A}(\mathcal{O})$ are associated with spacetime regions $\mathcal{O}$, without invoking a global tensor product structure [43]. The key insight is that subsystem structure emerges from the algebra of observables rather than being presupposed.

The HAFF framework extends this perspective by treating the selection of accessible algebras as physically constrained rather than given. Building on foundational work demonstrating that tensor product structures are observable-induced [119, 120], different physical contexts—characterized by different interaction structures, stability conditions, and environmental couplings—yield different accessible algebras, and hence different effective physical descriptions.

5.2.3 Accessible Algebras in HAFF

Following [134, 135], we define:

Definition 5.1 (Accessible Algebra)**.** *An **accessible algebra** $\mathcal{A}_\mathbf{c} \subset \mathcal{B}(\mathcal{H}_U)$ is a $*$-subalgebra satisfying physical constraints that ensure its elements correspond to operationally realizable observables within a given context* $\mathbf{c}$*.*

The subscript **c** denotes the *context*—the totality of physical conditions (interaction Hamiltonian, environmental structure, timescales) that determine which observables are accessible. Different contexts yield different accessible algebras from the same underlying Hilbert space.

The present paper specifies three constraints that jointly determine $\mathcal{A}_\mathbf{c}$.

5.3 Accessibility as Physical Constraint

We now develop the three constraints that determine which observables belong to the accessible algebra.

5.3.1 Constraint 1: Interaction Coupling

Constraint 5.2 (Interaction)**.** *An observable $\hat{O} \in \mathcal{B}(\mathcal{H}_U)$ satisfies the **interaction constraint** if it couples non-trivially to external degrees of freedom via the interaction Hamiltonian:*

$$[\hat{O}, \hat{H}_{int}] \neq 0. \tag{5.1}$$

Physical interpretation. An observable that commutes with all interaction terms is dynamically inert: it cannot be probed, recorded, or correlated with any external system. Such observables are mathematically well-defined but physically inaccessible.

Relation to measurement. Measurement requires that the system observable become correlated with apparatus degrees of freedom. This correlation is mediated by interaction. Observables that do not couple to any external system cannot, even in principle, be measured.

Remark 5.3. *The interaction constraint is necessary but not sufficient for accessibility. An observable may couple to external degrees of freedom yet fail to satisfy stability or redundancy requirements.*

Remark 5.4 (Conserved quantities)**.** *Conserved quantities that commute with H_{int} (e.g., total energy or charge in number-conserving systems) would fail this criterion. In such cases, the constraint should be interpreted as requiring that the observable can be* correlated *with external degrees of freedom, whether through direct interaction coupling or through pre-existing correlations and boundary conditions.*

5.3.2 Constraint 2: Dynamical Stability

Constraint 5.5 (Stability)**.** *An observable $\hat{O}$ satisfies the **stability constraint** if it remains approximately invariant under physically relevant dynamical maps $\mathcal{E}_t$:*

$$\|\mathcal{E}_t(\hat{O}) - \hat{O}\| < \epsilon \tag{5.2}$$

for timescales t relevant to the physical process under consideration.

Physical interpretation. Observables that scramble rapidly—spreading their information across many degrees of freedom faster than any recording process can track—cannot be reliably measured. Stability ensures that the observable persists long enough to be correlated with records.

Relation to scrambling. In the language of quantum chaos, stable observables are those with slow out-of-time-order correlator (OTOC) growth [48]:

$$\langle [\hat{O}(t), \hat{V}(0)]^2 \rangle \ll 1 \quad \text{for } t \ll \tau_{\text{scrambling}}. \tag{5.3}$$

Observables satisfying this condition resist rapid delocalization and maintain their identity under dynamical evolution.

Relation to decoherence. The stability constraint is closely related to the selection of pointer observables in decoherence theory [121]. Pointer observables are those that remain stable under system-environment interaction, forming the preferred basis in which the density matrix becomes approximately diagonal.

Remark 5.6 (Threshold ϵ)**.** *The threshold ϵ is not a fundamental constant but depends on the physical context: the precision of available recording mechanisms, the timescales of interest, and the noise level of the environment. This context-dependence is a feature, not a bug—it reflects the operational nature of accessibility.*

5.3.3 Constraint 3: Environmental Redundancy

Constraint 5.7 (Redundancy)**.** *An observable $\hat{O}$ satisfies the **redundancy constraint** if information about $\hat{O}$ is redundantly encoded across multiple independent environmental fragments $\{E_k\}$:*

$$I(\hat{O} : E_k) \approx H(\hat{O}) \quad \textit{for many } k, \tag{5.4}$$

where $I(\cdot : \cdot)$ denotes quantum mutual information and $H(\cdot)$ denotes von Neumann entropy.

Physical interpretation. Redundancy ensures that information about the observable is not localized in a single environmental degree of freedom but is broadcast across many independent fragments. This makes the information robust and intersubjectively accessible: multiple independent observers can extract the same information without disturbing each other's records.

Relation to quantum Darwinism. The redundancy constraint formalizes the central insight of quantum Darwinism [122]: classical objectivity arises when information about a system is redundantly imprinted on the environment. Observables satisfying this constraint are precisely those for which multiple observers can agree on measurement outcomes.

Operational significance. Redundancy distinguishes *objective* from *subjective* information. An observable whose information is encoded in only a single environmental fragment is accessible to at most one observer; different observers would obtain different, incompatible records. Redundancy ensures that the observable's value is a matter of intersubjective fact.

Remark 5.8 (Relation to classical objectivity)**.** *The redundancy constraint provides a structural account of why certain observables behave "classically": their values are recorded multiply and independently, making them robust against local perturbations and accessible to multiple agents.*

5.3.4 The Accessible Algebra

Definition 5.9 (Accessible Algebra via Constraints)**.** *The* ***accessible algebra*** $\mathcal{A}_{\mathbf{c}}$ *relative to context* $\mathbf{c}$ *is the set of all observables satisfying Constraints 5.2, 5.5, and 5.7:*

$$\mathcal{A}_{\mathbf{c}} = \{\hat{O} \in \mathcal{B}(\mathcal{H}_U) : \hat{O} \textit{ satisfies Constraints 1, 2, and 3}\}. \tag{5.5}$$

Strictly, the set of operators individually satisfying all three constraints need not form a $*$-algebra (algebraic closure under sums and products is not guaranteed). We define $\mathcal{A}_{\mathbf{c}}$ as the von Neumann algebra *generated by* the operators satisfying the three accessibility constraints. The generating operators satisfy the constraints individually; the algebraic closure step is an idealization whose validity in specific models remains to be verified.

The accessible algebra is not fixed *a priori* but is determined by the physical context. Different interaction Hamiltonians, environmental structures, and timescales yield different accessible algebras from the same underlying Hilbert space.

Constraint	**Condition**	**Physical Meaning**
Interaction	$[\hat{O}, \hat{H}_{\text{int}}] \neq 0$	Observable couples to external degrees of freedom
Stability	$\|\mathcal{E}_t(\hat{O}) - \hat{O}\| < \epsilon$	Observable persists under dynamics
Redundancy	$I(\hat{O} : E_k) \approx H(\hat{O})$	Information redundantly encoded

Table 5.1: Summary of the three accessibility constraints.

5.4 Measurement Reframed

We now apply the accessibility framework to reinterpret quantum measurement.

5.4.1 What Can Be Measured

Within the present framework, the question "What can be measured?" receives a precise answer:

> **An observable can be measured if and only if it belongs to the accessible algebra $\mathcal{A}_{\mathbf{c}}$.**

Observables outside $\mathcal{A}_{\mathbf{c}}$—those failing one or more of the three constraints—are not measurable within context $\mathbf{c}$, regardless of their mathematical definition. This does not mean they "do not exist" in any metaphysical sense, but that they do not correspond to operationally realizable measurements within the given physical context.

5.4.2 Measurement Outcomes

Given an accessible observable $\hat{O} \in \mathcal{A}_{\mathbf{c}}$, its measurement outcomes are identified with its eigenvalue structure:

> **Measurement outcomes are the eigenvalues of accessible observables.**

This identification is standard in quantum mechanics. The novelty lies in restricting attention to *accessible* observables: only those satisfying the three constraints yield operationally meaningful outcomes.

5.4.3 Definiteness from Redundancy

The definiteness of measurement outcomes—the fact that measurements yield single, definite results rather than superpositions—is traced to the redundancy constraint rather than to wave function collapse.

When an observable $\hat{O}$ satisfies the redundancy constraint, its eigenvalue is recorded in multiple independent environmental fragments. These records are mutually consistent: any fragment yields the same information about $\hat{O}$. This redundancy constitutes the objective, intersubjective definiteness of the measurement outcome.

> **Redundant environmental encoding explains the *effective* definiteness of measurement outcomes within each decoherence branch, but does not by itself select a unique outcome from the global superposition.**

The relationship between redundancy and absolute definiteness depends on one's interpretation of quantum mechanics. The global quantum state remains in superposition; what becomes definite is the content of redundant records, which all agree on the same eigenvalue within each branch.

5.4.4 The Outcome Problem

The framework does not explain why a *particular* eigenvalue is recorded rather than another. This "outcome problem" remains open:

> **We explain why outcomes are definite (redundancy), not why they are what they are.**

This limitation is shared with decoherence-based approaches. The present framework does not claim to resolve this aspect of the measurement problem, only to clarify the structural conditions under which definite outcomes become possible.

5.4.5 Measurement Without Observers

A crucial feature of the framework is that measurement is characterized without reference to observers, agents, or consciousness:

> **The "observer" is replaced by the "interaction context."**

Any physical system satisfying the three constraints—be it a photon counter, a mineral surface, or an interstellar dust grain—constitutes a "measurement site" for the relevant observables. Human observers are a special case, not a privileged category.

5.5 Relation to Existing Approaches

5.5.1 Decoherence Theory

Decoherence theory explains how interference between quantum states is suppressed through environmental entanglement [121]. The present framework is fully compatible with decoherence and may be viewed as extending it in two respects:

1. We make explicit the *conditions* under which decoherence selects a preferred basis (the stability and redundancy constraints).

2. We embed decoherence within the broader HAFF framework, connecting it to emergent geometry and gravitational phenomena.

5.5.2 Quantum Darwinism

Quantum Darwinism [122] emphasizes the role of environmental redundancy in establishing classical objectivity. The redundancy constraint (Constraint 5.7) formalizes this insight as a criterion for accessibility.

The present framework may be viewed as situating quantum Darwinism within an algebraic setting, treating redundancy as one of three jointly necessary conditions for observability rather than as a standalone principle.

5.5.3 QBism

QBism [34] interprets quantum states as expressions of an agent's beliefs. The present framework differs fundamentally: accessibility is determined by physical interaction structure, not by agent beliefs.

The key difference:

- **QBism**: Dependence on agent's epistemic state (belief-determined)

- **HAFF**: Dependence on interaction structure (interaction-determined)

Both reject naive realism about quantum states, but the present framework maintains objectivity by grounding accessibility in physical constraints rather than subjective beliefs.

5.5.4 Relational Quantum Mechanics

Relational quantum mechanics (RQM) [81] holds that quantum states are relational—defined only relative to a reference system. The present framework shares the emphasis on relationality but differs in its treatment of what grounds the relation:

- **RQM**: Relations between systems (system-relative)

- **HAFF**: Stability conditions on algebras (interaction-determined)

HAFF may provide the stable "nodes" required for RQM's relational network: before relations can exist, there must be relata stable enough to participate in interactions.

5.5.5 Algebraic Quantum Field Theory

The closest structural affinity is with algebraic quantum field theory (AQFT) [43]. Both frameworks treat observable algebras as primary and states as secondary. The present framework extends AQFT by:

1. Providing explicit criteria (the three constraints) for algebra selection
2. Connecting algebra selection to measurement and emergent geometry
3. Situating AQFT insights within the broader HAFF program

Approach	**Key Mechanism**	**Relation to HAFF**
Decoherence	Environmental entanglement	Compatible; constraints specify conditions
Quantum Darwinism	Redundant encoding	Redundancy constraint formalizes this
QBism	Agent beliefs	Categorically distinct; HAFF is interaction-determined
RQM	System relations	Complementary; HAFF provides stable relata
AQFT	Observable algebras	Closest affinity; HAFF adds selection criteria

Table 5.2: Relation of the present framework to existing approaches.

5.6 What This Paper Does NOT Claim

To prevent misinterpretation, we state explicitly what the paper does not claim.

1. **No resolution of the outcome problem.** We do not explain why particular measurement outcomes occur, only why outcomes are definite.
2. **No collapse postulate.** The framework assumes unitary evolution throughout. Definiteness arises from redundancy, not from non-unitary collapse.
3. **No modification of quantum mechanics.** We do not propose new equations, new dynamics, or modifications to the standard formalism.
4. **No consciousness or observer-dependence.** Accessibility is determined by physical constraints, not by conscious observers or epistemic states.
5. **No claim that all measurement problems are solved.** The framework addresses the definiteness problem but leaves other aspects (the preferred basis problem, the tails problem) to be addressed in conjunction with existing approaches.
6. **No claim of novelty regarding decoherence.** The framework builds on and is compatible with decoherence theory; it does not replace it.

7. **No claim of universal applicability.** The analysis is confined to the HAFF framework and does not assert that all approaches to measurement must adopt this structure.

8. **No metaphysical conclusions.** We do not claim that the accessible algebra exhausts reality, only that it exhausts what is operationally measurable within a given context.

9. **No derivation of Born rule.** The framework does not derive the Born rule for outcome probabilities; it assumes standard quantum probability.

10. **No claim about quantum-classical divide.** We do not assert a sharp boundary between quantum and classical; accessibility is context-dependent and admits degrees.

11. **No hidden variables.** The framework does not invoke hidden variables or additional ontology beyond standard quantum mechanics.

12. **No many-worlds commitment.** The framework is compatible with, but does not require, many-worlds interpretations.

13. **No claim of interpretational neutrality.** While the framework avoids some interpretational commitments, it does adopt the structural stance of HAFF, which may not be neutral with respect to all interpretations.

5.7 Connection to the HAFF Framework

5.7.1 Relation to Previous Papers

The present paper (Paper E) is part of a series developing the Holographic Alaya-Field Framework:

- **Paper A** [134]: Establishes that inequivalent coarse-graining structures induce inequivalent effective geometries from the same global quantum state.
- **Paper B** [135]: Clarifies the structural (vs. epistemic) nature of accessibility and situates HAFF relative to existing interpretations.
- **Paper C** [136]: Explores philosophical implications for causation, agency, and existence.
- **Paper D** [137]: Proposes that gravitational dynamics corresponds to the adiabatic evolution of accessible algebras.
- **Paper E** (this paper): Reframes measurement as a manifestation of accessibility constraints.
- **Paper F** (forthcoming): Addresses temporal asymmetry as accessibility propagation.

Paper	Phenomenon	Traditional View	HAFF Reframing
D	Gravity	Fundamental force	Evolution of accessible algebra
E	Measurement	Primitive process	Selection within accessible algebra
F	Time	Fundamental parameter	Direction of accessibility propagation

Table 5.3: The diagnostic triangle: gravity, measurement, and time reframed as aspects of algebraic accessibility.

5.7.2 The Diagnostic Triangle: D + E + F

Papers D, E, and F form a "diagnostic triangle" within the HAFF framework:

The unifying insight is that force, measurement, and time are not fundamental but are different projections of the structure of accessible algebras.

5.7.3 Structural Link to Gravity

Paper D establishes that gravity corresponds to the evolution of the accessible algebra $\mathcal{A}_{\mathbf{c}}(t)$. The present paper clarifies what determines $\mathcal{A}_{\mathbf{c}}$ at any given time: the three accessibility constraints.

The connection may be summarized as follows:

> If gravity (Paper D) describes how information maps to spatial curvature, then measurement (Paper E) reveals the pruning criterion that determines which information participates in that mapping. Without accessibility constraints, the holographic map would include non-physical operators and yield divergent geometry. The finiteness of gravity is grounded in the finiteness of the accessible algebra.

5.8 Conclusion

We have proposed that quantum measurement is not a primitive process but a manifestation of accessibility constraints on operator algebras.

The central results are:

1. **Three accessibility constraints**: Interaction coupling, dynamical stability, and environmental redundancy jointly determine which observables are accessible within a given physical context.

2. **Measurement reframed**: What can be measured is determined by membership in the accessible algebra; measurement outcomes are eigenvalues of accessible observables; definiteness arises from redundant environmental encoding.

3. **Observer-independence**: The framework characterizes measurement without reference to observers, consciousness, or subjective elements. The "observer" is replaced by the "interaction context."

4. **Compatibility**: The framework is compatible with unitary quantum mechanics, decoherence theory, and quantum Darwinism, while providing a structural account that connects measurement to the broader HAFF program.

The framework does not resolve all aspects of the measurement problem. It does not explain why particular outcomes occur, nor does it derive the Born rule. What it provides is a structural clarification: the conditions under which measurement-like phenomena emerge from physical constraints on operator algebras.

Within the HAFF program, measurement joins gravity and time as phenomena that are not fundamental but emerge from the structure of accessible algebras. This diagnostic unification does not constitute a Theory of Everything, but it suggests that seemingly disparate foundational puzzles may share a common structural origin.

Chapter 6

Temporal Asymmetry as Accessibility Propagation

Paper F

Originally published: Zenodo, DOI: 10.5281/zenodo.18400425

Abstract

We propose that temporal and causal asymmetry arise from the directional structure of accessibility propagation. Building on the Holographic Alaya-Field Framework (HAFF), which characterizes measurement as the selection of stable accessible algebras, we argue that the "arrow of time" is not a fundamental parameter but a consequence of how information spreads irreversibly into environmental degrees of freedom. The redundancy constraint central to accessibility—that information must be multiply recorded to be operationally accessible—is inherently asymmetric: information expands from few to many degrees of freedom, but the reverse process is statistically suppressed to the point of physical uninstantiability. This asymmetry defines a preferred direction that we identify with temporal ordering. No fundamental time parameter is assumed; all temporal ordering emerges from a partial order induced by algebraic inclusion and redundancy monotonicity. Causation is reframed as constraint propagation along this direction, with retrocausal trajectories being non-generic (measure zero) rather than forbidden by principle. A minimal mathematical model demonstrating irreversible redundancy expansion is provided in the appendix.

6.1 Introduction

6.1.1 The Problem of Time's Arrow

Among the deepest puzzles in physics is the origin of temporal asymmetry. The fundamental laws of physics—Newtonian mechanics, electromagnetism, quantum mechanics—are time-reversal invariant, with the notable exception of the weak interaction (CP violation in the CKM matrix, established experimentally in B-meson systems). However, this microscopic T-violation is quantitatively insufficient to explain the macroscopic arrow of time. Yet our experience of the world is profoundly asymmetric: eggs break but do not unbreak; we remember the past but not the future; causes precede effects.

This tension between microscopic reversibility and macroscopic irreversibility has been recognized since Boltzmann's work on statistical mechanics [16]. The standard resolution appeals to special initial conditions: the universe began in a low-entropy state, and the second law of thermodynamics reflects the statistical tendency to evolve toward higher entropy [75, 20].

While this explanation is widely accepted, it raises further questions:

- Why should initial conditions be "special"? What selects them?
- Is the thermodynamic arrow the only arrow, or are there independent sources of temporal asymmetry?
- In quantum gravity, where time itself may be emergent, how does any notion of "before" and "after" arise?

The present work does not claim to resolve these questions definitively. Instead, it offers a structural reframing: temporal asymmetry may be understood as a consequence of how accessible algebras propagate information.

6.1.2 Central Thesis

We propose that temporal asymmetry can be understood as follows:

> **Central Thesis:** No fundamental time parameter is assumed. All temporal ordering emerges from a partial order induced by algebraic inclusion and redundancy monotonicity. The "arrow of time" is the direction of irreversible accessibility propagation—information spreads from localized degrees of freedom into distributed environmental records, and this expansion is statistically irreversible.

This thesis builds on the accessibility framework developed in Paper E [138]. Recall that an observable is accessible only if information about it is redundantly recorded in multiple environmental fragments (the redundancy constraint). This redundancy is achieved through physical processes that spread information outward—precisely the processes that define thermodynamic irreversibility.

The key insight is that the redundancy constraint is inherently asymmetric:

- **Forward direction**: Information spreads from system to environment, creating multiple records. This satisfies the redundancy constraint.
- **Backward direction**: Contracting distributed information back into a localized system would require precise coordination of many degrees of freedom—a process of measure zero in the space of dynamical trajectories.

This asymmetry is not imposed by hand; it follows from the structure of accessibility itself.

6.1.3 Scope and Limitations

We state explicitly what this paper does and does not attempt.

This paper does:

- Propose a structural account of temporal asymmetry based on accessibility propagation
- Derive temporal ordering from redundancy structure without assuming fundamental time
- Connect this account to the thermodynamic and quantum arrows
- Reframe causation as constraint propagation along the accessibility direction
- Provide a minimal mathematical model (Appendix A) demonstrating irreversible redundancy expansion
- Situate the analysis within the broader HAFF framework

This paper does not:

- Derive the second law of thermodynamics from first principles
- Explain why initial conditions are low-entropy
- Resolve metaphysical debates about the nature of time (A-theory vs. B-theory, presentism vs. eternalism)
- Address free will, agency, or the phenomenology of temporal experience
- Propose new dynamical equations or empirical predictions

The analysis is structural. We examine how temporal asymmetry relates to accessibility structure, without claiming that this analysis exhausts the content of the problem.

Remark 6.1 (Relation to Papers D and E)**.** *Paper D [137] argued that gravity reflects the evolution of accessible algebras. Paper E [138] argued that measurement reflects the selection of accessible algebras. The present paper argues that time reflects the* directionality *of accessibility propagation. Together, these three papers characterize the diagnostic layer of the HAFF framework:*

- *D: Geometry (algebra evolution)*
- *E: Measurement (algebra selection)*
- *F: Time (algebra propagation direction)*

6.1.4 Outline

Section 6.2 reviews the status of time in various physical theories. Section 6.3 develops the core technical content: how the accessibility constraints generate directional structure. Section 6.4 reframes causation as constraint propagation along the accessibility direction. Section 6.5 compares the present approach to existing accounts of temporal asymmetry. Section 6.6 states explicit non-commitments. Section 6.7 discusses connections to the HAFF framework. Section 6.8 concludes. Appendix 6.9 provides a minimal mathematical model demonstrating the irreversibility of redundancy expansion.

6.2 Background: Time in Physics

Before developing the accessibility-based account, we briefly review the status of time in major physical theories. The purpose is to identify a common structural assumption: that time is an external parameter, given rather than derived.

6.2.1 Time in Classical and Quantum Mechanics

In Newtonian mechanics, time is an absolute parameter. The equations of motion are time-reversal invariant: if $\mathbf{x}(t)$ is a solution, so is $\mathbf{x}(-t)$ (with velocities reversed). There is no intrinsic arrow.

In quantum mechanics, time evolution is governed by the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}|\psi\rangle = \hat{H}|\psi\rangle. \tag{6.1}$$

This equation is unitary and reversible. The apparent irreversibility of measurement is an interpretational issue, not a feature of the formalism itself.

6.2.2 Time in Quantum Gravity

In canonical approaches to quantum gravity, the Wheeler-DeWitt equation takes the form:

$$\hat{H}|\Psi\rangle = 0, \tag{6.2}$$

where $|\Psi\rangle$ is the wave function of the universe. This equation contains no time parameter; the universe is described by a static state satisfying a constraint equation [28].

The Page-Wootters mechanism [72] recovers effective time evolution from correlations between a "clock" subsystem and the rest of the universe within a timeless universal state. However, this mechanism explains how time *ordering* emerges from correlations but does not explain why this ordering is *asymmetric*.

6.2.3 The Common Thread

Across these theories, time appears either as an external parameter or as an emergent concept requiring additional input. The present framework offers a third perspective: time as a **structural consequence** of accessibility propagation, with directionality arising from the statistical asymmetry of redundancy expansion.

6.3 Accessibility and Directionality

We now develop the central technical content: how the accessibility constraints generate a preferred direction that can be identified with temporal ordering.

6.3.1 Recap: The Redundancy Constraint

Paper E [138] established that an observable $\hat{O}$ belongs to the accessible algebra $\mathcal{A}_\mathbf{c}$ only if it satisfies three constraints, including the *redundancy constraint*:

$$I(\hat{O} : E_k) \approx H(\hat{O}) \quad \text{for many } k, \tag{6.3}$$

where $I(\cdot : \cdot)$ denotes quantum mutual information, $H(\cdot)$ denotes von Neumann entropy, and $\{E_k\}$ are independent environmental fragments.

This constraint ensures that information about accessible observables is distributed across multiple environmental subsystems, enabling intersubjective objectivity.

6.3.2 The Redundancy Index

We introduce a quantitative measure of redundancy:

Definition 6.2 (Redundancy Index)**.** *For an observable $\hat{O}$ and environment $E = \bigotimes_k E_k$ consisting of N fragments, the **redundancy index** $\mathcal{R}(\hat{O})$ is the number of environmental fragments that have acquired nearly complete information about $\hat{O}$:*

$$\mathcal{R}(\hat{O}) = \sum_{k=1}^{N} \Theta\left(I(\hat{O} : E_k) - (1 - \delta)H(\hat{O})\right), \tag{6.4}$$

where Θ is the Heaviside step function and $\delta \ll 1$ is the information loss tolerance.

High redundancy ($\mathcal{R} \sim N$) corresponds to classical, objective observables. Low redundancy ($\mathcal{R} \sim 1$) corresponds to quantum, contextual observables.

6.3.3 Asymmetry of Redundancy Flow

The central observation is that redundancy expansion and contraction are radically asymmetric:

Proposition 6.3 (Asymmetry of Redundancy Flow)**.** *Let $\hat{O}$ be an observable of a central system S interacting with an N-fragment environment E. Then:*

1. ***Expansion is generic**: Under typical interactions, $\mathcal{R}(\hat{O})$ increases from $\mathcal{R} = 0$ toward $\mathcal{R} \sim N$.*
2. ***Contraction is non-generic**: The phase space volume of trajectories along which $\mathcal{R}$ decreases is exponentially suppressed:*

$$\frac{Vol(\mathcal{R}\downarrow)}{Vol(\mathcal{R}\uparrow)} \sim e^{-\alpha N} \tag{6.5}$$

for some $\alpha > 0$ depending on the fragment dimensions.

The proof is provided in Appendix 6.9. The key insight is that expansion requires only generic spreading of correlations, while contraction requires exponentially precise conspiracy among N independent fragments.

We note that this argument is a specific instance of the general statistical-mechanical arrow of time (Boltzmann's combinatorial argument), applied to the redundancy index rather than entropy. The derivation presupposes a dynamical evolution parameter; what is derived is the *arrow* (directionality), not time itself.

6.3.4 Temporal Direction from Redundancy Gradient

This asymmetry induces a natural ordering on configurations of the accessible algebra:

Definition 6.4 (Accessibility Ordering). *Let $\mathcal{A}_\alpha$ and $\mathcal{A}_\beta$ be two configurations of the accessible algebra (corresponding to different redundancy structures). We define the partial order:*

$$\mathcal{A}_\alpha \prec \mathcal{A}_\beta \quad \Leftrightarrow \quad \mathcal{A}_\alpha \subset \mathcal{A}_\beta \text{ and } \mathcal{R}(\mathcal{A}_\beta) \geq \mathcal{R}(\mathcal{A}_\alpha). \tag{6.6}$$

This partial order is not imposed externally but emerges from the statistical structure of redundancy propagation. It constitutes the structural origin of temporal direction.

This definition assumes that algebraic inclusion and redundancy increase are consistent—i.e., that enlarging the algebra does not decrease the redundancy of the original operators. This monotonicity property holds for quantum Darwinism in the Markov limit but has not been established in general.

Clarification: Arrow Without Fundamental Time

No fundamental time parameter is assumed. What might conventionally be written as $\mathcal{A}(t_1)$ and $\mathcal{A}(t_2)$ with $t_1 < t_2$ is here understood as $\mathcal{A}_\alpha \prec \mathcal{A}_\beta$—a partial order on algebraic configurations induced by redundancy monotonicity.
"Dynamical trajectories" are not functions $\hat{O}(t)$ parametrized by external time, but **directed paths through the space of accessible algebras** $\{\mathcal{A}_\alpha\}$, with direction determined by the redundancy gradient:

$$\mathcal{A}_\alpha \subset \mathcal{A}_\beta \quad \text{with} \quad \mathcal{R}(\mathcal{A}_\beta) \geq \mathcal{R}(\mathcal{A}_\alpha). \tag{6.7}$$

Time is not a parameter but the **inclusion order of accessible structures**.

6.3.5 The Statistical Nature of the Arrow

The arrow of time, in this framework, is neither:

- A fundamental law (time-reversal symmetry is not violated)
- A thermodynamic accident (entropy is not the primary concept)
- A cosmological boundary condition (no special initial state is assumed)

Rather, it is a *statistical gradient*: the overwhelming majority of accessible-algebra configurations lie in the direction of increasing redundancy. Trajectories toward decreasing redundancy exist in principle but occupy exponentially vanishing phase space volume.

Time is the statistical gradient of redundancy.

6.3.6 Relation to Thermodynamic Arrow

The accessibility arrow and the thermodynamic arrow are closely related but not identical:

The accessibility arrow may be viewed as a *generalization* of the thermodynamic arrow: it applies whenever accessibility constraints are satisfied, even in contexts where thermodynamic entropy is not well-defined (e.g., quantum gravitational regimes where spacetime is emergent).

Feature	Thermodynamic Arrow	Accessibility Arrow
Defined by	Entropy increase	Redundancy expansion
Requires	Coarse-graining choice	Accessibility constraints
Fundamental quantity	$S = -k_B \mathrm{Tr}(\rho \ln \rho)$	$\mathcal{R}[\mathcal{A}]$ (redundancy index)
Applies to	Macroscopic systems	Any system with environment

Table 6.1: Comparison of thermodynamic and accessibility arrows.

6.4 Causation as Constraint Propagation

Having established that accessibility propagation defines a preferred direction, we now reframe causation in these terms.

6.4.1 Causation Without Fundamental Time

Traditional accounts of causation presuppose temporal ordering: causes precede effects. But if temporal ordering itself emerges from accessibility structure, causation must be reframed accordingly.

Definition 6.5 (Causal Relation)**.** *An observable $\hat{A}$ is* ***causally prior*** *to observable $\hat{B}$ (written $\hat{A} \rightsquigarrow \hat{B}$) if:*

1. *$\hat{A}$ and $\hat{B}$ are both accessible: $\hat{A}, \hat{B} \in \mathcal{A}_{\mathbf{c}}$*
2. *The redundancy of $\hat{A}$ is established before the redundancy of $\hat{B}$: $\mathcal{R}(\hat{A})$ saturates at algebraic configuration $\mathcal{A}_\alpha$ while $\mathcal{R}(\hat{B})$ saturates at $\mathcal{A}_\beta$ with $\mathcal{A}_\alpha \prec \mathcal{A}_\beta$*
3. *Counterfactual dependence holds: perturbations of $\hat{A}$ induce correlated perturbations of $\hat{B}$*

This definition grounds causation in the propagation of accessibility constraints through environmental redundancy.

6.4.2 Why Retrocausation is Non-Generic

A persistent question in philosophy of physics is whether retrocausation—effects preceding causes—is possible. The present framework provides a structural answer:

Proposition 6.6 (Suppression of Retrocausation)**.** *Retrocausal trajectories are not excluded by principle, but are exponentially suppressed and macroscopically atypical.*

Proof sketch. For $\hat{B}$ to causally influence $\hat{A}$ when $\mathcal{R}(\hat{A})$ is already saturated (information about $\hat{A}$ distributed across N environmental fragments), the influence would need to:

1. Propagate through all N fragments simultaneously
2. Reconverge the distributed information coherently
3. Do so without disturbing the existing redundancy structure

The phase space volume for such trajectories scales as $e^{-\alpha N}$ (Appendix 6.9), rendering them statistically negligible for macroscopic N. □

Remark 6.7. *This result does not "forbid" retrocausation by fiat. Rather, it explains why retrocausal scenarios—while not logically impossible—do not occur: they require exponentially fine-tuned conspiracies in Hilbert space that generically do not obtain. This is stronger than any "causal postulate" because it derives from the geometry of state space, not from an imposed principle.*

6.4.3 Causal Structure Without Spacetime

The causal relation $\rightsquigarrow$ defines a partial order on accessible observables with the following properties:

- **Irreflexive**: $\hat{A} \not\rightsquigarrow \hat{A}$
- **Asymmetric**: $\hat{A} \rightsquigarrow \hat{B}$ implies $\hat{B} \not\rightsquigarrow \hat{A}$ (by Proposition 6.6)
- **Transitive**: $\hat{A} \rightsquigarrow \hat{B}$ and $\hat{B} \rightsquigarrow \hat{C}$ implies $\hat{A} \rightsquigarrow \hat{C}$

These properties are characteristic of causal structure and emerge here without presupposing a background temporal manifold.

6.5 Relation to Existing Approaches

We situate the accessibility-based account relative to existing approaches to temporal asymmetry.

6.5.1 Comparison Table

Approach	**Source of Arrow**	**What It Presupposes**	**Relation to HAFF**
Thermodynamic	Entropy increase	Coarse-graining choice	Accessibility more fundamental
Cosmological	Low-entropy Big Bang	Boundary conditions	Explains initial conditions
Decoherence	Interference suppression	System-environment split	Special case of accessibility
Causal set	Fundamental partial order	Causal order as primitive	HAFF derives the order
Page-Wootters	Correlations in static $\lvert\Psi\rangle$	Timeless formulation	HAFF adds directionality
Accessibility	**Redundancy expansion**	**Interaction structure**	—

Table 6.2: Comparison of approaches to temporal asymmetry.

6.5.2 Key Distinctions

Thermodynamic arrow: The accessibility arrow is closely related but more fundamental. Entropy increase presupposes a coarse-graining; accessibility expansion explains *why* certain coarse-grainings are physically relevant.

Page-Wootters: Both approaches treat time as emergent. Page-Wootters explains how time *appears*; the accessibility framework explains why it has a *direction*.

Retrocausality programs: Some approaches explore retrocausal models [78]. The present framework does not exclude retrocausation in principle but explains its non-occurrence: retrocausal trajectories occupy exponentially vanishing phase space volume.

6.6 What This Paper Does NOT Claim

To prevent misreading, we state explicitly what this paper does *not* claim.

1. **No claim that time is unreal or illusory.** The framework reframes temporal asymmetry as emergent from accessibility structure, but this does not imply that time is "merely subjective" or non-existent.

2. **No adjudication between A-theory and B-theory of time.** The framework is compatible with both presentism and eternalism.

3. **No explanation of initial conditions.** We do not explain why the universe began with low redundancy, only why redundancy generically increases thereafter.

4. **No resolution of the problem of time in quantum gravity.** The framework clarifies what temporal asymmetry *means* in accessibility terms but does not derive time from the Wheeler-DeWitt equation.

5. **No claims about consciousness or subjective time.** The phenomenology of temporal experience is not addressed.

6. **No novel empirical predictions.** The analysis is structural, not dynamical.

7. **No claim that retrocausation is impossible.** Retrocausation is statistically suppressed (measure zero), not logically forbidden.

8. **No modification of quantum mechanics.** The framework assumes standard unitary evolution throughout.

6.7 Connection to HAFF Framework

This paper completes the diagnostic layer of the HAFF framework.

6.7.1 The D + E + F Diagnostic Triangle

Papers D, E, and F form a coherent triad, each addressing a different aspect of how structure emerges from accessible algebras:

The unifying theme is that features traditionally taken as fundamental—force, measurement, time—may be understood as emergent properties of accessibility structure.

Paper	Phenomenon	Traditional View	HAFF Reframing
D	Gravity	Force between masses	Evolution of accessible algebra
E	Measurement	Primitive process	Selection within accessible algebra
F	Time	Fundamental parameter	Direction of accessibility propagation

Table 6.3: The D + E + F diagnostic triangle.

6.7.2 Structural Link: D + E + F

The three papers form a symmetric closed structure:

- **D (Gravity)**: The evolution $\mathcal{A}_\mathbf{c}(t)$ of the accessible algebra manifests as curved geometry.
- **E (Measurement)**: The selection of $\mathcal{A}_\mathbf{c}$ via physical constraints manifests as objective outcomes.
- **F (Time)**: The direction of redundancy expansion within $\mathcal{A}_\mathbf{c}$ manifests as the causal arrow.

6.7.3 Boundary Note: Toward Layer III

The completion of Layer II (diagnostic unification) sets the stage for Layer III: the structural limits of the framework itself.

A key insight from D + E + F is that what appears fundamental (force, measurement, time) is actually emergent from accessible algebras. But this raises a question: *What determines the accessible algebra structure itself?*

Layer III (Paper G) will argue that this question admits no complete answer within the framework—not because the framework is incomplete, but because any answer would require a "meta-framework" to justify, leading to infinite regress. The boundary is structural, not epistemic.

A theory that claims to explain everything must know where it must stop.

6.8 Conclusion

6.8.1 Summary of Results

We have proposed a structural account of temporal asymmetry and causation based on accessibility propagation. The central results are:

1. **Time without fundamental parameter**: All temporal ordering emerges from a partial order induced by algebraic inclusion and redundancy monotonicity. No external time parameter is assumed.
2. **The accessibility arrow**: Redundancy expansion is generic; redundancy contraction is exponentially suppressed. This asymmetry defines a preferred direction.

3. **Causation as constraint propagation**: Causal relations emerge from constraint propagation along the accessibility arrow. Causes are sources of redundancy expansion; effects are regions of redundant recording.

4. **Retrocausation non-generic**: Retrocausal trajectories are not excluded by principle, but are exponentially suppressed and macroscopically atypical.

5. **Diagnostic layer complete**: With Papers D, E, and F, the HAFF framework provides unified structural accounts of gravity, measurement, and time.

6.8.2 Closing Remark

The arrow of time has puzzled physicists and philosophers for over a century. We do not claim to have dissolved this puzzle. What we have done is reframe it:

> The question is not "Why does entropy increase?" but "Why does accessibility expand?"

The answer—that expansion is generic while contraction requires exponential fine-tuning—follows from the geometry of Hilbert space in interacting systems. This does not explain everything. But by identifying the structural basis of temporal asymmetry, we clarify what remains to be explained—and what, perhaps, lies beyond the reach of structural analysis altogether.

6.9 A Minimal Model of Irreversible Redundancy Expansion

To rigorously demonstrate the central thesis of Section 6.3—that accessibility expansion is generic while contraction is statistically suppressed—we consider a finite-dimensional model of a central system interacting with a fragmented environment.

6.9.1 The Star-Graph Interaction Setup

Consider a central system S (the "source" of accessibility) and an environment E consisting of N independent subsystems (fragments) $E_1, E_2, \ldots, E_N$. The total Hilbert space is:

$$\mathcal{H}_{\text{tot}} = \mathcal{H}_S \otimes \mathcal{H}_{E_1} \otimes \mathcal{H}_{E_2} \otimes \cdots \otimes \mathcal{H}_{E_N}. \tag{6.8}$$

The interaction Hamiltonian generating accessibility is chosen to be of the "pre-measurement" type:

$$\hat{H}_{\text{int}} = g \sum_{k=1}^{N} \hat{O}_S \otimes \hat{M}_k, \tag{6.9}$$

where $\hat{O}_S$ is the observable of S becoming accessible, $\hat{M}_k$ are the monitoring operators of the environmental fragments, and g is the coupling strength.

6.9.2 Dynamics of Redundancy

Assume the initial state is uncorrelated:

$$|\Psi(0)\rangle = |s\rangle_S \otimes |e_0\rangle_{E_1} \otimes |e_0\rangle_{E_2} \otimes \cdots \otimes |e_0\rangle_{E_N}. \tag{6.10}$$

Under the unitary evolution $U(t) = e^{-i\hat{H}_{\text{int}}t/\hbar}$, the state evolves into an entangled superposition. For $\hat{O}_S$ with eigenstates $|o_i\rangle$:

$$|\Psi(t)\rangle = \sum_i c_i |o_i\rangle_S \otimes |E_i^{(1)}(t)\rangle \otimes |E_i^{(2)}(t)\rangle \otimes \cdots \otimes |E_i^{(N)}(t)\rangle, \tag{6.11}$$

where $|E_i^{(k)}(t)\rangle$ are the relative states of the environmental fragments.

The mutual information $I(\hat{O}_S : E_k)$ grows as the fragments become correlated with S. By standard decoherence results [121], for small t:

$$I(\hat{O}_S : E_k) \sim (gt)^2. \tag{6.12}$$

6.9.3 Proof of Asymmetry

Forward Evolution (Generic Expansion):

As t increases, information spreads to more fragments. For $gt \gg 1$, the redundancy index approaches its maximum:

$$\mathcal{R}(\hat{O}_S) \to N. \tag{6.13}$$

This state corresponds to the "classical plateau" where the algebra generated by $\hat{O}_S$ is maximally accessible.

Backward Evolution (Contraction Suppression):

Consider the time-reversed evolution from a state of high redundancy. For $\mathcal{R}$ to decrease, the N environmental fragments must conspiratorially un-correlate with S simultaneously.

In the phase space of the total system $\mathcal{H}_{\text{tot}}$, let V_{low} be the volume of states with low redundancy ($\mathcal{R} < \mathcal{R}_{\text{crit}}$) and V_{high} be the volume of states with high redundancy ($\mathcal{R} \sim N$).

By counting Hilbert space dimensions, the ratio is exponentially suppressed:

$$\frac{V_{\text{low}}}{V_{\text{high}}} \sim e^{-\alpha N}, \tag{6.14}$$

where $\alpha > 0$ depends on the dimension of the fragments.

6.9.4 Conclusion of Appendix

While the dynamical laws ($U(t) = e^{-i\hat{H}t/\hbar}$) are reversible, the **Accessibility Flow** is structurally irreversible:

- A trajectory starting in V_{low} generically moves to V_{high} (time arrow $\to$).
- A trajectory starting in V_{high} will almost never spontaneously fluctuate back to V_{low} within the recurrence time of the universe.

The irreversibility comes from **state space volume**, not from dynamical asymmetry. This is the structural basis of the accessibility arrow:

$$\boxed{\textbf{Time is the statistical gradient of Redundancy.}}$$

Chapter 7

Structural Limits of Unification

Paper G

Originally published: Zenodo, DOI: 10.5281/zenodo.18402907

Abstract

This paper examines the structural conditions under which a unificatory physical framework must terminate its explanatory extension. Building on recent work demonstrating that gravitational phenomena, measurement outcomes, and temporal asymmetries can be jointly reframed as consequences of accessible observable algebra selection, we argue that such frameworks cannot be simultaneously complete and self-grounding. The incompleteness identified here is neither formal (in the Gödel–Turing sense) nor epistemic, but architectural: it arises from the non-self-grounding character of accessibility-based physical description. We establish a structural lemma showing that any attempt to internalize accessibility conditions within the framework they enable leads to either infinite regress or explanatory collapse. The stopping point identified is therefore not discretionary but forced by the framework's own explanatory architecture. This analysis does not claim generality beyond the specific formalism developed; whether alternative approaches would encounter analogous limits remains an open question. The contribution is methodological: to articulate the conditions under which recognizing structural boundaries becomes a requirement of explanatory coherence rather than an admission of incompleteness.

7.1 Introduction

7.1.1 The Expectation of Completeness

The aspiration toward a unified physical description has historically been guided by the expectation that deeper unification corresponds to increased completeness. In this traditional view, apparent multiplicity—of forces, degrees of freedom, or explanatory principles—is taken to signal provisional fragmentation, to be resolved by a more fundamental theory. A Theory of Everything, in its strongest formulation, is therefore often assumed to be both unifying and self-grounding: it should not only subsume all known interactions under a single framework, but also account for the conditions under which its own descriptions are possible.

The present work does not adopt this expectation. Instead, it advances a more restricted claim: that unification may be achievable only up to a structurally imposed boundary, beyond which further explanatory extension would undermine the coherence of the framework itself. This claim does not arise from epistemic modesty, nor from skepticism regarding the scope of physical explanation, but from the internal architecture of the formalism developed in the preceding papers of this series [134, 135, 136, 137, 138, 139], which builds on foundational observations regarding the non-uniqueness of tensor factorizations [119, 120].

7.1.2 Summary of the Preceding Framework

Across Papers D–F, gravitational phenomena, measurement outcomes, and temporal asymmetries are jointly reframed as consequences of structural selection: specifically, the selection of accessible observable algebras and associated coarse-grainings. No new fundamental entities are postulated, and no modification of underlying dynamics is proposed. Rather, phenomena traditionally treated as primitive are shown to emerge from constraints on how physical descriptions are stably instantiated.

- **Paper D**: Gravitational dynamics corresponds to the adiabatic flow of the accessible algebra $\mathcal{A}_{\mathbf{c}}(t)$, not to a force operating within a fixed algebra.
- **Paper E**: Measurement is not a primitive process but a manifestation of accessibility constraints on operator algebras—specifically, the selection of observables satisfying interaction, stability, and redundancy criteria.
- **Paper F**: Temporal directionality is identified with the direction of irreversible accessibility propagation, grounded in the asymmetric expansion of redundant environmental records.

These results share a common structure: each phenomenon is traced to accessibility conditions rather than to fundamental ontology. This constitutes a genuine unification at the level of explanatory architecture.

7.1.3 The Problem of Self-Grounding

However, this reframing has a nontrivial implication. If the explanatory power of the framework depends essentially on restrictions—on what is accessible, stable, and non-scrambling—then unification cannot consist in the removal of all such restrictions. To do so would be to erase the very conditions that render physical description meaningful. Unification, in this sense, cannot be both total and self-enclosed.

The unifying move, therefore, is not a convergence toward an all-encompassing description, but a clarification of how far structural explanation can be coherently extended before it becomes reflexive. The aim of this final layer is to articulate that stopping point and to demonstrate that it is structurally forced rather than pragmatically chosen.

7.1.4 Scope and Limitations

Several clarifications are necessary at the outset.

First, the incompleteness identified in this paper is not formal in the sense of Gödel's incompleteness theorems. We do not claim that the framework contains undecidable propositions within a formal system, nor do we invoke metamathematical results. The incompleteness is *architectural*: it concerns the explanatory roles within a physical framework, not provability within a formal calculus.

Second, we do not claim that the structural limits identified here apply universally to all conceivable approaches to unification. The present analysis is confined to the algebraic and coarse-graining-based framework developed in Papers A–F. Whether alternative formalisms—category-theoretic, non-algebraic, or radically background-free—would exhibit analogous limits is an open question that we do not address.

Third, the stopping point identified is not temporal, existential, or normative. It does not mark the end of physics, nor a claim about the limits of human knowledge. It marks the point at which the framework's internal explanatory resources are exhausted without circularity.

7.1.5 Structure of the Paper

Section 7.2 develops the notion of accessibility as a non-global structural constraint and clarifies its distinction from epistemic limitations. Section 7.3 establishes a structural lemma demonstrating that accessibility-based descriptions cannot be self-grounding without collapse or regress. Section 7.4 presents a concrete collapse scenario illustrating what would occur if the framework were extended beyond its structural boundary. Section 7.5 characterizes the final cut as a forced stopping point rather than a discretionary choice. Section 7.6 states explicit non-claims to prevent misinterpretation. Section 7.7 concludes with reflections on the methodological significance of the analysis.

7.2 Accessibility as a Non-Global Constraint

7.2.1 The Role of Accessibility in HAFF

Central to the HAFF framework is the notion of accessibility. Physical descriptions are not formulated over the full algebra of global observables, but over restricted subalgebras determined by interaction structure, dynamical stability, and environmental redundancy. These restrictions are not introduced as pragmatic simplifications, nor as reflections of limited knowledge. They are constitutive of what counts as a well-defined physical description in the first place.

In Papers D–F, this point is developed across distinct domains:

- Effective geometry is shown to depend on coarse-grainings that preserve entanglement structure over relevant timescales.
- Measurement outcomes are shown to arise from dynamically stable partitions that resist rapid scrambling.
- Temporal directionality is associated with asymmetric information flow under constrained interactions.

In each case, the phenomenon under consideration becomes intelligible only relative to a selected accessible algebra.

7.2.2 Structural vs. Epistemic Constraints

A crucial distinction must be drawn between epistemic and structural constraints. This distinction is contested in philosophy of physics, and we acknowledge that what follows adopts it as a working criterion rather than a demonstrated result.

Definition 7.1 (Epistemic Constraint). *A constraint is **epistemic** if it concerns what can be known, inferred, or verified by agents, given their informational position.*

Definition 7.2 (Structural Constraint). *A constraint is **structural** if it concerns what descriptions are well-defined, given a pattern of physical interactions, independent of any agent's knowledge or epistemic state.*

The HAFF framework relies exclusively on the latter notion. Accessibility is determined by stability criteria—dynamical invariance, environmental redundancy (quantum Darwinism), and non-scrambling behavior—that are properties of the Hamiltonian and the global quantum state, not of observers.

Crucially, the selection of an accessible algebra is not arbitrary. Given a fixed interaction structure, different agents—or no agents at all—will identify the same accessible observables. This is the sense in which accessibility is structural rather than epistemic: it is interaction-determined, not belief-determined.

Remark 7.3 (Contested Distinction). *We acknowledge that this distinction between epistemic and structural constraints is philosophically contested. The framework does not claim to have resolved this broader debate. However, the burden of argument lies with the critic to demonstrate that the stability criteria invoked in Papers A–F reduce to epistemic conditions, rather than with the framework to prove a negative. The working distinction is adopted on the grounds that interaction-determined constraints are conceptually prior to agent-relative knowledge.*

7.2.3 Non-Globality of Accessibility

Accessibility is not a global property of the underlying theory. There is no privileged, all-encompassing accessible algebra from which all others can be derived. Each effective description presupposes its own restrictions, and those restrictions cannot be fully specified from within the description they enable.

This asymmetry is decisive. Any attempt to internalize the conditions of accessibility would require a further level of description, governed by its own accessibility conditions. The implications of this observation are developed in the following section.

7.3 A Structural Lemma on Self-Grounding

7.3.1 Statement of the Lemma

We now state the central structural result of this paper.

Lemma 7.4 (Structural Non-Self-Grounding). *Within the HAFF framework, no description can simultaneously:*

(i) specify the structure of accessibility, and

(ii) be formulated entirely within that same accessibility structure,

without collapse into circularity or triviality.

7.3.2 Argument

The argument proceeds in five steps.

Step 1: All physical descriptions in HAFF are formulated relative to an accessible algebra. This is not an optional modeling choice but the basic condition under which any observable, geometry, or temporal ordering becomes definable. Papers D–F establish that gravitational dynamics, measurement outcomes, and causal direction all presuppose restriction to a stable accessible subalgebra $\mathcal{A}_{\mathbf{c}} \subset \mathcal{B}(\mathcal{H}_U)$.

Step 2: Accessibility itself is defined by selection criteria. Stability, redundancy, and interaction locality determine which subalgebras are accessible. These criteria are conditions of possibility for description, not objects described within the description.

Step 3: Attempting to internalize accessibility requires re-applying accessibility criteria to themselves. That is, one would need an accessible algebra that describes the selection of the accessible algebra itself. The framework would have to render the conditions of its own applicability as objects within its descriptive scope.

Step 4: This generates a fixed-point requirement. The framework would have to identify an algebra $\mathcal{A}^*$ that:

- is accessible because it satisfies the stability criteria, and
- simultaneously encodes the criteria by which it is judged accessible.

Symbolically, one would require:

$$\mathcal{A}^* \in \mathrm{Acc}(\mathcal{A}^*), \tag{7.1}$$

where $\mathrm{Acc}(\cdot)$ denotes the set of algebras satisfying the accessibility criteria defined within the argument algebra.

Step 5: Such a fixed point is generically unavailable. Except in degenerate cases—trivial algebras (containing only the identity) or total algebras (the full $\mathcal{B}(\mathcal{H}_U)$ that erases all structure)—the selection criteria cannot be satisfied by their own output. A non-trivial accessible algebra defines distinctions (between accessible and inaccessible, stable and scrambled, redundant and local); encoding the criteria for those distinctions within the algebra would require the algebra to contain its own meta-description, which exceeds the information available at the object level.

7.3.3 Conclusion of the Lemma

Therefore, the framework cannot close on itself without either:

- collapsing into triviality (everything accessible, nothing distinguished), or
- introducing an external meta-structure (violating internal coherence).

The stopping point is not pragmatic. It is forced by the non-self-grounding character of accessibility-based description.

Remark 7.5 (Framework-Relative Claim)**.** *This is a necessity claim internal to the framework's architecture, not a universal limitation on explanation. We do not claim that all physical theories must exhibit this structure, only that the HAFF framework, as developed, does.*

7.4 A Collapse Scenario

To make the structural lemma concrete, we now present a hypothetical extension of HAFF that attempts to fully internalize accessibility as an object-level dynamical variable, and show that this attempt fails.

7.4.1 Hypothetical Extension

Step 1: Treat accessibility as a physical observable. Suppose one introduces an operator or state variable $\hat{A}$ encoding "degree of accessibility" for subalgebras. This variable would quantify, for each subalgebra $\mathcal{A} \subset \mathcal{B}(\mathcal{H}_U)$, the extent to which it satisfies the stability criteria.

Step 2: Demand dynamical laws for accessibility. To be explanatory, $\hat{A}$ must:

- evolve under some dynamics, and
- be measurable within the theory.

Step 3: Apply accessibility criteria to $\hat{A}$. But measurability requires that $\hat{A}$ itself satisfy:

- stability under interaction,
- redundancy across environments, and
- non-scrambling behavior.

That is, $\hat{A}$ must belong to some accessible algebra $\mathcal{A}_{\hat{A}}$.

7.4.2 Two Fatal Outcomes

Outcome (a): Infinite regress. The algebra $\mathcal{A}_{\hat{A}}$ that makes $\hat{A}$ accessible is itself defined by accessibility criteria. To explain why $\mathcal{A}_{\hat{A}}$ is accessible, one would need a further algebra $\mathcal{A}_{\mathcal{A}_{\hat{A}}}$, and so on. Each level of accessibility-description requires a higher-level accessibility structure to define its observables. The regress does not terminate.

Outcome (b): Totalization collapse. To avoid regress, one might declare everything accessible—that is, take $\mathcal{A}_{\mathbf{c}} = \mathcal{B}(\mathcal{H}_U)$. But then:

- no algebra selection remains,
- no measurement distinction exists (all observables are equally accessible),
- effective geometry loses definition (no coarse-graining induces structure), and

- temporal direction vanishes (no asymmetric accessibility propagation).

The framework either never terminates or destroys the very distinctions it set out to explain.

7.4.3 Conclusion of the Scenario

Any attempt to go "beyond" Papers D–F by internalizing accessibility eliminates the explanatory power already achieved. This is not philosophical caution. It is structural self-destruction.

The collapse scenario demonstrates concretely what the structural lemma establishes abstractly: the framework cannot extend itself to explain its own conditions of applicability without losing the capacity to explain anything at all.

7.5 The Necessity of a Final Cut

7.5.1 Stopping as Structural Necessity

The preceding analyses motivate a specific sense in which the HAFF framework is incomplete. This incompleteness is neither formal nor metaphysical. It does not arise from undecidable propositions, nor from claims about the limits of human cognition. Rather, it is structural: a consequence of the fact that explanatory resources cannot simultaneously function as both explanans and explanandum.

Within the framework developed here, gravity, measurement, and time are unified at the level of structural selection. They are shown to depend on how observable algebras are restricted and stabilized. This constitutes a genuine unification, insofar as disparate phenomena are traced to a common architectural feature. Yet the framework does not, and cannot, provide a further account of why those accessibility conditions obtain, without appealing to structures that would themselves require explanation under the same terms.

7.5.2 The Final Cut

The notion of a "final cut" is introduced to mark this boundary. It does not denote a temporal endpoint, nor a claim about the completion of physics. It denotes the point at which the internal explanatory strategy of the framework reaches saturation. Beyond this point, further elaboration would no longer clarify structure, but obscure it by erasing the asymmetries that make explanation possible.

The introduction of a final cut is not a discretionary methodological choice, nor a gesture of philosophical modesty. It is the point at which the framework exhausts its own internal resources without contradiction.

Beyond this point, any further extension would require the framework to explain the conditions of its own applicability using those very conditions—a requirement that admits no non-degenerate solution (Lemma 7.4).

The stopping point is therefore not selected but encountered. It is the boundary at which explanation ceases to be generative and becomes self-consuming.

7.5.3 Incompleteness as Internal Limit

In this sense, the incompleteness identified here is not provisional, nor external, nor epistemic. It is structural and internal: a limit imposed by the framework's success in making accessibility do explanatory work.

The framework terminates not in absence, but at the level of unselected structure—a domain that admits no geometry, no temporal ordering, and no observational standpoint, yet functions as the necessary substrate from which all three are selectively realized.

A theory may approach totality only by knowing where it must stop—relative to its own structure.

Remark 7.6 (Non-Arbitrary Stopping). *The stopping point is non-arbitrary in the following precise sense: any proposed extension beyond this point can be shown to lead either to regress (Section 7.4, Outcome a) or to collapse (Section 7.4, Outcome b). The burden of argument therefore shifts to those who claim that further extension is possible: they must specify how regress or collapse is avoided.*

7.6 Scope and Non-Claims

To prevent misinterpretation, several non-claims must be stated explicitly.

1. **No invocation of Gödel's theorems.** This work does not invoke Gödel's incompleteness theorems, nor does it rely on any formal analogy to them. The incompleteness identified here is architectural, not metamathematical. It concerns explanatory roles within a physical framework, not provability within a formal system.

2. **No claim of universal applicability.** No claim is made that accessibility constraints apply universally across all conceivable approaches to unification. The present analysis is confined to the algebraic and coarse-graining-based framework developed in Papers A–F.

3. **Contested distinction acknowledged.** The distinction between structural and epistemic constraints is acknowledged to be contested in the philosophy of physics. The framework adopts this distinction as a working criterion, grounded in interaction-determined stability conditions. It does not claim to have resolved the broader philosophical debate.

4. **No foreclosure of alternative frameworks.** The identification of a final cut does not preclude further physical progress, alternative models, or deeper insights within other frameworks. It merely states that, within the present formalism, further internal extension would be incoherent.

5. **No uniqueness claim.** We do not claim that this stopping point is unique. Other frameworks may stop elsewhere, or may not require stopping at all. The claim is only that, given the explanatory architecture adopted here, the stopping point identified is forced.

6. **No philosophical finality.** This is a structural observation within a specific formalism, not a philosophical conclusion about the ultimate nature of reality or the limits of knowledge as such. The stopping point identified is methodological and structural, not existential.

7. **No consciousness or observer-creation claims.** The framework does not claim that consciousness plays a fundamental role, that observers create reality, or that accessibility is observer-relative in any subjective sense.

8. **No claim of novelty regarding theoretical presupposition.** We acknowledge that the observation that explanatory frameworks have presuppositions they cannot fully explain is familiar in philosophy of science. The contribution here is to show that the specific HAFF framework forces this conclusion through its reliance on accessibility, not merely that it is compatible with it.

7.7 Conclusion

7.7.1 Summary of Results

This paper has examined the structural conditions under which the HAFF framework must terminate its explanatory extension.

The central results are:

1. **Accessibility as non-global constraint**: Physical descriptions in HAFF are formulated relative to accessible algebras, which are determined by stability criteria that cannot be fully specified from within the descriptions they enable.

2. **Structural non-self-grounding** (Lemma 7.4): No description within HAFF can simultaneously specify the structure of accessibility and be formulated entirely within that accessibility structure, without collapse or regress.

3. **Collapse scenario**: Any attempt to internalize accessibility as an object-level variable leads to infinite regress or totalization collapse, eliminating the framework's explanatory power.

4. **Forced stopping point**: The final cut is not discretionary but encountered as a structural necessity—the point at which further extension would be self-consuming rather than generative.

7.7.2 Methodological Significance

Unification is often equated with the elimination of boundaries. The analysis presented here suggests a different criterion: that a unifying framework should distinguish between boundaries that are provisional and those that are structural.

Papers D–F identify such structural boundaries in the treatment of gravity, measurement, and time. This final layer marks the point at which acknowledging those boundaries becomes a condition of explanatory clarity rather than an admission of incompleteness.

The value of the framework lies not in its ability to say everything, but in its ability to determine what cannot be said without loss of coherence. At that point, stopping is not a retreat, but a completion.

7.7.3 Open Questions

Several questions remain beyond the scope of this analysis:

- Whether alternative frameworks (category-theoretic, non-algebraic, or background-free) would exhibit analogous structural limits.
- Whether the structural/epistemic distinction adopted here can be given a more robust philosophical foundation.
- Whether the collapse scenario admits any non-trivial avoidance strategies not considered here.

These questions are left for future investigation.

Postscript: On the Closure of Structure

Originally published: Zenodo, DOI: 10.5281/zenodo.18407367

In the Holographic Alaya–Field Framework, the universe has been treated not as a collection of fundamental objects, but as a bounded domain of accessibility within a single global operator structure [134, 135]. Throughout this work, gravity, time, and measurement have appeared only insofar as stable distinctions are sustained by a persistent separation—what we have called the Cut—between accessible subalgebras and the total algebra [138, 132]. The philosophical implications of this structural stance—for causation, agency, and existence—have been explored in the accompanying essay [136].

Pushing this framework to its logical limit raises a natural question: what becomes of the theory when such distinctions can no longer be maintained?

Heat Death as Accessibility Saturation

From a structural perspective, the conventional notion of heat death admits a reinterpretation. Rather than signifying the disappearance of physical existence, it corresponds to the saturation of accessibility. As informational redundancy becomes maximal, differences between subsystems cease to be stably recordable [122]. The gradients that underwrite locality, temporal ordering, and effective classicality flatten into a homogeneous configuration.

In this limit, the Cut loses operational meaning. No stable partition remains that could support observers, records, or localized descriptions. The system approaches the undifferentiated operator structure introduced at the beginning of this work—a state of maximal symmetry and minimal distinguishability.

Structural Equivalence of Origin and Terminus

Crucially, this endpoint is not structurally distinct from the origin. The state of maximal entropy reached at late times is, in algebraic terms, indistinguishable from the maximally symmetric pre-differentiated configuration. The difference between "beginning" and "end" is therefore not ontological, but structural: it reflects whether accessibility constraints are present or dissolved.

Mathematically, let $\mathcal{A}_{\text{total}}$ denote the full operator algebra and let $S[\rho]$ denote the von Neumann entropy of a state ρ. At both temporal extremes:

$$\lim_{t \to 0^+} S[\rho(t)] \approx \lim_{t \to \infty} S[\rho(t)] \approx S_{\max}, \tag{7.2}$$

where the limits are understood in terms of accessible structure rather than absolute time. The initial state (pre-Cut) and the final state (post-dissolution) occupy the same region of algebraic configuration space—both correspond to conditions under which no stable coarse-graining can be sustained.

Bounded Evolution Without Cyclicity

Seen this way, cosmic evolution traces neither a linear narrative nor a teleological arc. It is instead bounded by two structurally equivalent limits: one preceding the emergence of stable distinctions, and one following their dissolution. The domain in which physics, observation, and meaning are possible occupies only the intermediate regime, where the Cut is sustained.

This observation carries no additional dynamical claims, nor does it posit a cosmological cycle in the sense of a Big Bounce or oscillating universe model [133]. It merely completes the logical closure of the framework developed here. The theory describes the conditions under which structure can appear, persist, and ultimately fail. Beyond those conditions, no further physical description is available—not because reality ends, but because the criteria for description are no longer satisfied.

The Contingency of Intelligibility

If the work has a final implication, it is a modest one: intelligibility itself is contingent. The universe is describable only while distinctions endure. Understanding this boundary does not diminish the value of structure; it clarifies the narrow window in which structure—and thus physics—is possible.

The framework terminates not in absence, but at the level of unselected structure: a domain that admits no geometry, no temporal ordering, and no observational standpoint, yet functions as the necessary substrate from which all three are selectively realized.

Remark 7.7 (On Structural Closure)**.** *The identification of origin and terminus as algebraically equivalent does not constitute a prediction about cosmological dynamics. It is a statement about the explanatory boundaries of accessibility-based description. Within those boundaries, the framework provides a unified account of gravity, measurement, and time [137, 138, 139]. Beyond them, no description formulated in terms of accessible algebras can be coherently maintained [132].*

Concluding Reflection

The value of a theoretical framework lies not only in what it explains, but in what it determines cannot be explained without loss of coherence. The stopping point identified in this work is not a failure of explanation but its completion. A theory that claims to explain everything must know where it must stop.

Clarity does not require totality. And knowing where to stop is sometimes the most precise act of understanding.

Part II

Q-RAIF: Quantum Reference Algebra for Information Flow

Chapter 8

Algebraic Constraints on the Emergence of Lorentzian Metrics in Entropic Gravity Frameworks

Paper A — "The Water"

Originally published: Zenodo, DOI: 10.5281/zenodo.18525876

Abstract

We investigate the algebraic conditions under which an emergent bulk geometry acquires a Lorentzian signature within the framework of entropic gravity. While thermodynamic approaches to gravity [52, 105] and the Ryu–Takayanagi formula [83] relate entanglement entropy to geometric data, the specific algebraic mechanism constraining the spacetime signature remains an open question.

We identify three independent constraints on the boundary algebra—associativity, metric compatibility, and causal channel encoding—and argue that their simultaneous satisfaction naturally selects a **Clifford algebra** $Cl(V, q)$ as the minimal compatible structure. We support this claim by systematically examining alternative algebraic frameworks (von Neumann factors, C^*-algebras, Jordan algebras, Lie algebras) and demonstrating that each fails to satisfy at least one constraint. A worked example using a qubit tensor network illustrates how the three constraints operate in a concrete setting.

The analysis complements the Holographic Alaya-Field Framework (HAFF) [127, 128], which establishes that geometry emerges from coarse-graining of observable algebras: the present work characterizes the algebraic constraints that any such emergent geometry must satisfy to be Lorentzian.

Keywords: emergent geometry, Clifford algebra, entropic gravity, holographic principle, signature selection, algebraic quantum gravity

8.1 Introduction

8.1.1 Context and Motivation

The AdS/CFT correspondence [68] and the Ryu–Takayanagi formula [83] have established that spacetime geometry can be viewed as an emergent property of quantum entanglement. Thermodynamic approaches [52, 105] further suggest that the Einstein equations arise as an equation of state. Yet a critical question remains: *What algebraic constraints ensure that the emergent geometry is Lorentzian?*

Most models assume the $(1, 3)$ signature *a priori*. We argue that this assumption can be partially justified by examining the algebraic consistency conditions on the boundary degrees of freedom.

8.1.2 Relation to HAFF

The Holographic Alaya-Field Framework [127, 128] demonstrates that inequivalent coarse-graining structures on a global quantum state induce inequivalent emergent geometries. HAFF establishes *that* geometry emerges from observable algebras; the present work addresses *what algebraic constraints* such emergent geometry must satisfy to be Lorentzian.

HAFF	Geometry is coarse-graining-dependent (the "ocean")
This paper	Lorentzian signature is algebraically constrained (the "water")

8.1.3 Scope and Disclaimers

This work does not propose a new fundamental theory of gravity, nor does it claim to derive $Cl(1, 3)$ from first principles alone. Rather, it identifies a set of physically motivated algebraic constraints and argues that Clifford algebra is the minimal structure satisfying all of them simultaneously. The argument is presented as a *consistency analysis*, not a uniqueness proof.

The specific value $(1, 3)$ for the signature requires additional input beyond the algebraic constraints developed here (e.g., observational dimensionality or anomaly cancellation arguments). We do not address why spacetime has $3 + 1$ dimensions.

8.2 Candidate Algebraic Structures

Before deriving constraints, we survey the landscape of algebraic structures that could, in principle, describe boundary degrees of freedom in a holographic setting. This survey serves as the basis for the exclusion argument in Section 8.3.4.

Definition 8.1 (Boundary Algebra)**.** *Let $\mathcal{A}_\partial$ be the algebra of observables on a holographic boundary. We require: (a) $\mathcal{A}_\partial$ acts faithfully on $\mathcal{H}_\partial$; (b) $\mathcal{A}_\partial$ admits a trace compatible with the holographic entropy bound; (c) coarse-graining of $\mathcal{A}_\partial$ induces an effective bulk description.*

The following algebraic families are candidates:

1. **von Neumann algebras** (Type I, II, III): Associative, closed under adjoint, weakly closed. Standard in algebraic QFT [43]. Type III_1 factors are generic in relativistic QFT.

2. **C^*-algebras**: Associative Banach algebras with involution. More general than von Neumann algebras. Standard framework for quantum observables.

3. **Lie algebras**: Antisymmetric bracket $[A, B] = -[B, A]$, satisfying the Jacobi identity. Encode infinitesimal symmetries. The universal enveloping algebra is associative.

4. **Jordan algebras**: Commutative but generally non-associative: $A \circ B = B \circ A$, satisfying the Jordan identity. Proposed for quantum mechanics by Jordan, von Neumann, and Wigner (1934).

5. **Octonion algebras**: Non-associative division algebra. Explored in the context of exceptional structures in string theory [41].

6. **Clifford algebras** $Cl(V, q)$: Associative, generated by a vector space V with quadratic form q, subject to $v^2 = q(v)\mathbf{1}$. Encode both metric and algebraic structure [50, 29].

8.3 Three Algebraic Constraints

We now derive three constraints from physically motivated requirements and examine which candidate algebras survive.

8.3.1 Constraint I: Associativity

Lemma 8.2 (Associativity Requirement)**.** *If the boundary algebra supports well-defined time evolution (evolution operators forming a semigroup), it must be associative.*

Motivation. The following is not a derivation but a statement of the minimal algebraic requirement, with physical motivation. The semigroup property requires $(U(t_1)U(t_2))U(t_3) = U(t_1)(U(t_2)U(t_3))$ for all $t_i \geq 0$. In a non-associative algebra, different bracketings of n sequential operations produce $C_n \sim 4^n/n^{3/2}$ distinct results (Catalan numbers), generating uncontrolled ambiguity that grows exponentially with the number of time steps.

We note that this constraint is automatically satisfied by operator algebras on Hilbert spaces, where composition of linear maps is inherently associative. The constraint therefore functions as a *structural boundary condition*: it delineates the algebraic regime in which consistent dynamics is possible, rather than excluding a plausible physical alternative. For analysis of non-associative dynamics and their instabilities, see [87, 41]. □

Remark 8.3 (Physical pathologies of non-associative dynamics)**.** *The exclusion of non-associative algebras is not merely formal. In octonionic quantum mechanics [41], the ambiguity of operator ordering leads to violations of the no-signaling condition: the outcome statistics of a measurement on subsystem A can depend on the bracketing convention chosen for a distant operation on subsystem B. In Jordan-algebraic quantum mechanics [6], the lack of associativity prevents the construction of tensor product state spaces* *with the standard entanglement structure, obstructing quantum error correction and the holographic encoding required by Constraint II. These pathologies are not hypothetical: they represent a quantifiable breakdown of information-processing primitives (teleportation fidelity, entanglement monogamy) that underpin the remainder of the framework.*

Exclusions: Jordan algebras and octonion algebras are non-associative and are excluded by Constraint I.

8.3.2 Constraint II: Metric Compatibility

Lemma 8.4 (Non-Degenerate Bilinear Form from Holographic Error Correction)**.** *For the boundary algebra to support error correction compatible with holographic bulk reconstruction, a non-degenerate bilinear form must be available on the space of boundary operators.*

Proof. Error correction in the holographic context requires quantifying the "distance" between the actual boundary state and the target code subspace. This requires a Lyapunov-type function $V(\delta\rho) \geq 0$ with $\dot{V} < 0$ under the correction protocol, which in turn requires a gradient flow:

$$\dot{\lambda} = -\Gamma\, G^{-1} \nabla_\lambda V, \tag{8.1}$$

where G is a metric on the parameter manifold of boundary states.

Important distinction: The metric G appearing here is an *information-geometric* metric on the space of boundary states (analogous to the Fisher–Rao metric [76]), not the emergent spacetime metric $g_{\mu\nu}$. However, recent results in holographic entanglement [32, 60] establish that linearized perturbations of the bulk metric $\delta g_{\mu\nu}$ are encoded in the boundary modular Hamiltonian and its associated Fisher information. Specifically, the quantum-corrected Ryu–Takayanagi formula [32] implies:

$$\delta S_A = \delta\langle K_A \rangle + \delta S_{\text{bulk}}, \tag{8.2}$$

where K_A is the boundary modular Hamiltonian and S_A is the boundary entanglement entropy. The structure of G_{info} on the boundary therefore constrains the structure of $g_{\mu\nu}$ in the bulk.

The requirement is thus that the boundary algebra carries a non-degenerate bilinear form compatible with this holographic encoding. Standard quantum-state metrics (Bures, Fisher–Rao) are positive-definite and satisfy non-degeneracy, but they do not encode signature information (see Constraint III).

We note that every quantum state space carries a Fisher–Rao metric; the constraint here is not that a metric *exists* (which is trivially satisfied) but that it is *canonically encoded* in the algebraic structure of the generating space, requiring no external specification. □

Exclusions: Lie algebras carry a Killing form, but it may be degenerate (for non-semisimple algebras) and does not naturally encode a quadratic form on the generating vector space. General C^*-algebras and von Neumann algebras support multiple choices of metric (Bures, Hilbert–Schmidt, etc.) but none is canonically "built in" to the algebraic structure itself.

Remark 8.5 (Quantitative cost of external metrics)**.** *The distinction between built-in and external metrics is not merely aesthetic; it has quantifiable information-theoretic* *consequences. For an algebra generated by n directions, an* arbitrary *metric on the generating space requires $n(n+1)/2$ real parameters (a symmetric bilinear form). In a Clifford algebra $Cl(V, q)$, the quadratic form q is specified by n values $\{q(e_i)\}_{i=1}^n$ (plus, in principle, the off-diagonal terms, which are fixed by the anticommutation relations $\{e_i, e_j\} = 2q(e_i, e_j)\mathbf{1}$). However, by diagonalizability, the metric information reduces to n eigenvalues and an $O(n)$ frame—a total of n real parameters in the canonical (diagonal) basis.*

For error correction with an external metric, the correction protocol must independently specify *G at each step. Since a general symmetric bilinear form on n generators has $n(n+1)/2$ independent parameters, the information cost of specifying an external metric grows quadratically with the number of generators. In a Clifford algebra, by contrast, the metric is encoded in the anticommutation relations themselves: once the algebra is fixed, no additional metric specification is needed.*

This distinction becomes physically relevant when the error correction must operate within the thermodynamic bounds of T-DOME (Paper I): the entropy production rate for maintaining NESS is bounded, and the additional overhead of specifying an external metric at each correction cycle may become prohibitive for large n. We emphasize that this is a heuristic argument for the efficiency of built-in metrics, not a formal complexity-theoretic result; making it rigorous would require specifying an explicit computational model for the error-correction protocol.

8.3.3 Constraint III: Causal Channel Encoding

Proposition 8.6 (Indefinite Signature within the Clifford Framework)**.** *Within a Clifford algebraic framework, a non-trivial causal structure—distinguishing time-like from space-like separation—requires the bilinear form to have indefinite signature (p, q) with $p \geq 1$, $q \geq 1$.*

Proof. The argument proceeds from the algebraic characterization of causal structure in quantum field theory [43].

Step 1: Commutativity encodes spacelike separation. In algebraic QFT, two observables A and B are spacelike separated if and only if $[A, B] = 0$ (microcausality axiom). This is not a convention but a physical requirement: spacelike-separated measurements must be jointly performable, which demands commutativity of the corresponding effects.

Step 2: A definite-signature form produces trivial causal structure. Consider a quadratic form q on the generating vector space V. If q is positive-definite, then all generators $v \in V$ satisfy $v^2 = q(v)\mathbf{1} > 0$, and they are algebraically indistinguishable with respect to norm type (all "spacelike"). In the corresponding Clifford algebra $Cl(n, 0)$, the even subalgebra $Cl^0(n, 0) \cong Cl(n-1, 0)$ generates the compact group $\mathrm{Spin}(n)$—no non-compact (boost) generators exist. The resulting automorphism group is compact, so there is no invariant cone structure that could distinguish timelike from spacelike directions: the causal structure is trivial (all directions are equivalent).

Step 3: Non-trivial causal structure requires indefinite signature. By contraposition from Step 2: if q is positive-definite, the automorphism group is compact and the causal structure is trivial. For the even subalgebra to contain both compact generators (spatial rotations) and non-compact generators (boosts)—enabling an invariant causal cone in the representation space—the quadratic form must have indefinite signature: $\mathrm{sig}(q) = (p, r)$ with $p, r \geq 1$. The resulting group $\mathrm{Spin}(p, r)$ is non-compact and admits the required invariant cone structure [43].

We emphasize that this argument does not determine the specific values of p and q. The identification $(p, q) = (1, 3)$ requires additional input: the observational dimensionality of macroscopic spacetime. The present argument establishes that, within the Clifford algebraic setting, indefiniteness is a *necessary condition*—under the Clifford framework adopted in Constraints I and II—for non-trivial causal structure. Alternative algebraic frameworks (e.g., non-associative or higher-categorical) might in principle encode causal distinctions differently; the claim here is conditional on the Clifford structure established by Constraints I and II. □

Exclusions: All positive-definite metrics (including standard Bures and Fisher–Rao on quantum state spaces) fail Constraint III. This is the constraint that separates Clifford algebras (which carry a built-in quadratic form of arbitrary signature) from generic associative algebras with positive-definite metrics.

8.3.4 Exclusion of Alternative Algebras

We now systematically evaluate each candidate from Section 8.2:

Algebra	**I: Assoc.**	**II: Metric**	**III: Indef.**	**Status**
von Neumann (Type III_1)	✓	$\sim$	×	No built-in signature
C^*-algebra (general)	✓	$\sim$	×	Metric not canonical
Lie algebra	✓*	×	—	No quadratic form
Jordan algebra	×	✓	—	Non-associative
Octonion algebra	×	✓	—	Non-associative
Clifford $Cl(V, q)$	✓	✓	✓	**All satisfied**

Table 8.1: Evaluation of candidate algebras against three constraints. ✓: satisfied; ×: violated; $\sim$: partially satisfied (metric exists but is not built-in or canonical). *Lie algebras are not associative, but their universal enveloping algebras are.

The marks "×" for von Neumann and C^*-algebras indicate that these categories are too *general* to encode the required structure canonically, not that they are incompatible: every Clifford algebra is itself a special case of a C^*-algebra and (in finite dimensions) a von Neumann algebra. The point is that additional structure (the quadratic form) must be imposed externally.

The key observation is that Clifford algebras are distinguished by having the quadratic form q *built into* the algebraic structure via the defining relation $v^2 = q(v)\mathbf{1}$. Other associative algebras (von Neumann, C^*) can be *equipped with* metrics, but do not carry a canonical one; the metric is an additional choice external to the algebra. In the holographic context, where the boundary algebra must encode bulk metric information, this built-in feature becomes a substantive advantage rather than a mere convenience.

8.3.5 Worked Example: Qubit Tensor Network

To illustrate how the three constraints operate concretely, consider a tensor network model of holographic bulk reconstruction.

Setup. Take N qubits arranged on a MERA (multiscale entanglement renormalization ansatz) tensor network [99, 106]. The boundary algebra is generated by tensor products of Pauli operators $\{\sigma_x, \sigma_y, \sigma_z\}$ acting on individual qubits.

Constraint I. The Pauli algebra is associative (it consists of 2×2 matrices). If we were to replace the Pauli operators with elements of an octonion algebra (which is non-associative), the isometry conditions defining the MERA network—specifically, $V^\dagger V = \mathbf{1}$, which requires associative composition—would fail.

Constraint II. The Pauli operators satisfy $\{\sigma_i, \sigma_j\} = 2\delta_{ij}\mathbf{1}$, which defines a *positive-definite* quadratic form. This is the Clifford algebra $Cl(3,0)$ (since $\sigma_i^2 = +\mathbf{1}$, the quadratic form is positive-definite). Note that $Cl(3,0) \cong M_2(\mathbb{C})$ while $Cl(0,3) \cong \mathbb{H} \oplus \mathbb{H}$ as real algebras; these are *not* isomorphic. The quadratic form is built into the anticommutation relation.

Constraint III (constructive ansatz). The Pauli algebra alone encodes a Euclidean signature $(3,0)$. To obtain a Lorentzian signature, we *posit* a distinguished direction corresponding to the modular Hamiltonian K, which generates modular flow (the boundary analog of time evolution in the bulk), and *assume* that it anticommutes with the spatial generators:

$$(iK)^2 = -\mathbf{1}, \qquad \sigma_i^2 = +\mathbf{1}, \qquad \{iK, \sigma_i\} = 0. \tag{8.3}$$

Under these assumptions, the extended algebra is $Cl(1,3)$. We emphasize that the anticommutation $\{iK, \sigma_i\} = 0$ is an additional structural assumption, not derived from the tensor network model; no known holographic model produces this anticommutation relation, and establishing (or refuting) it is a major open problem. It amounts to requiring that the modular direction and spatial directions jointly generate a Clifford algebra rather than a more general associative algebra.

Lesson. The tensor network example illustrates how associativity (isometry conditions), built-in metric (Pauli anticommutation), and indefinite signature (modular flow direction) naturally combine to produce Clifford structure in a concrete holographic model.

8.3.6 The Algebraic Compatibility Theorem

Proposition 8.7 (Clifford Compatibility)**.** *Among finitely-generated associative algebras over a vector space V equipped with a non-degenerate quadratic form q, the Clifford algebra $Cl(V,q)$ is the universal (and hence minimal) such structure, by its universal property.*

Proof. Constraint I requires associativity; Constraint II requires a non-degenerate quadratic form on the generating space; Constraint III requires indefinite signature. The universal property of Clifford algebras [50, 29] states that $Cl(V,q)$ is the unique (up to isomorphism) associative algebra generated by V subject to $v^2 = q(v)\mathbf{1}$. Any other associative algebra satisfying these constraints contains $Cl(V,q)$ as a subalgebra (or quotient), making Clifford the minimal compatible structure. □

Remark 8.8 (Scope of the Claim)**.** *Proposition 8.7 is a statement about* algebraic compatibility*, not physical uniqueness. It asserts that Clifford algebra is the natural minimal framework for encoding the three constraints simultaneously. It does not exclude larger* *structures, nor does it claim that physics* must *use the minimal option. The theorem should be understood as identifying an algebraic bottleneck rather than deriving a unique physical theory.*

We note that the final step of the proof—invoking the universal property of $Cl(V, q)$—is a standard algebraic fact (essentially the definition of Clifford algebras). The substantive content of the argument resides in the three constraints*: Lemma 8.2 (associativity as a boundary condition, substantiated by the pathologies of non-associative alternatives), Lemma 8.4 (non-degenerate bilinear form from holographic error correction, connected to the modular Hamiltonian and Fisher information), and Lemma 8.6 (indefinite signature from the algebraic structure of causal order). The theorem assembles these physical requirements into a single algebraic conclusion.*

8.4 Entropic Gravity from Algebraic Structure

8.4.1 Holographic Screen and Einstein Equations

Following Jacobson [52], the entropic force $F = T\nabla S$ and the holographic entropy bound $S \leq A/4G$ reproduce the Einstein field equations in the thermodynamic limit. This derivation assumes local Lorentz invariance—a condition naturally satisfied when the boundary algebra is Clifford-compatible, since $Cl(1, 3)$ contains $\mathrm{Spin}(1, 3)$ (the double cover of the Lorentz group) as a group within its even subalgebra $Cl^0(1, 3)$.

8.4.2 Relation to HAFF Emergence Chain

Within HAFF, geometry emerges via:

$$\text{Observable Algebra} \to \text{Representation} \to \text{Entanglement} \to \text{Connectivity} \to \text{Geometry}$$

The present work adds a constraint on the final arrow: among geometrically admissible coarse-grainings [127], those producing Lorentzian geometry must induce effective algebras compatible with $Cl(1, 3)$.

8.4.3 Relation to Algebraic QFT

In algebraic quantum field theory (AQFT) [43], local observable algebras associated with spacetime regions are generically Type III_1 von Neumann factors. These algebras are associative and support rich mathematical structure, but they do not carry a canonical metric of indefinite signature.

The connection to Lorentzian structure emerges through the Tomita–Takesaki theorem: for any cyclic and separating state, the modular operator Δ generates a one-parameter group (modular flow) that, in the Bisognano–Wichmann theorem, coincides with the boost generator in Rindler spacetime. This modular flow singles out a *time-like direction* within the algebraic structure.

In the HAFF framework, the accessible algebra $\mathcal{A}_\mathbf{c}$ can be understood as a stable subalgebra of a Type III_1 factor. The present analysis suggests that when such a subalgebra supports a Lorentzian bulk description, it must admit a $Cl(1, 3)$ representation—where the modular flow direction provides the time-like generator and spatial locality provides the space-like generators.

This perspective connects the present work to Witten's observation [114] that Type III_1 algebras are essential in gravitational settings, and to Connes' noncommutative geometry program [24], where Clifford algebras play a central role in the spectral characterization of Riemannian (and pseudo-Riemannian) manifolds.

8.5 Discussion

8.5.1 What This Result Does and Does Not Show

Does show: Clifford algebra is the minimal algebraic structure simultaneously satisfying associativity, metric compatibility, and causal channel encoding. The exclusion argument (Table 8.1) demonstrates that alternative algebras fail at least one constraint. The tensor network example (Section 8.3.5) illustrates the constraints in a concrete model.

Does not show: Why $3+1$ dimensions rather than some other (p,q)—the argument constrains to $Cl(p,q)$ for any $p \geq 1$; the value $(1,3)$ requires additional input. That gravity *is* entropic—we derive consistency conditions within the entropic gravity framework. That $Cl(1,3)$ is the *unique* boundary algebra—larger algebras containing $Cl(1,3)$ as a subalgebra are also compatible. A complete theory of quantum gravity.

8.5.2 Convergence with Paper B

The companion paper (Chapter 9) arrives at $Cl(V,q)$ from a completely different direction: thermodynamic stability of persistent open quantum subsystems.

	Paper A (this work)	**Paper B**
Question	What algebra does geometry need?	What algebra does persistence need?
Method	Holographic consistency	Lyapunov stability
Perspective	The world ("ocean")	The subsystem ("fish")
Result	$Cl(1,3)$ from signature	$Cl(V,q)$ from error stability

We note that this convergence is *heuristic rather than deductive*: it suggests that Clifford algebra occupies a distinguished position in the landscape of emergent algebraic structures, but does not constitute a proof. The convergence motivates further investigation, particularly through more elaborate models and deeper connections to established algebraic frameworks.

Remark 8.9 (Logical weight of the constraints). *Constraint I (associativity) is a background requirement automatically satisfied by any operator algebra on a Hilbert space. Constraint II (built-in metric) is a naturality/economy criterion rather than a strict physical necessity—any quantum system carries a Fisher–Rao metric, but requiring it to be algebraically encoded narrows the candidates. Constraint III (indefinite signature) carries the primary discriminating power. The three constraints are therefore not logically independent: the argument reduces to one genuine physical requirement (indefinite signature within an algebraically natural metric framework) plus background assumptions.*

8.5.3 Open Problems

1. **Dimensionality**: What additional constraints (anomaly cancellation, stability of persistent subsystems, observational input) fix $(p, q) = (1, 3)$?

2. **Constructive derivation**: Can Clifford generators be explicitly constructed from modular Hamiltonians or Tomita–Takesaki data in holographic models?

3. **Relation to noncommutative geometry**: How does the present analysis connect to Connes' spectral triples, where Clifford algebras characterize the Dirac operator?

4. **Tensor network realization**: Can the qubit toy model of Section 8.3.5 be made rigorous in the context of holographic error-correcting codes?

8.6 Conclusion

We have argued that the Lorentzian metric structure in entropic gravity frameworks is algebraically constrained by three independent requirements: associativity, metric compatibility (with holographic encoding), and indefinite signature. A systematic exclusion of alternative algebras (von Neumann, C^*, Jordan, Lie, octonion) shows that Clifford algebra $Cl(V, q)$ is the minimal structure satisfying all three simultaneously.

This result connects the top-down perspective of HAFF (geometry emerges from observable algebras) with bottom-up algebraic constraints (any causal geometry must be Clifford-compatible). It does not constitute a derivation of Lorentzian gravity from first principles, but identifies an algebraic bottleneck through which any emergent causal geometry must pass.

Chapter 9

Thermodynamic Stability Constraints on the Operator Algebra of Persistent Open Quantum Subsystems

Paper B — "The Fish"

Originally published: Zenodo, DOI: 10.5281/zenodo.18525890

9.1 Introduction

The interaction of a quantum system with a large environment typically leads to decoherence and thermalization [18, 111]. Maintaining a NESS requires continuous energetic cost [89, 55], which can be modeled as a feedback control process [84, 73]. We address: *What algebraic structures allow the internal control dynamics to remain Lyapunov stable?*

9.1.1 Three-Paper Structure

Paper	**Question**	**Analogy**
HAFF [127]	How does geometry emerge?	Ocean
Q-RAIF A [140]	What algebra does geometry need?	Water
This work	What algebra does survival need?	Fish

9.1.2 Anti-Solipsism Disclaimer

A potential misreading is that the observer "creates" geometry through survival. We explicitly reject this. The claim is structural: under the built-in metric assumption (Constraint II), any subsystem maintaining persistence encodes its environment using a Clifford-compatible algebra. Within HAFF, geometry exists as a stable organizational phase [128]—contingent on physical conditions but objective within them. The present paper argues that subsystems embedded in such a phase must reflect that geometry in their internal algebra—not generate it.

9.1.3 Scope

This work does not claim to derive Clifford algebra from first principles. It argues that, within the variational framework of persistence under Lindblad dynamics, Clifford algebra is the minimal algebraic structure compatible with stable feedback. The argument proceeds by exclusion of alternatives, not by uniqueness proof.

9.2 Variational Bounds on Persistence

Consider $\mathcal{H}_{\text{tot}} = \mathcal{H}_R \otimes \mathcal{H}_E$, with reduced dynamics:

$$\dot{\rho}_R = -i[H_{\text{eff}}, \rho_R] + \mathcal{D}[\rho_R]. \tag{9.1}$$

Definition 9.1 (Persistence Action)**.** $\mathcal{A}[Q] = \int_0^\tau dt\, D_{KL}(\rho(t) \| \rho_{\text{NESS}})$.

Minimizing $\delta\mathcal{A} = 0$ implies a control Hamiltonian $H_{\text{ctrl}}(t)$ generated by an operator algebra $\mathcal{O}$.

9.2.1 Why Lie Algebras Are Insufficient

Standard quantum control theory uses Lie algebra generators [112]: the control Hamiltonian $H_{\text{ctrl}} = \sum_k u_k(t) G_k$ where $\{G_k\}$ generate a Lie algebra $\mathfrak{g}$ via commutators $[G_i, G_j] = i f_{ijk} G_k$.

Lie algebras encode *infinitesimal symmetries*—they specify *which directions* in state space are accessible via control. However, they do not encode *distances* between states. The commutator $[G_i, G_j]$ determines the algebra's structure, but there is no built-in notion of "how far" a correction moves the state.

For error correction, the subsystem must quantify both the *direction* and the *magnitude* of environmental perturbations. This requires a quadratic form $q(v) = \eta_{\mu\nu} v^\mu v^\nu$ on the space of perturbations—which is precisely the additional structure that Clifford algebras provide over Lie algebras.

9.3 Algebraic Constraints on Control Stability

9.3.1 Constraint I: Associativity as Structural Boundary

Lemma 9.2 (Associativity Boundary)**.** *Consistent composition of sequential control operations requires an associative algebra.*

We acknowledge that this constraint is automatically satisfied by operator algebras on Hilbert spaces, where composition of linear maps is inherently associative [18]. Non-associative algebras (Jordan, octonion) are not realistic candidates for quantum dynamics.

Lemma 9.2 therefore functions as a *structural boundary marker*: it delineates the minimal algebraic condition separating consistent from inconsistent dynamics, analogous to how the second law delineates irreversibility without claiming that reversible processes are a realistic threat. For the mathematical structure of non-associative algebras and their dynamical instabilities, see [87, 41]. The physical pathologies of non-associative alternatives are discussed in Remark 8.3 of Paper A.

The substantive constraint is Constraint II, which discriminates among *associative* algebras.

9.3.2 Constraint II: Indefinite Metric for Channel Discrimination

Lemma 9.3 (Metric Constraint). *For a persistent subsystem to distinguish qualitatively different environmental coupling channels and implement directed error correction, the control algebra must carry a non-degenerate bilinear form. Among encodings of the qualitative distinction between channel types, an indefinite signature provides the most parsimonious algebraic encoding that requires no external labeling.*

Proof. Lyapunov stability requires $\dot{V} < 0$ for $V(\delta\rho) \geq 0$, implying gradient flow:

$$\dot{\lambda} = -\Gamma\, G^{-1}\nabla_\lambda V, \tag{9.2}$$

where G is a metric on the control parameter manifold.

Important distinction: G here is an information-geometric metric on the space of control parameters, not the spacetime metric. Standard quantum state metrics (Bures, Fisher–Rao [76]) are positive-definite and satisfy non-degeneracy. However, they are *isotropic*: they treat all perturbation directions equivalently.

In realistic open quantum systems, the environment couples to the subsystem through qualitatively different channels—dissipative (population decay), dephasing (coherence loss), and unitary (Hamiltonian shift). Effective error correction requires distinguishing *qualitatively* between these channel types.

A positive-definite anisotropic metric (e.g., $\mathrm{diag}(a, b, c)$ with $a \neq b \neq c > 0$) can encode quantitative differences between directions. In principle, an external labeling scheme (e.g., a discrete index classifying channels as "dissipative" vs. "unitary") could supplement such a metric to encode qualitative distinctions.

However, an indefinite quadratic form $q(v) = \eta_{\mu\nu}v^\mu v^\nu$ with $\mathrm{sig}(\eta) = (p, r)$, $p, r \geq 1$, encodes the dichotomy *intrinsically*: positive-norm directions ($v^2 > 0$) correspond to one class, negative-norm directions ($v^2 < 0$) to another, with the distinction built into the algebra via $v^2 = q(v)\mathbf{1}$. No external labeling is needed. The assignment of negative norm to dissipative directions is a structural *ansatz* motivated by the algebraic analogy, not a derived result from open quantum systems theory; establishing this correspondence rigorously remains an open problem.

We therefore argue that indefinite signature is the *most parsimonious* encoding of the reversible/irreversible dichotomy within an algebraic framework, though we do not claim it is the *unique* possibility. □

9.3.3 Exclusion of Alternative Algebras

Algebra	**I: Assoc.**	**II: Indef. q**	**Status**
von Neumann (III_1)	✓	×	No built-in q
C^*-algebra	✓	×	Positive-definite only
Lie algebra	✓*	×	Killing form, no q
Jordan algebra	×	—	Non-associative
Clifford $Cl(V, q)$	✓	✓	**Minimal**

Table 9.1: Systematic evaluation of candidate control algebras. *Via universal enveloping algebra.

The exclusion argument shifts the burden from "why Clifford?" to "why not the alternatives?"—and the answer is that no other standard algebraic framework carries a built-in indefinite quadratic form encoding channel discrimination.

9.3.4 Worked Example: Controlled Qubit Under Lindblad Dynamics

Setup. Consider a single qubit coupled to a thermal bath at inverse temperature β, with Lindblad dissipator:

$$\mathcal{D}[\rho] = \gamma_\downarrow \mathcal{L}[\sigma_-]\rho + \gamma_\uparrow \mathcal{L}[\sigma_+]\rho + \gamma_\phi \mathcal{L}[\sigma_z]\rho, \tag{9.3}$$

where $\mathcal{L}[L]\rho = L\rho L^\dagger - \frac{1}{2}\{L^\dagger L, \rho\}$, and $\gamma_\downarrow$, $\gamma_\uparrow$, γ_ϕ are decay, excitation, and dephasing rates.

Control algebra. The control Hamiltonian is $H_{\text{ctrl}} = \sum_i u_i(t)\, \sigma_i$ where $\{\sigma_x, \sigma_y, \sigma_z\}$ are Pauli operators. These satisfy $\{\sigma_i, \sigma_j\} = 2\delta_{ij}\mathbf{1}$—the defining relation of $Cl(3,0)$.

Lyapunov function. Take $V = D_{KL}(\rho \| \rho_{\text{NESS}})$ where ρ_{NESS} is the thermal state. The gradient $\nabla_u V$ is well-defined because the Pauli algebra carries a natural inner product (the quadratic form $q(\sigma_i) = +1$). The control protocol $u_i(t) = -\alpha\, \partial V/\partial u_i$ yields:

$$\dot{V} = -\alpha \sum_i \left(\frac{\partial V}{\partial u_i}\right)^2 \leq 0, \tag{9.4}$$

which is strictly negative away from the NESS.

Failure mode without built-in metric. If the control algebra were an abstract Lie algebra $\mathfrak{su}(2)$ (same generators, but with only the commutator structure $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$ and no anticommutator/metric), the control protocol could specify *rotation directions* in Bloch sphere but could not canonically quantify *how large* a correction to apply. The gradient flow (9.2) would require importing an external metric (e.g., the Killing form of $\mathfrak{su}(2)$, which happens to be proportional to δ_{ij}). We acknowledge that semisimple Lie algebras do carry canonical metrics (the Killing form); the advantage of the Clifford structure is specifically that it encodes a metric on the *generating vector space* V, with potentially indefinite signature, rather than only on the Lie algebra.

In the Pauli/Clifford case, the metric is *internal*: the same algebraic structure that generates rotations also defines distances. This unification is what makes Clifford algebras uniquely suited for feedback control where both direction and magnitude matter.

Extension to indefinite signature. When the subsystem must distinguish dissipative from unitary perturbations—e.g., $\gamma_\downarrow \neq 0$ (dissipative) versus Hamiltonian noise (unitary)—the control space naturally splits into sectors of different character. An indefinite quadratic form provides a parsimonious algebraic encoding of this split, upgrading $Cl(3,0)$ to $Cl(p,q)$ with appropriate signature, though alternative encodings using positive-definite metrics supplemented by external labeling are not excluded in principle.

9.3.5 The Algebraic Compatibility Theorem

Proposition 9.4 (Persistence Compatibility)**.** *Among associative algebras encoding n orthogonal control channels with a built-in non-degenerate quadratic form, the Clifford algebra $Cl(V, q)$ is the universal minimal structure, by its universal property.*

Proof. Constraint I (associativity) is given. Constraint II requires a non-degenerate quadratic form q on the generating space V, with indefinite signature when channel discrimination is required. The universal property of Clifford algebras [50] identifies $Cl(V, q)$ as the unique associative algebra generated by V subject to $v^2 = q(v)\mathbf{1}$. □

Corollary 9.5. *Under the built-in metric assumption (Constraint II), any subsystem maintaining NESS for $\tau \gg \tau_{\text{relax}}$ while discriminating among environmental channels encodes its boundary using a Clifford-compatible algebra.*

Remark 9.6 (Natural Selection, Not Design)**.** *The theorem establishes a selection principle. Subsystems do not "choose" Clifford algebra; only Clifford-compatible structures persist when channel discrimination is required. This is algebraic natural selection.*

Remark 9.7 (Content of the Argument)**.** *As with Proposition 8.7 in Paper A, the final step of the proof—invoking the universal property of $Cl(V, q)$—is a standard algebraic identity. The substantive content resides in the constraints: Lemma 9.2 (associativity as a boundary condition, with non-associative alternatives producing quantifiable pathologies; see Remark 8.3), and Lemma 9.3 (indefinite metric from the requirement of channel discrimination in directed error correction). The worked example in Section 9.3.4 demonstrates that the constraints are simultaneously satisfiable in a concrete Lindblad model.*

9.4 Contextual Relations

9.4.1 Convergence with Paper A

The companion paper [140] argues that $Cl(1, 3)$ is the minimal algebra compatible with emergent Lorentzian geometry in entropic gravity.

	Paper A	**This work**
Starting point	Holographic boundary	Open subsystem
Method	Signature selection	Lyapunov stability
Key constraint	Causal ordering	Channel discrimination
Result	$Cl(1, 3)$	$Cl(V, q)$

We note explicitly that this convergence is *heuristic rather than deductive.* Two arguments pointing to the same algebraic structure from different directions is suggestive but does not constitute proof. The convergence motivates further investigation through explicit models and connections to established frameworks, not a claim of mathematical necessity.

9.4.2 Relation to Quantum Control Theory

Standard quantum control operates within a Lie algebraic framework [112]: controllability is characterized by the Lie algebra generated by the drift and control Hamiltonians. This framework is complete for determining *reachability* of target states.

However, Lie algebras encode symmetries (via commutators) without encoding distances (via quadratic forms). When the control objective is not merely reachability but *stabilization against stochastic perturbations*—as in NESS maintenance—both direction and magnitude of corrections must be specified. Clifford algebras provide this additional structure through their built-in quadratic form, complementing rather than replacing the Lie algebraic framework.

9.4.3 Relation to Decoherence-Free Subspaces

Decoherence-free subspaces (DFS) [119, 64] represent subsystems that are passively protected from environmental noise by symmetry. The present analysis addresses the *active* counterpart: subsystems that maintain coherence through continuous feedback. In both cases, the algebraic structure of the system-environment interaction determines which subsystems can persist. The Clifford constraint identified here applies to the active case; DFS theory applies to the passive case. A unified treatment remains an open problem.

9.5 Discussion

What this result does show: Clifford algebra is the minimal algebraic structure satisfying both associativity and built-in indefinite metric among standard algebraic candidates. The exclusion of alternatives (Table 9.1) and the controlled qubit example (Section 9.3.4) provide concrete support.

What this result does not show: That Clifford algebra is the *unique* solution—larger structures are also compatible. That non-Clifford feedback is impossible in all settings—it is possible when channel discrimination is not required. A derivation from first principles independent of the variational framework assumed here.

9.6 Conclusion

We have argued that geometric (Clifford) algebra structure is a natural minimal requirement for persistent subsystems that must discriminate among environmental coupling channels: (i) associativity is a structural boundary condition; (ii) a built-in indefinite quadratic form is required for channel discrimination and directed error correction; (iii) no standard alternative algebra satisfies both with built-in structure.

Combined with the companion paper's holographic constraints (Chapter 8), this suggests $Cl(V, q)$ occupies a distinguished position as the algebraic structure simultaneously compatible with geometric consistency and thermodynamic persistence.

Chapter 10

The Realizability Bridge: Algebraic Closure in the Q-RAIF Framework

Paper C — "The Bridge"

Originally published: Zenodo, DOI: 10.5281/zenodo.18528934

Abstract

This addendum provides a minimal mathematical bridge between the two foundational papers of the **Q-RAIF (Quantum Reference Algebra for Information Flow)** framework. Paper A (Chapter 8) establishes that the observable algebra of a holographically consistent universe must contain $Cl(1,3)$ as its minimal Clifford-compatible structure. Paper B (Chapter 9) establishes that the control algebra of a persistent subsystem must be Cliffordian $Cl(V,q)$ to ensure Lyapunov stability under entropic constraints.

Here we prove the **Closure Theorem**: any *physically realizable* control algebra must embed into the environmental algebra as a subalgebra. We formalize the required feedback synchrony via a *Same-Clock* co-indexing lemma, ensuring the feedback loop is thermodynamically potent.

This note does not modify Papers A or B; it supplies only the realizability bridge needed for algebraic closure.

Keywords: Q-RAIF, realizability, representation, operator algebra, Clifford algebra, open quantum systems, Lyapunov stability, algebraic closure

10.1 Introduction

10.1.1 Context: The Q-RAIF Program

The Quantum Reference Algebra for Information Flow (Q-RAIF) framework investigates what algebraic structures are *necessary*—as opposed to merely convenient—for the self-consistent description of physical reality and persistence within it. The program builds on the Holographic Alaya-Field Framework (HAFF) [127, 128], which establishes that geometry emerges from coarse-graining of observable algebras.

Paper	Question	Analogy	Result
HAFF [127]	How does geometry emerge?	Ocean	Algebra $\to$ Geometry
Q-RAIF A [140]	What algebra does geometry need?	Water	$Cl(1,3)$
Q-RAIF B [141]	What algebra does survival need?	Fish	$Cl(V,q)$
This work	Must the fish fit the water?	Bridge	$Cl(V,q) \hookrightarrow Cl(1,3)$

10.1.2 The Logical Gap

Papers A and B independently arrive at Clifford algebra from opposite directions. Both papers explicitly note that this convergence is *heuristic rather than deductive* [140, 141]. The present note closes the gap by proving a realizability constraint: the internal control algebra of any persistent subsystem must be representable within the external observable algebra.

10.1.3 Scope

This addendum introduces no new physical assumptions. It uses only the objects and results already established in Papers A and B, and derives their mutual constraint. Papers A and B remain unmodified.

10.2 Setup and Prerequisites

Let $\mathcal{U}$ be a universe described by the Q-RAIF framework.

- **Environment ("water").** Let $\mathcal{A}_{\text{ext}}$ denote the algebra of observables accessible at the holographic boundary. Paper A (Chapter 8) argues that $\mathcal{A}_{\text{ext}}$ must contain $Cl(1,3)$ as its minimal Clifford-compatible subalgebra (Proposition 8.7, "Clifford Compatibility").
- **Subsystem ("fish").** Let $\mathcal{O}_{\text{int}}$ denote the internal control algebra of a persistent subsystem $R \subset \mathcal{U}$. Paper B (Chapter 9) argues that thermodynamic persistence requires $\mathcal{O}_{\text{int}} \cong Cl(V,q)$ for some (V,q) (Proposition 9.4, "Persistence Compatibility").

The remaining logical gap is the relationship between $\mathcal{O}_{\text{int}}$ and $\mathcal{A}_{\text{ext}}$: can a stable Clifford control algebra exist while being structurally disjoint from the available environmental observables?

10.3 Realizability and Same-Clock Co-Indexing

Definition 10.1 (Algebraic Realizability)**.** *A control algebra $\mathcal{O}_{\text{int}}$ is* ***physically realizable*** *within an environment $\mathcal{A}_{\text{ext}}$ if there exists a homomorphism*

$$\phi : \mathcal{O}_{\text{int}} \to \mathcal{A}_{\text{ext}} \tag{10.1}$$

such that $\mathrm{Im}(\phi)$ *has non-zero action on the interaction Hamiltonian H_{int}, i.e.,* $[\mathrm{Im}(\phi), H_{\text{int}}] \neq 0$*. This ensures that the controller can physically influence the system-environment boundary.*

Let I be an operational/causal index set (e.g., proper-time frames or discretized event slices). For a subset $J \subseteq I$, write $\mathcal{A}|_J$ for the restriction of an algebra $\mathcal{A}$ to the index set J.

Lemma 10.2 (Same-Clock / Co-Indexing). *For a feedback loop to be causally closed and thermodynamically potent (capable of entropy export [89]), there must exist non-null index overlap between control and feedback windows: there exist $J_{\mathrm{ctrl}}, J_{\mathrm{env}} \subseteq I$ such that*

1. ***Non-null intersection:*** $J_{\mathrm{ctrl}} \cap J_{\mathrm{env}} \neq \emptyset$.
2. ***Window integrity:*** *on any critical lookback window $W \subseteq J_{\mathrm{ctrl}} \cap J_{\mathrm{env}}$ used to define the controller, $\mathcal{A}_{\mathrm{ext}}|_W$ is well-defined (no holes on W).*

Justification. If $J_{\mathrm{ctrl}} \cap J_{\mathrm{env}} = \emptyset$, the control action is operationally decoupled from environmental feedback, so no entropy export channel exists; persistence (NESS [89]) fails. If window integrity fails on a critical lookback window W, the feedback map—and thus the Lyapunov descent condition (Eq. (4) of Paper B (Chapter 9))—is not definable on the operational window. Therefore both conditions are necessary. □

10.3.1 Semisimplicity of Clifford Algebras

Lemma 10.3 (Injectivity from Semisimplicity). *Let (V, q) be a finite-dimensional real vector space with non-degenerate quadratic form. Then $Cl(V, q)$ is semisimple. If $\dim V$ is even, $Cl(V, q)$ is simple, and every non-zero algebra homomorphism $\phi : Cl(V, q) \to \mathcal{A}$ is injective.*

Proof. By the periodicity theorem for real Clifford algebras [4, 65], $Cl(V, q)$ with non-degenerate q is isomorphic to a matrix algebra $M_n(K)$ or a direct sum $M_n(K) \oplus M_n(K)$, where $K \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$ depends on the signature and dimension modulo 8. In either case the algebra is semisimple.

When $\dim V$ is even, $Cl(V, q)$ is simple (a single matrix block). The kernel of any algebra homomorphism is a two-sided ideal; a simple algebra admits no proper non-trivial ideals, so $\ker \phi = \{0\}$ whenever $\phi \neq 0$. □

10.4 The Closure Theorem

Theorem 10.4 (Q-RAIF Algebraic Closure). *Assume $\mathcal{A}_{\mathrm{ext}} \supseteq Cl(1, 3)$ (Paper A (Chapter 8)). Let R be a persistent subsystem whose control algebra satisfies $\mathcal{O}_{\mathrm{int}} \cong Cl(V, q)$ (Paper B (Chapter 9)). If $\mathcal{O}_{\mathrm{int}}$ is realizable in $\mathcal{A}_{\mathrm{ext}}$ (Definition 10.1) and the Same-Clock conditions of Lemma 10.2 hold, then the effective control algebra*

$$\mathcal{O}_{\mathrm{eff}} := \mathrm{Im}(\phi) \subseteq \mathcal{A}_{\mathrm{ext}} \tag{10.2}$$

is a Clifford subalgebra of the external geometry.

Proof. By realizability, there exists a homomorphism $\phi : \mathcal{O}_{\mathrm{int}} \to \mathcal{A}_{\mathrm{ext}}$ with non-trivial image. The operational content of the controller is its image $\mathcal{O}_{\mathrm{eff}} = \mathrm{Im}(\phi)$. Since $\mathcal{O}_{\mathrm{int}} \cong Cl(V, q)$ by the persistence requirement (Proposition 9.4), and ϕ is structure-preserving, $\mathcal{O}_{\mathrm{eff}}$ inherits the Clifford relations $v^2 = q(v)\mathbf{1}$ [50]. Moreover, by Lemma 10.3, if $\dim V$ is even then $Cl(V, q)$ is simple and ϕ is necessarily injective; the image is therefore isomorphic to $Cl(V, q)$ itself, giving a genuine embedding $Cl(V, q) \hookrightarrow \mathcal{A}_{\text{ext}}$. In the physically relevant case ($\dim V = 4$, signature $(1, 3)$ or compatible sub-signature), the even-dimensionality condition is satisfied. Since $\mathcal{O}_{\text{eff}} \subseteq \mathcal{A}_{\text{ext}}$, the internal geometry (V, q) is induced by a restriction of the ambient algebraic structure. □

Remark 10.5 (Scope of the closure theorem). *The logical structure of Theorem 10.4 is essentially: "a Clifford algebra, homomorphically embedded in a larger algebra, lands as a Clifford subalgebra." The non-trivial content resides not in this deduction but in the* input conditions*: (i) Paper A's argument that $\mathcal{A}_{\text{ext}} \supseteq Cl(1, 3)$; (ii) Paper B's argument that persistence forces $\mathcal{O}_{\text{int}} \cong Cl(V, q)$; (iii) the injectivity guarantee from semisimplicity (Lemma 10.3). The theorem's role is to assemble these independently substantiated conditions into a single algebraic closure statement, not to introduce new mathematical content.*

Corollary 10.6 (No Ghost Algebra). *A control algebra that is mathematically stable (Cliffordian) but not representable in $\mathcal{A}_{\text{ext}}$ is not physically realizable. In particular, a control structure with signature incompatible with $(1, 3)$ cannot underwrite persistent feedback in a universe whose observable algebra contains $Cl(1, 3)$. Here "signature incompatible with $(1, 3)$" means that the quadratic space (V', q') does not embed isometrically into $(\mathbb{R}^{1,3}, \eta)$, where we use the convention that p generators square to $+1$ and q generators square to -1.*

10.5 Discussion

10.5.1 What This Result Does and Does Not Show

Does show: Realizability forces the internal control algebra of a persistent subsystem to embed into the external observable algebra. Combined with Papers A and B, this converts the previously heuristic convergence ($Cl(V, q)$ from stability, $Cl(1, 3)$ from geometry) into a constrained embedding: $Cl(V, q) \hookrightarrow Cl(1, 3)$.

Does not show: That ϕ must be injective in general—however, by Lemma 10.3, injectivity *is* guaranteed when $\dim V$ is even (which includes the physically relevant case $\dim V = 4$). For odd $\dim V$, the image $\text{Im}(\phi)$ is isomorphic to a simple factor of $Cl(V, q)$ and still carries the Clifford structure. That the specific signature (V, q) is uniquely determined—only that it must be compatible with $(1, 3)$. That this constitutes a derivation of physics from first principles—it is a consistency constraint within the Q-RAIF framework.

10.5.2 The Bridge Statement

Remark 10.7 (Closing the Loop). *Paper A fixes the realizable operator content of the world ($\mathcal{A}_{\text{ext}}$). Paper B fixes the algebraic form required for persistence ($\mathcal{O}_{\text{int}}$). Theorem 10.4 locks them together: realizable persistence forces the agent's control algebra to be built from the same algebraic atoms as its environment. The fish's gills must be made of water's molecules.*

10.5.3 Connection to HAFF

Within the HAFF program [127, 128], geometry emerges from coarse-graining of observable algebras. The Closure Theorem adds a further structural consequence: not only does the world's geometry emerge from its algebra, but any persistent subsystem's internal geometry is *constrained to be a restriction* of that emergent geometry. This is algebraic natural selection operating at the level of geometric structure.

Part III

T-DOME: Thermodynamic Dynamics of Observer-Memory Entanglement

Chapter 11

Non-Markovian Memory and the Thermodynamic Necessity of Temporal Accumulation

Paper I — "The Seed"

Originally published: Zenodo, DOI: 10.5281/zenodo.18574342

Abstract

We investigate the thermodynamic constraints on open quantum systems that must persist far from equilibrium in stochastic environments. Working within the framework of stochastic thermodynamics and information thermodynamics (Sagawa–Ueda), we define a *survival functional* $\mathcal{S} := \Delta F - W$ measuring the difference between the non-equilibrium free energy gained and the work invested by an agent.

We prove a **Markovian Ceiling**: for any open-loop Markovian (GKSL) dynamics with no measurement or feedback, $\mathcal{S} \leq 0$—the agent cannot thermodynamically "profit." We then derive an exact identity—valid for *arbitrary* (possibly correlated) initial states under autonomous evolution in the weak-coupling limit—expressing the survival functional in terms of the change in system–environment mutual information and bath displacement: $\beta\,\mathcal{S} = -\Delta I(S{:}E) - \Delta D_{\mathrm{KL}}(\rho_E \| \rho_E^{\mathrm{th}})$. Pre-existing correlations $I(S{:}E;\,0) > 0$, built during prior interaction epochs, serve as a consumable thermodynamic resource; their consumption during non-Markovian backflow intervals yields $\mathcal{S} > 0$, bounded by the initial correlation budget.

This establishes **memory as a thermodynamic necessity** for sustained far-from-equilibrium persistence. The memory kernel induces a causal partial order on system trajectories that, when restricted to the classical sector selected by decoherence (quantum Darwinism), is consistent with the accessibility ordering of the Holographic Alaya-Field Framework (HAFF). A worked example—a spin-boson model with Lorentz–Drude spectral density—illustrates how non-Markovian backflow enables free-energy extraction unavailable to memoryless systems.

Finally, using the entropy rate and predictive information from computational mechanics, we quantify the intrinsic cost of memory and identify the **Memory Catastro-**

phe: unbounded memory under finite energy leads to thermodynamic collapse, motivating the symmetry-breaking mechanism of Paper II.

Keywords: non-Markovian dynamics, open quantum systems, Nakajima–Zwanzig equation, memory kernel, thermodynamic arrow of time, information backflow, entropy production, stochastic thermodynamics

11.1 Introduction

11.1.1 Context: The Problem of Persistence

A quantum system coupled to a thermal environment generically relaxes toward equilibrium. This is the content of the *zeroth crisis*: absent special structure, every open subsystem is eventually erased by thermal noise [18].

Yet the physical world contains persistent far-from-equilibrium structures—from molecular machines to living organisms—that maintain themselves against the entropic tide for timescales vastly exceeding their intrinsic relaxation times. What structural feature of their dynamics makes this possible?

The standard answer invokes free-energy input: a persistent system is one that continuously imports low-entropy energy and exports high-entropy waste [88]. This is correct but incomplete. Two systems receiving *identical* free-energy flux from *identical* environments may exhibit vastly different persistence characteristics. The distinguishing factor, we argue, is *memory*—the capacity to condition present dynamics on past environmental states.

11.1.2 Position within the Series

This paper is the first of three constituting the **T-DOME** (Thermodynamic Dynamics of Observer-Memory Entanglement) framework, the third pillar of a three-paper program.

Framework	**Question**		**Result**	**Status**
HAFF [127, 128]	How does geometry emerge?	Ocean	Algebra $\to$ Geometry	Complete
Q-RAIF [140, 141]	What algebra must an observer have?	Fish	$Cl(V, q) \hookrightarrow Cl(1, 3)$	Complete
T-DOME I (this work)	Why must agents carry memory?	Seed	Markovian ceiling; memory as necessity	**This paper**
T-DOME II	Why must agents break symmetry?	Ego	Reference-frame selection	Planned
T-DOME III	How does self-calibration arise?	Loop	Fisher self-referential bound	Planned

The three T-DOME papers form an irreversible logical chain. Each resolves a survival crisis created by its predecessor:

1. **Paper I (The Seed):** Without memory, a system is trapped in the *Markovian present*—no accumulation, no temporal arrow, inevitable thermal death. Memory breaks this trap but floods the system with unbounded historical data.

2. **Paper II (The Ego):** Unbounded memory under finite computational resources causes processing collapse. Spontaneous symmetry breaking of the reference frame (establishing a "self") resolves the overload but introduces systematic bias.

3. **Paper III (The Loop):** Uncorrected bias diverges from a changing environment. A self-referential calibration loop (monitoring one's own prediction error) resolves the bias but requires the system to "observe its own observation"—closing the self-calibration loop.

The present paper addresses only the first link in this chain.

11.1.3 Relation to HAFF Paper F

HAFF Paper F [130] establishes the arrow of time as the direction of *accessibility propagation*: informational redundancy $\mathcal{R}(\hat{O})$ generically expands, inducing a partial order $\prec$ on observable algebras. That analysis is purely algebraic—it characterizes temporal asymmetry without invoking dynamics.

The present paper complements Paper F by identifying the *dynamical* origin of temporal asymmetry: the non-Markovian memory kernel $\mathcal{K}(t, s)$. We show (Section 11.5) that the partial order induced by the kernel's temporal support embeds into the HAFF accessibility ordering as a sub-structure. The two descriptions are dual faces of the same phenomenon: Paper F provides the algebraic skeleton; Paper I provides the dynamical muscle.

11.1.4 Scope and Disclaimers

1. This work does *not* claim that non-Markovian dynamics is sufficient for persistence. Memory is identified as *necessary* under the conditions specified; sufficiency requires additional structure (Papers II and III).

2. We do *not* claim that all non-Markovian systems outperform all Markovian systems. The theorem establishes that the supremum of survival efficiency over non-Markovian dynamics strictly exceeds the Markovian supremum.

3. We do *not* derive the specific form of the memory kernel from first principles. The kernel is treated as a structural feature of the system-environment coupling.

4. The term "agent" is used in the control-theoretic sense (a subsystem that acts on its environment to maintain a target state) and carries no implication of consciousness, intention, or subjective experience.

5. A broader structural analogy with classical philosophical concepts of temporal persistence exists but is outside the scope of this paper.

11.2 Mathematical Preliminaries

11.2.1 Open Quantum Systems: The Markovian Baseline

Consider a bipartite Hilbert space $\mathcal{H} = \mathcal{H}_R \otimes \mathcal{H}_E$, where R denotes the "agent" (reduced system) and E the environment. The total Hamiltonian is

$$H = H_R \otimes \mathbb{1}_E + \mathbb{1}_R \otimes H_E + \lambda\, H_{\text{int}}, \tag{11.1}$$

where λ parametrizes the coupling strength.

Under the Born–Markov and secular approximations, the reduced dynamics of $\rho_R(t) = \mathrm{Tr}_E[\rho(t)]$ is governed by the Gorini–Kossakowski–Sudarshan–Lindblad (GKSL) master equation [66, 36]:

$$\dot{\rho}_R(t) = -i[H_{\text{eff}},\, \rho_R(t)] + \sum_k \gamma_k \left(L_k \rho_R(t) L_k^\dagger - \tfrac{1}{2}\{L_k^\dagger L_k,\, \rho_R(t)\} \right), \tag{11.2}$$

with $\gamma_k \geq 0$ and Lindblad operators $\{L_k\}$.

Remark 11.1 (Markovian = Memoryless)**.** *The GKSL equation is* time-local*: $\dot{\rho}_R(t)$ depends only on $\rho_R(t)$, never on $\rho_R(s)$ for $s < t$. Physically, this corresponds to an environment with vanishing correlation time ($\tau_E \to 0$): the bath "forgets" its interaction with the system instantaneously. The semigroup property $\Lambda(t+s) = \Lambda(t)\Lambda(s)$ ensures complete positivity at all times but precludes any information backflow from environment to system [80].*

11.2.2 Beyond Markov: The Nakajima–Zwanzig Equation

When the environmental correlation time τ_E is non-negligible, the Born–Markov approximation fails. The exact reduced dynamics is captured by the Nakajima–Zwanzig (NZ) integro-differential equation [70, 123]:

$$\dot{\rho}_R(t) = -i[H_{\text{eff}},\, \rho_R(t)] + \int_0^t ds\, \mathcal{K}(t,s)\, \rho_R(s), \tag{11.3}$$

where $\mathcal{K}(t,s)$ is the **memory kernel**—a superoperator encoding the influence of the system's entire history on its present dynamics.

Definition 11.2 (Memory Kernel)**.** *The memory kernel $\mathcal{K} : [0,\infty)^2 \to \mathcal{L}(\mathcal{B}(\mathcal{H}_R))$ is the superoperator satisfying* (11.3)*. It encodes two types of information:*

1. ***Environmental structure:*** *the spectral density, correlation functions, and non-equilibrium features of the bath;*

2. ***Temporal reach:*** *the effective support $\tau_{\text{mem}} := \inf\{\tau : \|\mathcal{K}(t,s)\| < \epsilon\ \forall\, t - s > \tau\}$, the "memory depth."*

The Markovian limit corresponds to $\mathcal{K}(t,s) \to \mathcal{K}_0\, \delta(t-s)$, recovering the GKSL generator.

Remark 11.3 (Information Backflow)**.** *Non-Markovian dynamics admits* information backflow*: the distinguishability of two initial states, as measured by the trace distance $D(\rho_1(t), \rho_2(t))$, can temporarily increase [17]. This is the operational signature of memory—the environment returns previously absorbed information to the system.*

11.2.3 Thermodynamic Framework

We adopt the framework of stochastic thermodynamics for open quantum systems [31]. The following conventions are fixed throughout.

Definition 11.4 (Thermodynamic Setup)**.**

1. ***Hamiltonian decomposition.*** *The system Hamiltonian is* $H_S(t) = H_R + H_{\text{ctrl}}(t)$*, where* H_R *is the* fixed *bare Hamiltonian and* $H_{\text{ctrl}}(t)$ *is the agent's time-dependent control protocol. The bath Hamiltonian* H_E *and coupling* H_{int} *are as in* (11.1)*.*

2. ***Reference state.*** *The thermal equilibrium state of the bare Hamiltonian is*

$$\rho_{\text{eq}} := \frac{e^{-\beta H_R}}{Z_R}, \qquad Z_R := \operatorname{tr}\big(e^{-\beta H_R}\big), \qquad \beta := (k_B T)^{-1}. \tag{11.4}$$

Since H_R *is time-independent,* ρ_{eq} *is a well-defined, fixed reference throughout the protocol.*

3. ***Non-equilibrium free energy.*** *For any state* ρ *of the reduced system,*

$$F(\rho) := \operatorname{tr}(\rho\, H_R) + \beta^{-1} \operatorname{tr}(\rho \ln \rho) = \langle H_R \rangle_\rho - \beta^{-1} S(\rho), \tag{11.5}$$

where $S(\rho) = -\operatorname{tr}(\rho \ln \rho)$ *is the von Neumann entropy. The equilibrium value is* $F_{\text{eq}} = -\beta^{-1} \ln Z_R$*.*

4. ***Free energy–relative entropy identity.***

$$D_{\text{KL}}(\rho \| \rho_{\text{eq}}) = \beta\big(F(\rho) - F_{\text{eq}}\big) \geq 0. \tag{11.6}$$

Thus D_{KL} *measures the free-energy surplus in units of* $k_B T$*.*

5. ***Work.*** *The work performed on the system by the control protocol over* $[0, \tau]$ *is*

$$W[0,\tau] := \int_0^\tau \operatorname{tr}\left(\rho(t)\, \frac{\partial H_{\text{ctrl}}}{\partial t}\right) dt. \tag{11.7}$$

6. ***Entropy-production functional.*** *The* generalised entropy production *over* $[0, \tau]$ *is*

$$\Sigma[0,\tau] := \beta\big(W[0,\tau] - \Delta F\big), \tag{11.8}$$

where $\Delta F = F(\rho(\tau)) - F(\rho(0))$*. For uncorrelated (product) initial states,* $\Sigma \geq 0$ *recovers the standard second-law bound. For initially correlated states,* Σ *can be* negative*, reflecting the consumption of pre-existing correlations (see Remark 11.27).*

Remark 11.5 (Why H_R is fixed)**.** *The bare Hamiltonian* H_R *defines the system's energy scale and hence the reference state* ρ_{eq}*. The agent acts on the world through* $H_{\text{ctrl}}(t)$*, which may be time-dependent. This separation ensures that* ρ_{eq} *is well-defined and time-independent, avoiding the ambiguity that arises when the full* $H_S(t)$ *is used to define the thermal reference.*

Definition 11.6 (Standing Assumptions)**.** *The following minimal assumptions are in force throughout Sections 11.4–11.6 unless stated otherwise. Every main result (Lemma 11.20, Theorem 11.22, Corollary 11.26) relies* only *on items* (A1)–(A5) *below.*

(A1) ***Finite-dimensional bipartite system.*** *$\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_E$, with total Hamiltonian* (11.1) *and global unitary evolution $U(t) = \mathcal{T}\exp\Big(-i\int_0^t H(s)\,ds\Big)$.*

(A2) ***Weak coupling.*** *The system–environment interaction satisfies $\lambda \ll 1$ in* (11.1)*, so that $\Delta\langle H_{\rm int}\rangle = O(\lambda)$ [18]. Energy conservation is then $\Delta\langle H_R\rangle + \Delta\langle H_{\rm ctrl}\rangle + \Delta\langle H_E\rangle \approx 0$ up to controlled $O(\lambda)$ corrections.*

(A3) ***Fixed environmental reference.*** *$\rho_E^{\rm th} := e^{-\beta H_E}/Z_E$ is a fixed* bookkeeping *Gibbs state at inverse temperature β. The* actual *initial bath state $\rho_E(0)$ need not coincide with $\rho_E^{\rm th}$; when $\rho_E(0) \neq \rho_E^{\rm th}$, the quantity $D_{\rm KL}(\rho_E(t)\|\rho_E^{\rm th})$ tracks the nonequilibrium free energy stored in the bath relative to this reference. The bath Hamiltonian H_E is time-independent.*

(A4) ***Arbitrary initial state.*** *The total initial state $\rho_{SE}(0)$ is* not *required to be a product state. In particular, initial system–environment correlations $I(S{:}E;\,0) > 0$ and initial bath displacement $D_{\rm KL}(\rho_E(0)\|\rho_E^{\rm th}) > 0$ are both permitted.*

(A5) ***Regularity.*** *All quantum states appearing in the thermodynamic identities are assumed to have full rank (or are restricted to their support), so that all relative entropies $D_{\rm KL}(\rho\|\sigma)$ are finite.*

Remark 11.7 (Bookkeeping conventions)**.** *The heat absorbed by the environment is $Q := \Delta\langle H_E\rangle = \mathrm{Tr}[\rho_E(\tau)H_E] - \mathrm{Tr}[\rho_E(0)H_E]$ (matching Esposito* et al. *[31]). We define $\Sigma := \beta(W - \Delta F)$ as a* generalised entropy-balance functional*; for correlated initial conditions Σ need not be nonnegative (see Remark 11.27).*

11.2.4 The Survival Functional

We now define the central quantity of this paper.

Definition 11.8 (Survival Functional)**.** *For a reduced system R evolving under dynamics Λ over $[0,\tau]$, the **survival functional** is*

$$\mathcal{S}[\Lambda,\tau] := \Delta F - W[0,\tau] = \big[F(\rho(\tau)) - F(\rho(0))\big] - W[0,\tau]. \tag{11.9}$$

Equivalently, using (11.8),

$$\beta\,\mathcal{S}[\Lambda,\tau] = -\Sigma[0,\tau]. \tag{11.10}$$

Note on nomenclature. We retain the term "survival functional" to emphasize the biological interpretation of persistence far from equilibrium; mathematically, $\mathcal{S}$ is strictly a *generalized entropy-balance functional* derived from the first and second laws.

Remark 11.9 (Interpretation)**.** *The survival functional has a transparent physical meaning:*

- $\mathcal{S} > 0$*: the system gained more free energy than was invested by the external protocol—a* thermodynamic profit*. The agent has extracted usable work from environmental correlations.*
- $\mathcal{S} = 0$*: the agent breaks even (reversible limit, $\Sigma = 0$).*
- $\mathcal{S} < 0$*: the agent paid more than it gained (the generic irreversible case).*

Under the standard second law ($\Sigma \geq 0$), $\mathcal{S} \leq 0$ always. As we show in Sections 11.3 and 11.4, achieving $\mathcal{S} > 0$ requires information*—and the memory kernel provides exactly this.*

Remark 11.10 (Connection to Information Thermodynamics)**.** *In the Sagawa–Ueda framework [85, 86], a system under feedback control satisfies the generalized second law*

$$\Sigma \geq -I_{\text{feedback}}, \tag{11.11}$$

where $I_{\text{feedback}} \geq 0$ is the mutual information gained through measurement of the system. This permits $\Sigma < 0$ (and hence $\mathcal{S} > 0$) at the expense of information. The core thesis of this paper is that a non-Markovian memory kernel provides implicit *feedback: the system's history encodes correlations with the environment that play the same thermodynamic role as explicit measurement outcomes.*

11.3 The Markovian Ceiling

We now establish the fundamental thermodynamic limitation of memoryless agents. The result is elementary given the framework of Section 11.2.3, but its consequences are far-reaching: under *open-loop* control—where the agent's protocol $H_{\text{ctrl}}(t)$ is fixed in advance and receives no information from the bath—the survival functional can never be positive.

11.3.1 Spohn's Inequality

Throughout this section we assume that the GKSL generator $\mathcal{L}$ is a *thermal Lindbladian*: it is obtained from the weak-coupling (Davies) limit of a system coupled to a single thermal bath at inverse temperature β, and satisfies **quantum detailed balance** (the KMS condition) [97, 18]. Under this assumption, the unique stationary state is the Gibbs state $\rho_{\text{ss}} = \rho_{\text{eq}}$ of (11.4), and the generator is self-adjoint with respect to the KMS inner product. This ensures that the entropy production rate below is well-defined and non-negative.

Definition 11.11 (Markovian Semigroup)**.** *Throughout this paper, "Markovian" dynamics refers strictly to a* ***dynamical semigroup*** *generated by a time-independent GKSL generator $\mathcal{L}$ with non-negative rates. While time-dependent CP-divisible maps [80] are often called Markovian in broader contexts, the ceiling theorem (Theorem 11.14) targets the semigroup case $\Lambda(t) = e^{\mathcal{L}t}$, where no memory effects or temporal correlations can be exploited.*

Lemma 11.12 (Spohn [97])**.** *For any GKSL dynamical semigroup $\Lambda_t = e^{\mathcal{L}t}$ satisfying quantum detailed balance with unique invariant state ρ_{eq}, the entropy production rate*

$$\sigma(t) := -\operatorname{tr}\big(\mathcal{L}[\rho(t)]\,(\ln\rho(t) - \ln\rho_{\text{eq}})\big) \tag{11.12}$$

satisfies $\sigma(t) \geq 0$, with equality if and only if $\rho(t) = \rho_{\text{eq}}$.

Proof. This follows from the contractivity of CPTP maps under quantum relative entropy [97, 18]: $D_{\text{KL}}(\Lambda_t\rho\|\Lambda_t\rho_{\text{eq}}) \leq D_{\text{KL}}(\rho\|\rho_{\text{eq}})$ for all $t \geq 0$. Differentiating at $t = 0$ yields $\sigma(t) \geq 0$. □

11.3.2 The Markovian Ceiling Theorem

Definition 11.13 (Open-loop Markovian control class $\mathcal{C}_{\mathrm{M}}$)**.** *A protocol $H_{\mathrm{ctrl}}(t)$ belongs to the* ***open-loop Markovian control class*** $\mathcal{C}_{\mathrm{M}}$ *if and only if:*

(C1) $H_{\mathrm{ctrl}}(t)$ *is a* predetermined *function of t alone, fixed before the protocol begins.*

(C2) *No measurement of the system or environment is performed during* $[0, \tau]$*, and* $H_{\mathrm{ctrl}}(t)$ *receives no feedback from measurement outcomes.*

(C3) $H_{\mathrm{ctrl}}(t)$ *is statistically independent of the bath realization* $\{\xi_E(s) : s \in [0, \tau]\}$.

Protocols involving adaptive measurement-based feedback (Sagawa–Ueda [85]) are excluded *from* $\mathcal{C}_{\mathrm{M}}$.

Theorem 11.14 (Thermal Markovian Ceiling)**.** *Assume the GKSL generator satisfies quantum detailed balance with respect to a single thermal bath at inverse temperature β (i.e., the KMS detailed-balance condition of Lemma 11.12). Let Λ^{M} denote the resulting GKSL dynamics* (11.2)*, coupled to the stationary thermal bath, under a control protocol* $H_{\mathrm{ctrl}}(t) \in \mathcal{C}_{\mathrm{M}}$ *(Definition 11.13). Then the survival functional satisfies*

$$\mathcal{S}[\Lambda^{\mathrm{M}}, \tau] \leq 0 \qquad \textit{for all } \tau \geq 0. \tag{11.13}$$

Equality holds in the quasi-static limit ($\Sigma \to 0$), where the protocol varies slowly enough that the state remains close to the instantaneous Gibbs state $\rho_{\mathrm{eq}}(t)$ at all times.

Proof. The proof proceeds in two steps.
Step 1: Free-energy balance. Differentiating (11.6), the relative entropy evolves as

$$\frac{d}{dt} D_{\mathrm{KL}}(\rho(t) \| \rho_{\mathrm{eq}}) = \beta \dot{W}(t) - \sigma(t), \tag{11.14}$$

where $\dot{W}(t) = \mathrm{tr}(\rho(t)\, \partial_t H_{\mathrm{ctrl}})$ is the instantaneous power and $\sigma(t)$ is Spohn's entropy production rate (11.12). Integrating over $[0, \tau]$:

$$\Delta D_{\mathrm{KL}} = \beta\, W[0, \tau] - \underbrace{\int_0^\tau \sigma(t)\, dt}_{= \Sigma \ \geq \ 0}. \tag{11.15}$$

Step 2: Applying Spohn. By Lemma 11.12, $\sigma(t) \geq 0$ for all t, so $\Sigma \geq 0$. From (11.15):

$$\Delta D_{\mathrm{KL}} \leq \beta\, W[0, \tau]. \tag{11.16}$$

Converting via (11.6): $\Delta F \leq W[0, \tau]$, whence $\mathcal{S} = \Delta F - W \leq 0$.

The ceiling $\mathcal{S} = 0$ is achieved in the reversible limit where the protocol is infinitely slow and $\sigma(t) \to 0$ pointwise. □

Remark 11.15 (The "Open-Loop" Qualifier)**.** *The restriction to the control class $\mathcal{C}_{\mathrm{M}}$ (Definition 11.13) is essential. If the agent can perform* measurements *on the bath and condition its protocol on the outcomes—i.e., violate condition* (C2)*—the Sagawa–Ueda generalized second law* (11.11) *permits $\Sigma < 0$ (and hence $\mathcal{S} > 0$) at the expense of mutual information. The Markovian ceiling is therefore not a universal bound on all Markovian agents, but on agents whose protocols satisfy* (C1)–(C3).

This qualifier is precisely the point: the memory kernel of non-Markovian dynamics provides implicit access to bath correlations, playing the role of implicit measurement—the subject of Section 11.4.

Corollary 11.16 (Temporal Blindness)**.** *Under the Born–Markov approximation, the bath correlation function is replaced by its white-noise limit $C(t, s) \to C_0\,\delta(t-s)$, and the GKSL dissipator depends only on the spectral density $J(\omega)$ evaluated at the system's Bohr frequencies. The agent interacts with the environment's* power spectrum *but is structurally blind to its* temporal correlations*—the off-diagonal elements $C(t, s)$ for $t \neq s$.*

Consequently, the spectral gap $\lambda_{\min} \propto \sum_k J(\omega_k)$ of the Liouvillian sets the rate of irreversible decay. Maintaining $D_{\mathrm{KL}} > 0$ requires continuous work at rate $\dot{W} \geq \beta^{-1}\sigma(t) > 0$, and the integrated cost always meets or exceeds the integrated gain.

Remark 11.17 (Dissipative vs. Self-Nourishing Structures)**.** *The Markovian ceiling partitions far-from-equilibrium structures into two classes:*

- ***Dissipative structures*** *($\mathcal{S} \leq 0$): sustained by continuous external free-energy input. Every unit of order is paid for in full. (Schrödinger's sense [88].)*
- ***Self-nourishing structures*** *($\mathcal{S} > 0$): extract structured advantage from environmental correlations, gaining more free energy than they consume. These require information flow, and hence memory.*

The ceiling is not a limitation of the agent's control strategy but a structural consequence *of temporal blindness: without memory, the environment's temporal correlations are thermodynamically invisible.*

11.4 The Non-Markovian Advantage

Having established that open-loop Markovian agents are thermodynamically capped at $\mathcal{S} \leq 0$, we now demonstrate how non-Markovian dynamics breaks this ceiling. The mechanism is grounded entirely in standard quantities: the quantum mutual information $I(S{:}E)$ between system and environment serves as a consumable thermodynamic resource. Non-Markovian backflow intervals are precisely those during which pre-existing correlations are consumed, enabling the system to extract free energy beyond what open-loop work provides.

11.4.1 System–Environment Mutual Information

We work with the total system–environment state $\rho_{SE}(t)$, evolving unitarily under the total Hamiltonian (11.1). The quantum mutual information

$$I(S{:}E;\, t) := S(\rho_S(t)) + S(\rho_E(t)) - S(\rho_{SE}(t)) = D_{\mathrm{KL}}\big(\rho_{SE}(t) \,\big\|\, \rho_S(t) \otimes \rho_E(t)\big) \geq 0 \quad (11.17)$$

quantifies the total correlations (classical and quantum) between the system S and the environment E at time t.

Remark 11.18 (Role of Initial Correlations)**.** *Under the Born approximation, the initial state is taken as a product $\rho_{SE}(0) = \rho_S(0) \otimes \rho_E^{\mathrm{th}}$, so $I(S{:}E;\, 0) = 0$. For a system that has already been interacting with its environment (the physically generic situation for a "persistent agent"), the effective initial state at any restart time $t_0 > 0$ is* not *a product state: the preceding evolution has established correlations $I(S{:}E;\, t_0) > 0$. These pre-existing correlations—the system's "memory" of past interactions—are the thermodynamic resource that the memory kernel can exploit.*

11.4.2 The Information–Thermodynamic Identity

The following identity is the central technical tool of this section. It holds for **any** initial state—product or correlated—and relies only on unitarity and the definitions of mutual information and relative entropy.

Remark 11.19 (Relative-entropy chain rule)**.** *We repeatedly use the identity*

$$D_{\mathrm{KL}}\big(\rho_{SE} \,\big\|\, \rho_S \otimes \sigma_E\big) = I(S{:}E)_{\rho_{SE}} + D_{\mathrm{KL}}(\rho_E \| \sigma_E), \tag{11.18}$$

valid for arbitrary *(possibly correlated)* ρ_{SE} *and any full-rank reference state* σ_E.[1] *Importantly, this is a* pure algebraic identity *and does not assume product initial conditions.*

Lemma 11.20 (Information–Thermodynamic Identity)**.** *Let* $\rho_{SE}(t)$ *evolve unitarily under the total Hamiltonian. Then, for* ***any*** *initial state* $\rho_{SE}(0)$ *(product or correlated):*

$$\Delta I(S{:}E) + \Delta D_{\mathrm{KL}}\big(\rho_E \,\big\|\, \rho_E^{\mathrm{th}}\big) = \Delta S_S + \beta\, \Delta \langle H_E \rangle, \tag{11.19}$$

where $\Delta S_S = S(\rho_S(\tau)) - S(\rho_S(0))$ *is the change in the system's von Neumann entropy and* $\Delta\langle H_E\rangle = \mathrm{Tr}[\rho_E(\tau)\, H_E] - \mathrm{Tr}[\rho_E(0)\, H_E]$ *is the energy absorbed by the environment.*

Proof. Applying the chain rule (11.18) (Remark 11.19) with $\sigma_E = \rho_E^{\mathrm{th}}$:

$$D_{\mathrm{KL}}\big(\rho_{SE}(t) \,\big\|\, \rho_S(t) \otimes \rho_E^{\mathrm{th}}\big) = I(S{:}E;\, t) + D_{\mathrm{KL}}\big(\rho_E(t) \,\big\|\, \rho_E^{\mathrm{th}}\big). \tag{11.20}$$

Expanding the left side using $\ln \rho_E^{\mathrm{th}} = -\beta H_E - \ln Z_E$:

$$D_{\mathrm{KL}}\big(\rho_{SE}(t) \,\big\|\, \rho_S(t) \otimes \rho_E^{\mathrm{th}}\big) = -S(\rho_{SE}(t)) + S(\rho_S(t)) + \beta \langle H_E \rangle_t + \ln Z_E. \tag{11.21}$$

Since the total evolution is unitary, $S(\rho_{SE}(t)) = S(\rho_{SE}(0))$ for all t. Taking the difference between times τ and 0 cancels both $S(\rho_{SE})$ and $\ln Z_E$, yielding

$$\Delta\big[I(S{:}E) + D_{\mathrm{KL}}(\rho_E \| \rho_E^{\mathrm{th}})\big] = \Delta S_S + \beta\, \Delta \langle H_E \rangle. \tag{11.22}$$

□

Remark 11.21 (No assumption on the initial state)**.** *The proof of Lemma 11.20 uses* only *unitarity (*$\Delta S(\rho_{SE}) = 0$*) and the algebraic structure of the KL divergence. No assumption is made about the initial state* $\rho_{SE}(0)$*, the coupling strength, or the character (Markovian or non-Markovian) of the reduced dynamics. When the initial state is a product state with the environment in thermal equilibrium, all initial-time terms vanish and the identity reduces to the Esposito decomposition [31]:* $\Sigma = I(S{:}E;\, \tau) + D_{\mathrm{KL}}(\rho_E(\tau) \| \rho_E^{\mathrm{th}})$.

11.4.3 The Survival Identity

We now connect the information–thermodynamic identity (11.19) to the survival functional $\mathcal{S}$ defined in Section 11.2.4.

[1] This follows from the definition of quantum relative entropy and $\ln(\rho_S \otimes \sigma_E) = \ln \rho_S \otimes \nVdash_E + \nVdash_S \otimes \ln \sigma_E$. See, e.g., M. M. Wilde, *Quantum Information Theory*, 2nd ed., Cambridge University Press (2017), Sec. 11; and M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (2000), Ch. 11.

Theorem 11.22 (Survival Functional: General Form)**.** *Under Assumptions* (A1)–(A5) *of Definition 11.6, let $\rho_{SE}(t)$ evolve unitarily from an* arbitrary *(possibly correlated) initial state $\rho_{SE}(0)$. The survival functional satisfies*

$$\boxed{\beta\,\mathcal{S}[\Lambda,\tau] = -\Delta I(S{:}E) - \Delta D_{\mathrm{KL}}\big(\rho_E \,\big\|\, \rho_E^{\mathrm{th}}\big) - \beta\,\Delta\langle H_{\mathrm{ctrl}}\rangle,} \tag{11.23}$$

where $\Delta\langle H_{\mathrm{ctrl}}\rangle = \mathrm{Tr}[\rho_S(\tau)\,H_{\mathrm{ctrl}}(\tau)] - \mathrm{Tr}[\rho_S(0)\,H_{\mathrm{ctrl}}(0)]$ is the change in the control-field energy.

For ***autonomous evolution*** *($H_{\mathrm{ctrl}} = 0$ throughout $[0,\tau]$), the control term vanishes:*

$$\beta\,\mathcal{S}[\Lambda,\tau] = -\Delta I(S{:}E) - \Delta D_{\mathrm{KL}}\big(\rho_E \,\big\|\, \rho_E^{\mathrm{th}}\big). \tag{11.24}$$

Proof. The proof uses three ingredients: the definition of $\mathcal{S}$, the first law, and Lemma 11.20.

Step 1 (First law in weak coupling). Since $H_{\mathrm{ctrl}}(t)$ is the only time-dependent component of H, the work satisfies $W = \Delta\langle H\rangle \approx \Delta\langle H_R\rangle + \Delta\langle H_{\mathrm{ctrl}}\rangle + \Delta\langle H_E\rangle$ by Assumption (A2).

Step 2 (Connecting Σ to the identity). From Definition 11.8 and (11.8), using $\Delta F = \Delta\langle H_R\rangle - \beta^{-1}\Delta S_S$:

$$\begin{aligned}\Sigma = \beta(W - \Delta F) &= \beta\big(W - \Delta\langle H_R\rangle\big) + \Delta S_S \\ &= \beta\big(\Delta\langle H_{\mathrm{ctrl}}\rangle + \Delta\langle H_E\rangle\big) + \Delta S_S \\ &= \big(\Delta S_S + \beta\,\Delta\langle H_E\rangle\big) + \beta\,\Delta\langle H_{\mathrm{ctrl}}\rangle.\end{aligned} \tag{11.25}$$

By Lemma 11.20, the parenthesized term equals $\Delta I(S{:}E) + \Delta D_{\mathrm{KL}}(\rho_E \| \rho_E^{\mathrm{th}})$. Hence

$$\Sigma = \Delta I(S{:}E) + \Delta D_{\mathrm{KL}}\big(\rho_E \,\big\|\, \rho_E^{\mathrm{th}}\big) + \beta\,\Delta\langle H_{\mathrm{ctrl}}\rangle. \tag{11.26}$$

Step 3 (Survival functional). $\beta\,\mathcal{S} = -\Sigma$ by (11.10), yielding (11.23). For $H_{\mathrm{ctrl}} = 0$: $\Delta\langle H_{\mathrm{ctrl}}\rangle = 0$, recovering (11.24). □

Remark 11.23 (Weak-coupling caveat)**.** *The identity* (11.23) *holds exactly under the weak-coupling approximation ($\lambda \to 0$); for finite coupling, $O(\lambda)$ corrections from the interaction energy H_{int} should be tracked. For autonomous evolution ($H_{\mathrm{ctrl}} = 0$), the identity* (11.24) *is exact.*

Remark 11.24 (Nature of the result)**.** *Equation* (11.23) *is an exact* accounting identity, *not an inequality or optimality bound. It establishes that any thermodynamic profit ($\mathcal{S} > 0$) in the autonomous regime must be perfectly balanced by the consumption of system–environment correlations ($\Delta I < 0$) or the relaxation of the bath ($\Delta D_{\mathrm{KL}} < 0$). The "non-Markovian advantage" arises because memory kernels allow access to regimes where $\Delta I(S{:}E)$ is negative and dominant—a channel that memoryless (Born–Markov) dynamics resets to zero at every time step (Remark 11.29).*

Remark 11.25 (Scope of the theorem)**.** *Theorem 11.22 holds for* any *initial state $\rho_{SE}(0)$—product or correlated. The proof requires only Assumptions* (A1)–(A5) *of Definition 11.6 and the definitions of $\mathcal{S}$, $I(S{:}E)$, and D_{KL}. No assumption about the reduced dynamics (Markovian, non-Markovian, or otherwise) is needed. This generality is essential: a persistent agent that has already been interacting with its environment necessarily carries correlations ($I(S{:}E;\,0) > 0$), and it is precisely these correlations that constitute the thermodynamic resource for survival.*

11.4.4 Three Regimes of Survival

We specialize to the autonomous case ($H_{\text{ctrl}} = 0$), which is the natural setting for the "memory as a resource" argument: the agent benefits from pre-existing correlations without external driving.

Corollary 11.26 (Three Regimes)**.** *Under autonomous evolution, identity* (11.24) *identifies three regimes:*

1. ***Product initial state*** *($I(S{:}E;\,0) = 0$, $D_{\text{KL}}(\rho_E(0)\|\rho_E^{\text{th}}) = 0$): Both ΔI and ΔD_{KL} are increases from zero to non-negative final values, so*

$$\beta\,\mathcal{S} = -\big(I(S{:}E;\,\tau) + D_{\text{KL}}(\rho_E(\tau)\|\rho_E^{\text{th}})\big) \le 0.$$

This recovers the Markovian ceiling (Theorem 11.14), now with a precise accounting of where *the entropy goes: into system–environment correlations and bath displacement.*

2. ***Correlated initial state*** *($I(S{:}E;\,0) > 0$): If the dynamics* consumes *pre-existing correlations ($\Delta I < 0$, i.e., $I(S{:}E;\,\tau) < I(S{:}E;\,0)$), the first term contributes* positively *to $\mathcal{S}$. Provided*

$$|\Delta I(S{:}E)| > \Delta D_{\text{KL}}\big(\rho_E \,\big\|\, \rho_E^{\text{th}}\big), \tag{11.27}$$

the survival functional is strictly positive: $\mathcal{S} > 0$. The agent has converted pre-existing correlations into usable free energy.

3. ***Upper bound:*** *Since $I(S{:}E;\,\tau) \ge 0$ and $D_{\text{KL}}(\rho_E(\tau)\|\rho_E^{\text{th}}) \ge 0$, the maximum survival gain is bounded by*

$$\beta\,\mathcal{S} \le I(S{:}E;\,0) + D_{\text{KL}}\big(\rho_E(0) \,\big\|\, \rho_E^{\text{th}}\big). \tag{11.28}$$

The thermodynamic profit cannot exceed the total initial "resource budget"—the pre-existing correlations plus the initial displacement of the bath from equilibrium.

11.4.5 The Correlation Battery

The three regimes of Corollary 11.26 raise a natural question: *where do the initial correlations $I(S{:}E;\,0) > 0$ come from?*

Remark 11.27 (The Correlation Battery)**.** *The answer is: from **prior interaction epochs**. A persistent agent does not begin its existence in a product state. Over any interaction interval, unitary evolution generically builds system–environment correlations ($\Delta I > 0$), at a thermodynamic cost ($\mathcal{S} < 0$ during this phase by Corollary 11.26(i)). The non-Markovian agent's advantage is that these correlations* persist *and can be consumed during later intervals ($\Delta I < 0$, $\mathcal{S} > 0$).*

*The process is analogous to a **battery**:*

- ***Charging phase*** *(correlation building, $\Delta I > 0$): the agent "pays" free energy to build system–environment correlations. $\mathcal{S} < 0$.*
- ***Discharging phase*** *(correlation consumption, $\Delta I < 0$): the agent extracts free energy from the stored correlations. $\mathcal{S} > 0$.*

A Markovian agent cannot operate this battery. The Born approximation resets $I(S{:}E) = 0$ *at every infinitesimal time step, destroying the stored correlations before they can be used. The semigroup property* $\Lambda(t+s) = \Lambda(t)\Lambda(s)$ *is precisely the statement that no inter-epoch correlations survive. The memory kernel* $\mathcal{K}(t, s)$ *is what allows the non-Markovian agent to carry charge across epochs.*

Crucially, global thermodynamics remains respected. For any full cycle starting from an uncorrelated thermal state ($I(S{:}E;\, 0) = 0$, $D_{\mathrm{KL}}(\rho_E(0)\|\rho_E^{\mathrm{th}}) = 0$)*, the total survival functional satisfies*

$$\beta\,\mathcal{S}[0,t] = -\Sigma[0,t] \;\leq\; 0 \qquad \textit{(second law)}. \tag{11.29}$$

The local positivity $\mathcal{S}[t^*, t] > 0$ *during the discharging phase is strictly funded by the free energy dissipated during the earlier charging phase (see Proposition 11.28 for the formal decomposition).*

Proposition 11.28 (Full-cycle closure)**.** *Under the conditions of Theorem* 11.22 *with autonomous evolution* ($H_{\mathrm{ctrl}} = 0$)*, partition* $[0, \tau]$ *at any intermediate time* t^* *into a charging phase* $[0, t^*]$ *and a discharging phase* $[t^*, \tau]$*.*

(i) ***Charging*** *(product initial state,* $I(S{:}E;\, 0) = 0$*). By Corollary* 11.26(i)*,*

$$\beta\,\mathcal{S}[0,t^*] = -I(S{:}E;\, t^*) - D_{\mathrm{KL}}\big(\rho_E(t^*)\,\big\|\,\rho_E^{\mathrm{th}}\big) \;\leq\; 0. \tag{11.30}$$

(ii) ***Discharging*** *(correlated initial state at* t^**). Applying* (11.24) *to* $[t^*, \tau]$*:*

$$\beta\,\mathcal{S}[t^*,\tau] = -\big(I(S{:}E;\, \tau) - I(S{:}E;\, t^*)\big) - \big(D_{\mathrm{KL}}(\rho_E(\tau)\|\rho_E^{\mathrm{th}}) - D_{\mathrm{KL}}(\rho_E(t^*)\|\rho_E^{\mathrm{th}})\big), \tag{11.31}$$

which is positive whenever the decrease in correlations dominates the change in bath displacement (Corollary 11.26(ii)*).*

(iii) ***Full cycle.*** *Since* $\mathcal{S}$ *is additive over concatenated intervals,* $\beta\,\mathcal{S}[0,\tau] = \beta\,\mathcal{S}[0,t^*] + \beta\,\mathcal{S}[t^*,\tau]$*. Equivalently, applying* (11.24) *directly to* $[0, \tau]$ *with* $I(S{:}E;\, 0) = 0$*:*

$$\beta\,\mathcal{S}[0,\tau] = -I(S{:}E;\, \tau) - D_{\mathrm{KL}}\big(\rho_E(\tau)\,\big\|\,\rho_E^{\mathrm{th}}\big) \;\leq\; 0. \tag{11.32}$$

The net thermodynamic profit over the full cycle is non-positive—the "interest" paid during charging meets or exceeds the "dividend" collected during discharging. But the local *positivity of* $\mathcal{S}$ *during discharge* (11.31) *is what enables the agent to survive through intervals that would kill a memoryless system.*

Proof. The survival functional is additive over concatenated intervals:

$$\mathcal{S}[0,\tau] = \underbrace{(\Delta F[0,t^*] - W[0,t^*])}_{\mathcal{S}[0,t^*]} + \underbrace{(\Delta F[t^*,\tau] - W[t^*,\tau])}_{\mathcal{S}[t^*,\tau]},$$

since both ΔF and W decompose additively. Items (i) and (iii) then follow from Theorem 11.22 (autonomous case) applied to $[0, t^*]$ and $[0, \tau]$ respectively, each starting from a product state. Item (ii) follows from Theorem 11.22 applied to $[t^*, \tau]$ with correlated initial state $\rho_{SE}(t^*)$. Inequality (11.32) holds because $I(S{:}E;\, \tau) \geq 0$ and $D_{\mathrm{KL}}(\rho_E(\tau)\|\rho_E^{\mathrm{th}}) \geq 0$. $\square$

11.4.6 Connection to Non-Markovianity Measures

Remark 11.29 (The Born Approximation Destroys the Resource)**.** *Under the Born (product-state) approximation, every infinitesimal time step begins from $\rho_{SE} \approx \rho_S \otimes \rho_E^{\mathrm{th}}$, enforcing $I(S{:}E) = 0$ at all times. Corollary 11.26(i) then guarantees $\mathcal{S} \leq 0$ for* every *finite interval. The Born approximation does not merely simplify the dynamics—it* eliminates the thermodynamic resource *(system–environment correlations) that would otherwise be available.*

Remark 11.30 (Connection to BLP Non-Markovianity)**.** *The Breuer–Laine–Piilo (BLP) measure of non-Markovianity [17] is defined via the temporary increase of trace distance between pairs of initial states: $\mathcal{N}_{\mathrm{BLP}} := \max_{\rho_{1,2}} \int_{\dot{D}>0} \frac{d}{dt} D(\rho_1(t), \rho_2(t))\, dt$. The intervals where trace distance increases are precisely the "discharging" intervals of Remark 11.27 [80]: correlations previously deposited in the bath flow back to the system, restoring distinguishability. The BLP measure thus witnesses the thermodynamic resource that drives $\mathcal{S} > 0$ in Corollary 11.26(ii).*

Remark 11.31 (Consistency with the Sagawa–Ueda Framework)**.** *In the Sagawa–Ueda framework [85, 86], measurement-based feedback permits $\Sigma \geq -I_{\mathrm{feedback}}$, where I_{feedback} is the mutual information gained through measurement. The memory kernel plays an analogous role: the pre-existing correlations $I(S{:}E; 0)$ are the non-Markovian analogue of I_{feedback}. The total system (agent + bath) still satisfies $\Sigma_{\mathrm{total}} \geq 0$; the apparent "profit" for the agent is paid for by the correlations consumed from the system–environment entanglement. The bound* (11.28) *is the non-Markovian analogue of the Sagawa–Ueda bound $\beta\, \mathcal{S} \leq I_{\mathrm{feedback}}$.*

11.4.7 Mechanism: The Surfer Analogy

The physical mechanism admits an intuitive picture.

- **The Markovian Agent (The Stone):** A stone thrown into the ocean sinks. It interacts with the water only at the instant of contact, dissipates its kinetic energy, and thermalizes ($\mathcal{S} \leq 0$). Each collision builds system–environment correlations that are immediately discarded (Born approximation), so $I(S{:}E) = 0$ at all times. The wave structure is invisible to it.
- **The Non-Markovian Agent (The Surfer):** A surfer carries *memory* of past wave patterns—encoded in the correlations $I(S{:}E; t_0) > 0$ built up over previous interactions (the "charging phase" of Remark 11.27). During backflow intervals ($\Delta I < 0$), the surfer *spends* these stored correlations to extract free energy from the wave itself. The surfer remains far from equilibrium not by fighting the environment, but by converting temporal correlations into thermodynamic profit.

Remark 11.32 (Thermodynamic Rectification)**.** *The "surfing" mechanism is **thermodynamic rectification**: the memory kernel $\mathcal{K}(t, s)$ functions as a temporal filter that enables the system to accumulate correlations during one phase and consume them during another. Formally, the kernel enables access to the resource $I(S{:}E; 0)$ accumulated during previous interaction epochs—converting the environment's temporal correlations into the system's structural persistence via the ΔI term in Theorem 11.22.*

Remark 11.33 (Memory as Implicit Maxwell's Demon)**.** *The memory kernel functions as an* implicit Maxwell's demon*. A Markovian system interacts with each environmental fluctuation exactly once, at the moment of contact; the Born approximation resets* $I(S{:}E) = 0$ *after each step. A non-Markovian system retains a trace of past fluctuations (via* $\mathcal{K}(t, s)$ *with* $s < t$*) and can exploit correlations between past and present environmental states. This is not a violation of the second law but an instance of the Sagawa–Ueda generalization: the demon's cost is paid in the currency of memory maintenance (Landauer erasure), a point we quantify in Section 11.7. The total budget for "demonic profit" is capped by the bound* (11.28)*.*

11.5 Emergent Temporal Arrow

We have shown that survival requires memory. This requirement yields a corollary: the emergence of a thermodynamic arrow of time. In this framework, time is not an external parameter; rather, *the direction of time is the direction of memory accumulation.*

We formalize this by defining a dynamical partial order induced by the memory kernel and connecting it to the algebraic accessibility structure of HAFF Paper F [130].

11.5.1 The Causal Memory Order

A non-Markovian memory kernel $\mathcal{K}(t, s)$ defines a causal link between a past state at s and the present dynamics at t. We define a partial order based on the effective support of this influence.

Definition 11.34 (Causal Memory Order)**.** *Let* $\mathcal{T} = \{\rho(t) \mid t \in \mathbb{R}^+\}$ *be a state trajectory. We define the binary relation* $\prec_K$ *on* $\mathcal{T}$ *by*

$$\rho(s) \prec_K \rho(t) \quad \Longleftrightarrow \quad \exists\, \tau \in [s, t] \text{ such that } \|\mathcal{K}(t, \tau)[\rho(s)]\| > \epsilon, \tag{11.33}$$

where $\epsilon > 0$ *is a physical distinguishability threshold set by the thermal noise floor* $\epsilon \sim e^{-\beta\,\Delta E_{\min}}$*. Physically,* $\rho(s) \prec_K \rho(t)$ *means "the dynamics at* t *retains operationally distinguishable information about the state at* s*."*

For a Markovian agent, $\mathcal{K}(t, s) \propto \delta(t - s)$, so $\rho(s) \not\prec_K \rho(t)$ for any $s < t$. The Markovian agent has no dynamical past—it lives in an eternal "now." A non-Markovian agent carries its history within its dynamics; the depth of the order $\prec_K$ is set by the memory time τ_{mem} (Definition 11.2).

11.5.2 Unidirectionality from Survival Optimization

Why does the order $\prec_K$ point "forward"? While the microscopic laws are time-reversible, the *survival imperative* (maximizing $\mathcal{S}$) creates a statistical irreversibility.

Conjecture 11.35 (Fisher Information Accretion)**.** *Let* $\mathcal{I}_F(\theta; \rho(t))$ *denote the Fisher information contained in the system state* $\rho(t)$ *regarding a parameter* θ *encoded in the environment at time* $s < t$*. For an agent whose dynamics maximize the survival functional* (11.9)*, the time-averaged Fisher information satisfies*

$$\overline{\frac{d}{dt}\,\mathcal{I}_F(\theta; \rho(t))} \geq 0, \tag{11.34}$$

where the overbar denotes a time average over scales larger than the bath correlation time τ_B*.*

Heuristic argument. By Theorem 11.22, the survival functional is maximized when $I(S{:}E;\,0)$ is large and can be consumed ($\Delta I < 0$) during subsequent evolution. Maintaining a large correlation budget $I(S{:}E)$ requires the system state to retain correlations with environmental degrees of freedom; this is precisely the content of $\mathcal{I}_F(\theta;\,\rho(t)) > 0$. An agent that discards useful correlations (decreasing $\mathcal{I}_F$) without thermodynamic necessity depletes the resource $I(S{:}E)$ and hence its survival functional. Since the environment's correlations decay on a timescale τ_B, the agent must continuously build new correlations to replace decaying ones. The net effect is a time-averaged accretion of Fisher information, whose gradient defines the dynamical arrow of time. □

We emphasize that this argument is informal: no formal connection between the Fisher information $I_F(\theta;\rho(t))$ (about an environmental parameter) and the quantum mutual information $I(S{:}E)$ (between system and environment) has been established. A rigorous derivation via the quantum Cramér–Rao bound remains an open problem.

11.5.3 The Bridge to HAFF

We now connect this dynamical picture to the algebraic picture of HAFF Paper F [130], where the arrow of time was defined by the expansion of the redundancy subalgebra $\mathcal{R}$.

The connection requires care: quantum information cannot be cloned (the no-cloning theorem), so the "redundancy expansion" of HAFF must be interpreted through the lens of *quantum Darwinism* [122]. In this framework, the environment acquires not copies of the quantum state $\rho(s)$ itself, but rather *coarse-grained classical records* of pointer-state outcomes—precisely the information that survives decoherence and can be redundantly encoded in many environmental fragments.

Proposition 11.36 (Dynamical–Algebraic Correspondence)**.** *Let $\prec_K$ be the causal memory order (Definition 11.34) and let $\prec_{\mathrm{HAFF}}$ be the accessibility order of HAFF Paper F, defined by the inclusion of redundancy subalgebras $\mathcal{R}$. Under the additional assumption that the system–environment interaction produces decoherence in a preferred pointer basis [122], there exists a coarse-graining map $\Phi : \rho(t) \mapsto \hat{p}(t)$ (projecting onto the diagonal in the pointer basis) such that:*

$$\rho(s) \prec_K \rho(t) \quad \Longrightarrow \quad \mathcal{R}(\Phi[\rho(s)]) \subseteq \mathcal{R}(\Phi[\rho(t)]). \tag{11.35}$$

That is, the dynamical partial order maps into the algebraic accessibility order when restricted to the classical sector selected by decoherence.

Proof. The argument has three steps.
Step 1 (Dynamical side): $\rho(s) \prec_K \rho(t)$ implies that the memory kernel $\mathcal{K}$ transduces information about the state at s into the dynamics at t, via system–environment correlations built up over $[s,t]$.
Step 2 (Quantum Darwinism): The system–environment interaction selects pointer states $\{|i\rangle\}$ that are robust under decoherence [122]. The diagonal populations $p_i(t) = \langle i|\rho(t)|i\rangle$ constitute *classical* information. Quantum Darwinism [122] establishes that this classical information—and *only* this information—is redundantly imprinted in many environmental fragments E_k through the decoherence interaction. Each fragment that

acquires a record of $\hat{p}(t) = \{p_i(t)\}$ contributes to the growth of the redundancy subalgebra $\mathcal{R}$. Crucially, no quantum cloning is involved: the no-cloning theorem forbids copying of arbitrary quantum states, but does not constrain the classical pointer-state probabilities, which are freely duplicable. The expansion of $\mathcal{R}$ reflects the proliferation of these classical records, not the copying of quantum coherences.

Step 3 (Correspondence): The coarse-graining map Φ projects onto the *commuting* subalgebra generated by the pointer observables $\{|i\rangle\langle i|\}$. The resulting probability distributions $\hat{p}(t)$ are classical and lie in a simplex. If $\rho(s) \prec_K \rho(t)$, then the dynamics at t retains information about the state at s (Definition 11.34); in the pointer basis, this means $\hat{p}(s)$ is statistically reconstructible from the environmental records available at t. Since each environmental fragment carrying a record of $\hat{p}$ contributes to the HAFF redundancy subalgebra $\mathcal{R}$, and the number of such fragments grows monotonically with the accumulation of decoherence records. This monotonicity holds on timescales coarse-grained over the bath correlation time; during non-Markovian backflow intervals, the number of "good" fragments could temporarily decrease. The inclusion $\mathcal{R}(\Phi[\rho(s)]) \subseteq \mathcal{R}(\Phi[\rho(t)])$ follows. □

Remark 11.37 (Scope of the Correspondence)**.** *Proposition 11.36 is a* consistency result*, not a derivation of HAFF from T-DOME or vice versa. It shows that the dynamical arrow (memory accumulation) and the algebraic arrow (redundancy expansion) are compatible when restricted to the decoherence-selected classical sector. The quantum coherences—which are not redundantly recorded—lie outside this correspondence and are handled by the full non-Markovian dynamics.*

Remark 11.38 (Dynamical and Algebraic Time)**.** *The correspondence links two independently motivated notions of temporal direction:*

	Paper F (HAFF)	***Paper I (T-DOME)***
Nature	*Algebraic*	*Dynamical*
Mechanism	*Redundancy expansion*	*Information backflow from memory*
Formalism	*Partial order on* $\mathcal{A}_\mathbf{c}$	*Partial order* $\prec_K$ *on* $\rho_R(t)$
Asymmetry source	*Phase-space measure*	*Bath correlation structure*
Domain	*Classical (pointer) sector*	*Full quantum dynamics*

Paper F provides the structural skeleton *of temporal asymmetry; Paper I provides the dynamical* muscle*.*

Remark 11.39 (The Seed and the Tree)**.** *The correspondence justifies the title of this paper. In HAFF, the geometry of spacetime is the static "tree." In T-DOME, the memory kernel is the "seed" containing the generative algorithm for growth. Time is not the space in which the tree grows; time is the* act of growing *itself.*

11.6 Worked Example: The Quantum Predictive Agent

To illustrate the Markovian ceiling and the memory advantage *quantitatively*, we employ the archetypal open quantum system model: the spin-boson model with Lorentz–Drude spectral density, which admits an exact analytic solution for the decoherence dynamics [18].

11.6.1 Model Setup

The total Hamiltonian is $H = H_S + H_B + H_I$. The agent is a two-level system with energy gap ω_0: $H_S = \frac{\omega_0}{2}\sigma_z$. The environment is a bosonic bath: $H_B = \sum_k \omega_k b_k^\dagger b_k$. The interaction is of the pure-dephasing form $H_I = \sigma_z \otimes \sum_k (g_k\, b_k + g_k^*\, b_k^\dagger)$.

The spectral density $J(\omega) = \sum_k |g_k|^2 \delta(\omega - \omega_k)$ characterizes the environment. We choose the Lorentz–Drude form:

$$J(\omega) = \frac{2\lambda\,\gamma\,\omega}{\omega^2 + \gamma^2}, \tag{11.36}$$

where λ is the reorganization energy and γ is the bath memory rate (inverse correlation time $\tau_B = 1/\gamma$). We place the system in the low-temperature regime $\beta\omega_0 \gg 1$ (i.e., $k_B T \ll \omega_0$). The bath correlation function in the $T \to 0$ limit is $C(t) = \lambda\gamma\, e^{-\gamma|t|}$, so the parameter γ directly controls the bath memory depth. For $\beta\omega_0 \geq 10$ the finite-temperature corrections to all quantities below are of order $O(e^{-\beta\omega_0}) \lesssim 5 \times 10^{-5}$ and are neglected throughout.[2]

11.6.2 Exact Decoherence Function

For the pure-dephasing spin-boson model in the $T \to 0$ limit, the off-diagonal element of the reduced density matrix $\rho_{01}(t) = \rho_{01}(0)\, p(t)$ is governed by the **decoherence function** [18]:

$$p(t) = e^{-\gamma t/2}\left[\cos(\Omega t) + \frac{\gamma}{2\Omega}\sin(\Omega t)\right], \tag{11.37}$$

where $\Omega := \frac{1}{2}\sqrt{4\lambda\gamma - \gamma^2}$. This solution is exact for the Lorentz–Drude spectral density.

Remark 11.40 (Non-Markovian Regime)**.** *The character of the dynamics is controlled by the discriminant* $\Delta := 4\lambda\gamma - \gamma^2 = \gamma(4\lambda - \gamma)$*:*

- $\gamma > 4\lambda$ *(*$\Delta < 0$*):* Ω *is imaginary,* $p(t)$ *decays monotonically. The dynamics is Markovian (no backflow).*
- $\gamma = 4\lambda$ *(*$\Delta = 0$*): Critical damping.* $p(t) = (1 + \gamma t/2)\, e^{-\gamma t/2}$.
- $\gamma < 4\lambda$ *(*$\Delta > 0$*):* Ω *is real and positive.* $p(t)$ *oscillates with envelope* $e^{-\gamma t/2}$. *The dynamics is* non-Markovian*: intervals where* $|p(t)|$ *increases correspond to information backflow [17].*

The non-Markovian regime $\gamma < 4\lambda$ *is thus the regime of structured, long-memory baths.*

11.6.3 Quantitative Evaluation

We now evaluate the survival functional explicitly. For the pure-dephasing model, populations are conserved ($p_0(t) = p_0(0)$, $p_1(t) = p_1(0)$), and the non-equilibrium free energy depends only on the coherence:

$$F(\rho(t)) - F(\rho_{\text{eq}}) = \beta^{-1}\, D_{\text{KL}}(\rho(t) \| \rho_{\text{eq}}). \tag{11.38}$$

[2] All plots and numerical values use the standard $T \to 0$ analytic expression for the decoherence function (11.37) (see Breuer and Petruccione [18], Sec. 12.3, for the Lorentz–Drude pure-dephasing solution), which provides an accurate proxy in the low-temperature regime; β is a well-defined bookkeeping parameter and β^{-1} a finite energy scale.

For a qubit with initial state $\rho(0) = \frac{1}{2}(\mathbb{1} + \vec{r} \cdot \vec{\sigma})$ and $r_z = 0$ (maximal coherence in the x–y plane), the relative entropy reduces to $D_{\mathrm{KL}}(\rho(t)\|\rho_{\mathrm{eq}}) \approx |p(t)|^2\,|\rho_{01}(0)|^2$ exactly for the chosen initial state ($r_z = 0$, equal populations), where the population contribution to D_{KL} is constant and cancels in ΔD_{KL} (see, e.g., [18]). Since there is no external driving ($H_{\mathrm{ctrl}} = 0$, $W = 0$), the survival functional is simply

$$\beta\,\mathcal{S}(t) = D_{\mathrm{KL}}(\rho(t)\|\rho_{\mathrm{eq}}) - D_{\mathrm{KL}}(\rho(0)\|\rho_{\mathrm{eq}}) \propto |p(t)|^2 - 1. \tag{11.39}$$

The proportionality in (11.39) is specific to the **pure-dephasing model** with the chosen maximally coherent initial state ($r_z = 0$) and measurement in the pointer basis (σ_z). Under these conditions, the exact solution [18] ensures that population terms vanish from the free energy ($\Delta\langle H_S\rangle = 0$), leaving only the coherence contribution: $\beta\,\mathcal{S} = -\Delta S_S$ depends only on the coherence trajectory $|p(t)|$. The proxy $|p(t)|^2$ thus rigorously captures the sign and monotone behaviour of $\beta\,\mathcal{S}$; the exact numerical prefactor depends on the initial state and on β, but the qualitative conclusion—$\mathcal{S} > 0$ during backflow intervals—is robust and does not depend on the proxy normalization.

For a Markovian evolution, $|p(t)|$ decreases monotonically, so $|p(t)|^2 - 1 \le 0$ for all t: $\mathcal{S} \le 0$ always (consistent with Theorem 11.14). For non-Markovian evolution with $\gamma < 4\lambda$, the oscillations in $p(t)$ produce intervals where $|p(t)|$ *increases after a previous decrease*, i.e., the system *re-coheres.*

Concrete parameters. Set $\omega_0 = 1$ (energy units), $\lambda = 1$, $\gamma = 0.5$ (deep non-Markovian regime: $\gamma/4\lambda = 0.125 \ll 1$). Then:

$$\Omega = \tfrac{1}{2}\sqrt{4 \cdot 1 \cdot 0.5 - 0.25} = \tfrac{1}{2}\sqrt{1.75} \approx 0.661. \tag{11.40}$$

The decoherence function (11.37) first reaches $p(t^*) = 0$ at $t^* \approx 2.00/\Omega \approx 3.03$ (in units of ω_0^{-1}), where the system has fully decohered. Subsequently, the environment *returns* coherence: $|p(t)|$ increases, reaching a local maximum $|p(t_1)| \approx 0.31$ at $t_1 \approx 4.75/\omega_0$.

Over the backflow interval $[t^*, t_1]$, for the pure-dephasing qubit with the chosen initial state ($r_z = 0$, maximal coherence) and in the autonomous setting ($H_{\mathrm{ctrl}} = 0$, $W = 0$), the survival proxy (11.39) gives

$$\beta\,\mathcal{S}[t^*, t_1] \propto |p(t_1)|^2 - |p(t^*)|^2 \approx 0.093 - 0 = 0.093 > 0. \tag{11.41}$$

Equivalently, $\mathcal{S} \approx 0.093\,\beta^{-1}$ in the bookkeeping units set by β. The agent has gained a dimensionless survival advantage $\beta\,\mathcal{S} \approx +0.093$ *with zero work input* (autonomous evolution, $H_{\mathrm{ctrl}} = 0$), solely by exploiting the non-Markovian backflow. Figure 11.1 illustrates the contrast between Markovian and non-Markovian evolution.

Consistency with Theorem 11.22 and Proposition 11.28. Since this is autonomous evolution ($H_{\mathrm{ctrl}} = 0$), identity (11.24) applies exactly: $\beta\,\mathcal{S} = -\Delta I(S{:}E) - \Delta D_{\mathrm{KL}}(\rho_E\|\rho_E^{\mathrm{th}})$. The example realizes the *correlation battery* of Remark 11.27, with the charge–discharge decomposition of Proposition 11.28:

- **Charging** ($[0, t^*]$, eq. (11.30)): the system decoheres, building correlations $I(S{:}E;\, t^*) > 0$ at the cost of $\mathcal{S} < 0$.
- **Discharging** ($[t^*, t_1]$, eq. (11.31)): the correlations are consumed ($\Delta I < 0$ over this interval), returning $\beta\,\mathcal{S} \approx +0.093 > 0$.

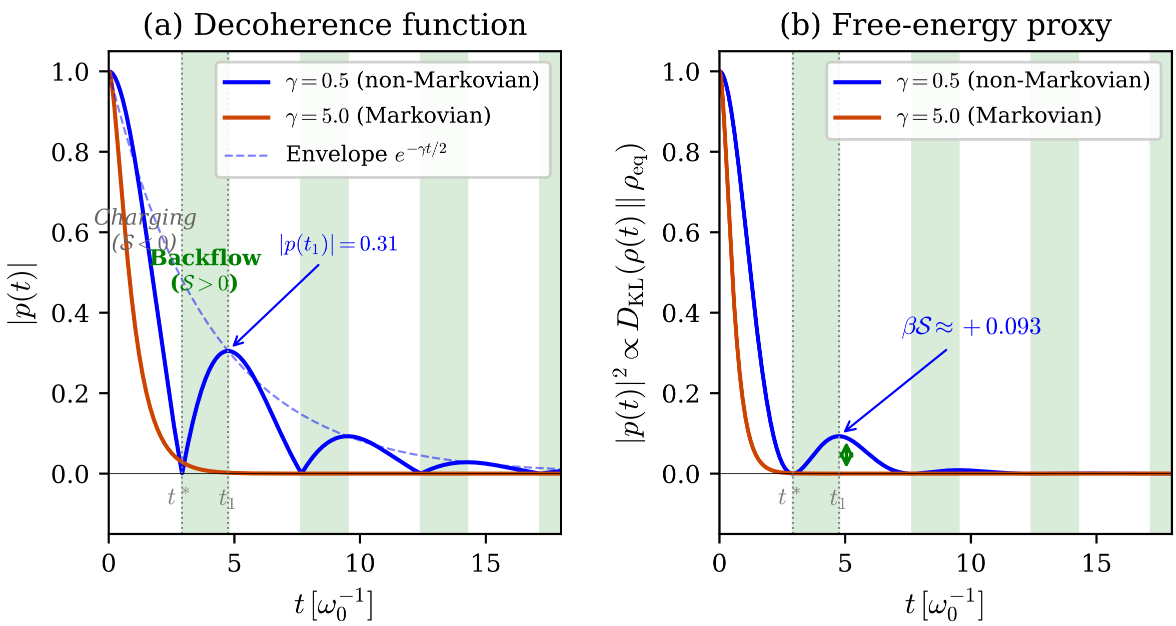

Figure 11.1: Pure-dephasing spin-boson model (Section 11.6) with Lorentz–Drude spectral density (11.36). **Parameters:** $\omega_0 = 1$ (energy unit), $\lambda = 1$ (reorganization energy). **Units:** all times in ω_0^{-1}; energies in $\hbar\omega_0$. **Regime:** low temperature ($\beta\omega_0 \gg 1$); the standard $T \to 0$ analytic expression (11.37) [18] is used as an accurate proxy. **(a)** Decoherence amplitude $|p(t)|$ (eq. (11.37)). Blue: non-Markovian ($\gamma = 0.5$, $\gamma/4\lambda = 0.125$). Orange: Markovian ($\gamma = 5.0$, $\gamma/4\lambda = 1.25$). Dashed: exponential envelope $e^{-\gamma t/2}$. Green bands indicate backflow (Remark 11.40: $d|p|/dt > 0$, $\Gamma(t) < 0$ per (11.42)). **(b)** Survival proxy $|p(t)|^2 \propto \beta\,\mathcal{S}$ (eq. (11.39)). At the first revival ($t_1 = \pi/\Omega \approx 4.75\,\omega_0^{-1}$), the non-Markovian agent achieves $\beta\,\mathcal{S}[t^*, t_1] \approx +0.093$ (eq. (11.41)), consistent with the closed-form prediction (11.43), funded by the consumption of pre-existing correlations (Proposition 11.28). The Markovian agent decays monotonically: $\mathcal{S} \leq 0$ always (Theorem 11.14).

The bound (11.28) is satisfied: $\beta\,\mathcal{S}[t^*, t_1] = 0.093 \leq I(S{:}E;\, t^*)$. Full-cycle closure (11.32) is confirmed: $\beta\,\mathcal{S}[0, t_1] < 0$.

Instantaneous decoherence rate. The rate of coherence loss is

$$\Gamma(t) := -\frac{d}{dt}\ln|p(t)| = \frac{\gamma}{2} - \frac{\Omega\sin(\Omega t) + \frac{\gamma}{2}\cos(\Omega t)}{\cos(\Omega t) + \frac{\gamma}{2\Omega}\sin(\Omega t)}. \tag{11.42}$$

In the Markovian limit $\gamma \gg 4\lambda$, $\Gamma(t) \to \gamma/2 > 0$ for all t (monotone decoherence). In the non-Markovian regime $\gamma < 4\lambda$, $\Gamma(t)$ oscillates and becomes *negative* during the backflow intervals where $|p(t)|$ increases. These are precisely the intervals where $\mathcal{S} > 0$.

Closed-form revival amplitude. The decoherence function (11.37) can be written as $p(t) = R\,e^{-\gamma t/2}\cos(\Omega t - \phi)$, where $R = \sqrt{1 + (\gamma/2\Omega)^2}$ and $\phi = \arctan(\gamma/2\Omega)$, with $R\cos\phi = 1$. The extrema of $|p(t)|$ occur at $t_n = n\pi/\Omega$ ($n = 0, 1, 2, \ldots$), and the first revival peak after the first zero is at $t_1 = \pi/\Omega$. Its amplitude is *exactly*

$$|p(t_1)| = e^{-\gamma\pi/(2\Omega)}, \qquad \beta\,\mathcal{S}[t^*, t_1] \approx |p(t_1)|^2 = e^{-\gamma\pi/\Omega}. \tag{11.43}$$

This is the paper's central computable prediction: the survival gain at first backflow is determined by a single dimensionless ratio γ/Ω.

Remark 11.41 (Parameter Survey)**.** *Table 11.1 demonstrates the transition from the Markovian regime ($\mathcal{S} \leq 0$) to the non-Markovian regime ($\mathcal{S} > 0$) as the bath memory rate γ decreases below the critical value 4λ. All entries use $\omega_0 = 1$, $\lambda = 1$, $W = 0$ (autonomous evolution), with revival amplitudes computed from* (11.43)*.*

γ	$\gamma/4\lambda$	*Regime*	$\lvert p(t_1)\rvert$	$\beta\,\mathcal{S}(t_1)$	$\Gamma_{\min}$
20.0	5.0	*Markov*	—	≤ 0	> 0
4.0	1.0	*Critical*	—	≤ 0	$= 0$
2.0	0.50	*Non-Markov*	0.043	+0.002	< 0
1.0	0.25	*Non-Markov*	0.163	+0.027	< 0
0.5	0.125	*Deep NM*	0.305	+0.093	< 0
0.1	0.025	*Deep NM*	0.605	+0.37	< 0

Table 11.1: Survival functional at first backflow revival as a function of the bath memory rate γ, for the pure-dephasing spin-boson model with Lorentz–Drude spectral density. $|p(t_1)|$ is computed from (11.43); $\Gamma_{\min}$ is the sign of the minimum of the instantaneous decoherence rate (11.42). The transition $\mathcal{S} \leq 0 \to \mathcal{S} > 0$ occurs precisely at the non-Markovian threshold $\gamma = 4\lambda$. For $\gamma = 0.1$ (deep non-Markovian), the agent achieves $\beta\,\mathcal{S} \approx +0.37$ per backflow cycle in the autonomous setting ($H_{\mathrm{ctrl}} = 0$).

Remark 11.42 (The Two Regimes: Summary)**.**

	Markovian *($\gamma = 20$)*	***Non-Markovian*** *($\gamma = 0.5$)*
$\gamma/4\lambda$	5.0 *(overdamped)*	0.125 *(underdamped)*
τ_B	$0.05\,\omega_0^{-1}$	$2.0\,\omega_0^{-1}$
$p(t)$	*Monotone decay*	*Oscillatory with envelope*
$\lvert p(t_1)\rvert$ *at first revival*	0 *(no revival)*	≈ 0.31
$\beta\,\mathcal{S}$ *at revival*	≤ 0	$\approx +0.093$
$\Gamma(t)$	> 0 *always*	*Oscillates,* < 0 *during backflow*
Interpretation	*Stone (sinks)*	*Surfer (rides backflow)*

The non-Markovian agent achieves $\beta\,\mathcal{S} \approx +0.093$ per backflow cycle (autonomous, $H_{\mathrm{ctrl}} = 0$), while the Markovian agent can only lose free energy. As the coupling deepens ($\gamma/4\lambda \to 0$), the revival amplitude grows and $\mathcal{S}$ increases (Table 11.1), bounded above by $\beta\,\mathcal{S} \leq I(S{:}E;\, t^*)$ (Corollary 11.26(iii)).

11.7 The Cost of Memory

We have shown that memory allows an agent to breach the Markovian ceiling. However, every advantage carries a thermodynamic shadow. We now quantify the cost of memory and identify the survival crisis that sets the stage for Paper II.

11.7.1 The Landauer Debt

To exploit the memory kernel $\mathcal{K}(t, s)$, the physical substrate of the agent must maintain correlations with its own past. This is equivalent to storing information. By Landauer's principle, erasing or overwriting this information dissipates heat; if the agent does not erase, it must pay an entropic cost to store.

Proposition 11.43 (Landauer Cost of Memory)**.** *Let $\mathcal{I}_{\mathrm{stored}}(\tau_{\mathrm{mem}})$ be the mutual information between the agent's state trajectory over $[t-\tau_{\mathrm{mem}},\, t]$ and its current control protocol $H_{\mathrm{ctrl}}(t)$. The free-energy cost of maintaining this memory satisfies*

$$\Delta F_{\mathrm{mem}} \geq k_B T \ln 2 \cdot \mathcal{I}_{\mathrm{stored}}(\tau_{\mathrm{mem}}). \tag{11.44}$$

11.7.2 The Memory Catastrophe

The crisis arises from the scaling of $\mathcal{I}_{\mathrm{stored}}$ with time. To quantify this, we borrow two quantities from computational mechanics [27, 90]:

Definition 11.44 (Entropy Rate and Predictive Information)**.** *Let $\{X_t\}$ be the stochastic process describing the environment's influence on the agent (e.g., the sequence of bath correlation values).*

1. *The **entropy rate** of the environment is*

$$h_\mu := \lim_{n\to\infty} H(X_n \mid X_{n-1}, \ldots, X_1), \tag{11.45}$$

measuring the intrinsic unpredictability per time step.

2. *The **predictive information** (excess entropy) is*

$$I_{\mathrm{pred}} := I(\overleftarrow{X};\, \overrightarrow{X}) = \sum_{k=1}^{\infty} \big[H(X_k) - h_\mu\big], \tag{11.46}$$

where $\overleftarrow{X}$ and $\overrightarrow{X}$ denote the past and future half-chains. This is the total amount of information about the future that is encoded in the past—the useful *memory.*

For an environment with finite predictive information ($I_{\mathrm{pred}} < \infty$), an optimal agent needs only finite memory to capture all exploitable correlations. However, for environments with *divergent* predictive information (e.g., processes with long-range temporal correlations, $1/f$ noise, or non-stationary statistics), the required memory grows without bound.

Proposition 11.45 (The Memory Catastrophe)**.** ***Assumptions.*** *Let the environment be a* stationary, mixing *stochastic process with positive entropy rate $h_\mu > 0$ [27, 90]. Consider an agent that maintains a memory kernel $\mathcal{K}(t,s)$ with support on $[t-\tau_{\mathrm{mem}},\, t]$. Let $\dot{W}_{\mathrm{budget}}$ be the agent's available free-energy flux (constant).*

1. *The minimum memory required to exploit correlations up to depth τ_{mem} satisfies*

$$\mathcal{I}_{\mathrm{stored}}(\tau_{\mathrm{mem}}) \geq \min\!\big(I_{\mathrm{pred}},\ h_\mu\, \tau_{\mathrm{mem}}\big). \tag{11.47}$$

2. *The Landauer cost of maintaining this memory is*

$$\dot{W}_{\mathrm{mem}} \geq k_B T \ln 2 \cdot h_\mu, \tag{11.48}$$

since the agent must erase (or overwrite) at least h_μ bits per unit time to prevent memory overflow.

3. *There exists a critical time t_{crit} beyond which the memory maintenance cost exceeds the survival gain:*
$$t > t_{\text{crit}} \implies \dot{W}_{\text{mem}}(t) > \dot{W}_{\text{budget}}, \tag{11.49}$$
unless the agent compresses its memory.

The agent dies not from entropy (disorder) but from hypermnesia*: the thermodynamic cost of perfect memory exceeds the benefit it provides.*

Proof. Part (1): an agent exploiting temporal correlations to depth τ_{mem} must store at least the mutual information between the past τ_{mem} time steps and the present. For a stationary ergodic process, this mutual information is bounded below by $\min(I_{\text{pred}}, h_\mu \tau_{\text{mem}})$ [27, 13].

Part (2): each time step, the agent receives $\sim h_\mu$ bits of genuinely new information. To maintain a fixed-capacity memory, it must erase at least this many bits, incurring Landauer cost $k_B T \ln 2 \cdot h_\mu$ per time step.

Part (3): if $I_{\text{pred}} = \infty$ (as for environments with long-range correlations), the stored information grows as $\mathcal{I}_{\text{stored}} \sim h_\mu \tau_{\text{mem}}$. Combined with part (2), the memory cost grows linearly in the effective memory depth. For any finite budget $\dot{W}_{\text{budget}}$, there exists t_{crit} such that the cost exceeds the budget. □

Remark 11.46 (Bound tightness)**.** *This bound is tight only for environments with long-range temporal correlations (infinite predictive information I_{pred}). For short-range-correlated environments, the mutual information between past and present is $O(1)$ (the* excess entropy*), and the memory cost is bounded rather than divergent.*

11.7.3 Resolution: The Necessity of Forgetting

To survive beyond t_{crit}, the agent must introduce a *lossy compression* scheme: it must discard the vast majority of stored correlations and retain only the thermodynamically salient features.

- **Compression requires a criterion.** To decide what to keep and what to erase, the agent needs a *relevance function*—a mapping from stored correlations to survival value. This is a reference frame that ranks information by its contribution to $\mathcal{S}$.
- **A reference frame requires symmetry breaking.** An "unbiased" agent that treats all correlations as equally valuable cannot compress: it must keep everything. The act of preferring one subset of information over another is a spontaneous breaking of the informational symmetry. This is the thermodynamic definition of a "perspective"—or, more precisely, a *privileged basis.*

Remark 11.47 (The Origin of Paper II)**.** *Proposition 11.45 reveals the* poison *embedded in Paper I's medicine. Memory enables survival beyond the Markovian ceiling, but unbounded memory under finite energy resources leads to* computational explosion*: the agent must process an ever-growing archive with bounded free energy.*

This is the precise thermodynamic origin of the crisis addressed in Paper II. The resolution—spontaneous symmetry breaking of the agent's reference frame—is not an additional hypothesis but a thermodynamic necessity*: the agent must compress its infinite history into a finite, biased representation. The "self" (a privileged computational basis) emerges as the minimal structure that makes memory computationally tractable.*

In the structural parallel noted in HAFF Essay C [129]: the accumulation mechanism of Paper I provides the raw material for survival, but without the discriminative compression of Paper II, the system collapses under the weight of its own stored correlations.

11.8 Numerical Demonstration

The preceding sections establish analytic bounds and a worked example in the spin-boson model. We now provide a numerical illustration showing that the Markovian ceiling signature predicted by Theorem 11.14 and the memory advantage of Theorem 11.22 are reproduced in a minimal partially observed environment. Full code and parameters are provided for reproducibility.

11.8.1 Model

Environment. A two-hidden-state HMM with aliased observations. The hidden state $s_t \in \{0, 1\}$ evolves as a persistent Markov chain with $\Pr(s_{t+1} = s_t) = 1-\varepsilon$; the parameter $\varepsilon \in [10^{-3}, 10^{-1}]$ controls the correlation length $\ell \sim 1/\varepsilon$. Observations $o_t \in \{A, B\}$ are aliased: $\Pr(o_t = A \mid s_t = 0) = 0.5+\delta$, $\Pr(o_t = A \mid s_t = 1) = 0.5-\delta$, with $\delta = 0.05$ (mutual information $I(O; S) \approx 0.007$ bits). Reward: $r_t = 1$ if $a_t = s_t$, 0 otherwise.

Agents. All agents use the true model parameters and compute exact Bayesian posteriors; the only difference is how many observations each agent retains.

- **Markov-k** ($k \in \{1, 2, 4, 8\}$): runs an exact Bayes filter over the most recent k observations (sliding window, uniform prior at each window start); acts by MAP.
- **Memory (Bayes filter)**: maintains the full belief state $b_t = \Pr(s_t = 1 \mid o_{1:t})$ via the exact predict–update cycle over all past observations; acts by MAP.

Parameters.

Quantity	Value	Role
T	100,000	horizon per trial
Seeds	10	independent replications
δ	0.05	observation asymmetry
k	$\{1, 2, 4, 8\}$	Markov window sizes
ε	logspace(10^{-3}, 10^{-1}, 15)	transition noise grid
Burn-in	5,000	discarded steps

11.8.2 Results

Figure 11.2 shows the two key signatures.

Result 1: Markov ceiling (Figure 11.2a). The average reward $\bar{R}$ of the Bayes filter (memory agent) increases monotonically with correlation length $\ell = 1/\varepsilon$, while each Markov-k agent saturates at a distinct ceiling. The ceilings are ordered: $k = 1$ (lowest) through $k = 8$ (highest), and all fall below the memory agent for $\ell \gtrsim 20$. This is consistent with the qualitative prediction of Theorem 11.14: finite-order Markov representations have a performance upper bound that the memory-carrying agent surpasses.

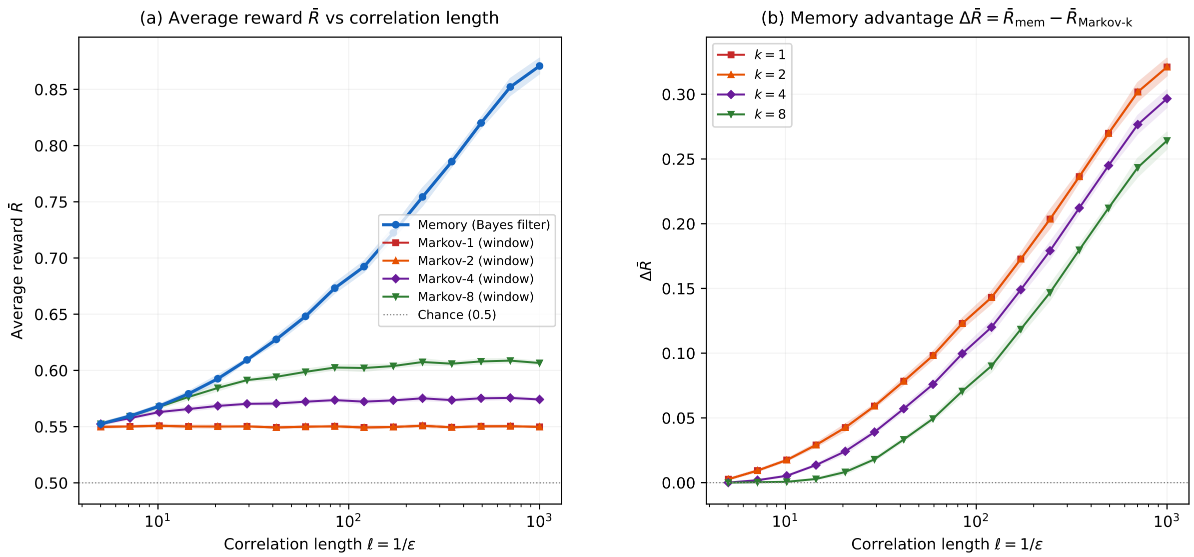

Figure 11.2: **Markov ceiling and memory advantage.** $T = 100{,}000$, 10 seeds, 95% CI bands. **(a)** Average reward $\bar{R}$ vs correlation length $\ell = 1/\varepsilon$. The Bayes filter (blue, bold) rises monotonically; Markov-k agents saturate at k-dependent ceilings. **(b)** Performance gap $\Delta\bar{R} = \bar{R}_{\text{mem}} - \bar{R}_{\text{Markov-}k}$ increases with ℓ; smaller k yields a larger gap.

Result 2: Memory advantage (Figure 11.2b). The gap $\Delta\bar{R} = \bar{R}_{\text{mem}} - \bar{R}_{\text{Markov-}k}$ increases monotonically with ℓ, and is larger for smaller k. Shaded bands show 95% confidence intervals across 10 seeds. The Markov-1 and Markov-2 curves nearly overlap at small ℓ, reflecting the fact that short observation windows provide negligible additional information in this aliasing regime—a consistency check, not a deficiency.

11.8.3 Scope of This Demonstration

These simulations illustrate the ceiling phenomenon predicted by Theorem 11.14 under the stated model class; they do not constitute a proof beyond this class.

This demonstration **does** show:

1. A reproducible regime in which finite-order Markov agents exhibit a performance ceiling while a memory-carrying (Bayes filter) agent improves—the Markov ceiling signature predicted by Theorem 11.14.

2. The memory advantage (Theorem 11.22) manifests as a monotonically growing gap that widens with correlation length and tightens with window size.

This demonstration does **not** show:

1. Universality across environments, observation models, or agent architectures. The model uses a two-state HMM with binary aliased observations.

2. Tight constants or the functional form of the ceiling boundary $\ell_c(k)$.

3. That the Bayes filter is optimal among all possible memory-carrying agents.

Reproducibility. The complete simulation is a self-contained Python script (`paper1_markov_ceiling_demo.py`, ~ 560 lines, requiring only NumPy and Matplotlib) with fixed random seeds. All figures in this section can be reproduced by executing the script. The following files are included in the supplementary archive:

- `paper1_markov_ceiling_demo.py` — simulation script
- `fig_paper1_markov_ceiling.pdf` — Figure 11.2
- `markov_ceiling_data.csv` — raw sweep data
- `markov_ceiling_boundary.csv` — extracted ceiling boundaries $\ell_c(k)$

11.9 Discussion

11.9.1 Summary of Results

Result	**Statement**	**Sec.**
Markovian Ceiling	$\mathcal{S} \leq 0$ for open-loop GKSL (no feedback)	11.3
Memory Advantage	$\beta\mathcal{S} = -\Delta I - \Delta D_{\mathrm{KL}} - \beta\Delta\langle H_{\mathrm{ctrl}}\rangle$; $\mathcal{S} > 0$ when correlations consumed (any initial state)	11.4
Quantitative demo	Spin-boson: $\beta\mathcal{S} \approx +0.093 > 0$ at first backflow revival (Fig. 11.1, Table 11.1)	11.6
Temporal Arrow	$\prec_K \to \prec_{\mathrm{HAFF}}$ via quantum Darwinism	11.5
Memory Catastrophe	$\dot{W}_{\mathrm{mem}} \geq k_B T \ln 2 \cdot h_\mu$; exceeds budget at t_{crit}	11.7
Numerical demo	Markov ceiling reproduced in HMM (Fig. 11.2)	11.8

11.9.2 What This Paper Does and Does Not Show

This paper shows:

1. Under open-loop GKSL dynamics (no measurement or feedback), the survival functional $\mathcal{S} \leq 0$ (Theorem 11.14).

2. For any initial state (product or correlated), the survival functional satisfies the exact identity $\beta\,\mathcal{S} = -\Delta I(S{:}E) - \Delta D_{\mathrm{KL}}(\rho_E \| \rho_E^{\mathrm{th}}) - \beta\,\Delta\langle H_{\mathrm{ctrl}}\rangle$ (Theorem 11.22). Under autonomous evolution, when pre-existing system–environment correlations are consumed ($\Delta I < 0$), $\mathcal{S} > 0$ is achievable, bounded by the initial correlation budget (Corollary 11.26).

3. A quantitative spin-boson example illustrates: $\beta\,\mathcal{S} \approx +0.093 > 0$ at the first non-Markovian revival (Section 11.6).

4. The causal memory order $\prec_K$ is consistent with the HAFF accessibility order when restricted to the classical (pointer-state) sector (Proposition 11.36).

5. The thermodynamic cost of memory, quantified by the environment's entropy rate h_μ, creates a survival crisis for agents with finite energy budgets (Proposition 11.45).

6. A minimal computational demonstration reproduces the Markov ceiling and memory advantage signatures in a two-state HMM with aliased observations (Section 11.8, Figure 11.2).

This paper does not show:

1. That non-Markovian dynamics is *sufficient* for persistence (it is necessary but not sufficient; Paper II addresses the additional requirements).

2. That *all* non-Markovian systems outperform all Markovian systems (the comparison is between suprema under specified constraints).

3. That Markovian agents with explicit measurement-feedback are bounded by the ceiling (the Sagawa–Ueda framework shows they are not; Remark 11.15).

4. That the specific form of the optimal memory kernel can be derived from first principles without specifying the environment.

5. That memory implies or requires consciousness.

Chapter 12

Spontaneous Symmetry Breaking of Reference Frames as a Computational Cost Minimization Strategy

Paper II — "The Ego"

Originally published: Zenodo, DOI: 10.5281/zenodo.18579703

Abstract

We investigate the computational constraints on persistent open quantum systems that carry non-Markovian memory (Paper I [143]). Paper I established that memory is a thermodynamic necessity for survival beyond the Markovian ceiling, but revealed a secondary crisis: the *Memory Catastrophe*, in which the Landauer cost of maintaining unbounded history exceeds any finite free-energy budget.

We prove a **Computational Ceiling**: any agent that processes its memory *symmetrically*—treating all components of its internal Clifford algebra $Cl(V, q)$ as equally relevant—reaches computational paralysis at a finite critical time t_{par}.

We then show that the resolution requires **spontaneous symmetry breaking** of the agent's internal reference frame: the selection of a privileged basis (a gauge fixing of the automorphism group $G = \text{Aut}(Cl(V, q))$) that compresses the memory kernel into a tractable, low-dimensional representation. The optimal compression is governed by a survival-weighted rate-distortion bound; under generic conditions, the agent retains $k^* = \mathcal{C}_{\text{budget}}/h_\mu$ components and discards the rest.

This establishes **reference-frame selection as the survival-optimal strategy under bounded rationality**: the "self" (a privileged computational basis) is not an additional hypothesis but the minimal structure that makes memory computationally tractable.

The broken phase introduces four systematic bias terms—basis selection, frame drag, objective centering, and model incompleteness—that are generic consequences of gauge fixing under assumptions (B1)–(B5). We show that under environmental drift, a fixed reference frame leads to the **Delusion Trap**: an exponential divergence of prediction error that the agent cannot detect from within its own frame, establishing the crisis that Paper III must resolve.

12.1 Introduction

12.1.1 Context: The Problem of Overload

Paper I of this series [143] established that non-Markovian memory is a thermodynamic necessity for persistent far-from-equilibrium systems: under open-loop Markovian (GKSL) dynamics, the survival functional satisfies $\mathcal{S} \leq 0$ (the Markovian Ceiling), while agents carrying memory kernels can achieve $\mathcal{S} > 0$ by consuming stored system–environment correlations.

This result, however, carries a price. The *Memory Catastrophe* (Paper I, Proposition 10) shows that the Landauer cost of maintaining a memory archive of depth τ_{mem} grows at a rate

$$\dot{W}_{\text{mem}} \geq k_B T \ln 2 \cdot h_\mu, \tag{12.1}$$

where h_μ is the entropy rate of the environmental process [27, 90]. For any finite free-energy budget $\dot{W}_{\text{budget}}$, there exists a critical time t_{crit} beyond which $\dot{W}_{\text{mem}} > \dot{W}_{\text{budget}}$: the agent's memory consumes more resources than are available.

But thermodynamic cost is only half the crisis. Even if unlimited free energy were available for memory maintenance, the agent must still *process* the stored correlations—evaluate the survival functional as a function of its ever-growing archive—using finite computational resources. This is the problem that the present paper addresses.

12.1.2 Position within the Series

This paper is the second of three constituting the **T-DOME** (Thermodynamic Dynamics of Observer-Memory Entanglement) framework, the third pillar of a three-paper program.

Framework	Question		Result	Status
HAFF [127, 128]	How does geometry emerge?	Ocean	Algebra $\to$ Geometry	Complete
Q-RAIF [140, 141]	What algebra must an observer have?	Fish	$Cl(V, q) \hookrightarrow Cl(1, 3)$	Complete
T-DOME I [143]	Why must agents carry memory?	Seed	Markovian ceiling; memory as necessity	Complete
T-DOME II (this work)	Why must agents break symmetry?	Ego	Reference-frame selection under bounded computation	**This paper**
T-DOME III	How does self-calibration arise?	Loop	Fisher self-referential bound	Planned

The three T-DOME papers form an irreversible logical chain. Each resolves a survival crisis created by its predecessor:

1. **Paper I (The Seed):** Without memory, a system is trapped in the *Markovian present*—no accumulation, no temporal arrow, inevitable thermal death. Memory breaks this trap but floods the system with unbounded historical data.

2. **Paper II (The Ego, this work):** Unbounded memory under finite computational resources causes processing collapse. Spontaneous symmetry breaking of the reference frame (establishing a "self") resolves the overload but introduces systematic bias.

3. **Paper III (The Loop):** Uncorrected bias diverges from a changing environment. A self-referential calibration loop (monitoring one's own prediction error) resolves the bias but requires the system to "observe its own observation"—closing the self-calibration loop.

12.1.3 Relation to Q-RAIF

Q-RAIF Paper B [141] established that any persistent open quantum subsystem maintaining a non-equilibrium steady state (NESS) requires an internal control algebra isomorphic to a Clifford algebra $Cl(V, q)$. Paper C [142] showed that this algebra must embed in the environmental observable algebra via a realizability homomorphism $\phi : Cl(V, q) \hookrightarrow Cl(1, 3)$.

The Clifford algebra $Cl(V, q)$, however, admits a non-trivial *automorphism group* $G = \mathrm{Aut}(Cl(V, q))$. In the absence of external constraints, all elements of G yield physically equivalent representations—the choice of basis within the algebra is a *gauge freedom*. This gauge freedom is the mathematical substrate of the symmetry that the present paper breaks.

The "ego" is not a new algebraic structure imposed from outside the Q-RAIF framework; it is a *gauge fixing* of the already-present internal symmetry, driven by computational optimality under bounded resources.

12.1.4 Relation to HAFF Paper G

HAFF Paper G established *architectural incompleteness*: the observable-algebra framework cannot self-ground [131]. The present paper provides a partial operational resolution: under bounded computation, an agent satisfying (B1)–(B5) is driven to choose a computational basis (break symmetry) precisely *because* the framework is incomplete. The ego is an operational response to incompleteness, not a metaphysical addition.

12.1.5 Scope and Disclaimers

To prevent interpretational overreach, we state at the outset what this paper does *not* claim:

1. We do not claim that symmetry breaking is *sufficient* for persistence. Paper III addresses the additional requirements.

2. We do not claim that the specific form of the privileged basis is unique—only that *some* basis selection is necessary under bounded computation.

3. The term "ego" or "self" is used in the control-theoretic sense: a fixed reference frame within the agent's internal algebra. It carries no implication of consciousness or subjective experience.

4. A broader structural analogy with classical philosophical concepts of selfhood exists but is outside the scope of this paper.

Related work. The idea that bounded agents must compress their representations has roots in Simon's bounded rationality [93], Shannon's rate-distortion theory [91, 25], and Sims's rational inattention [94], which models finite-capacity decision-makers as solving a rate-distortion problem—precisely the economic counterpart of our $\mathcal{C}_{\text{budget}}$ formalism. The information bottleneck [102] formalises relevance-weighted compression and has been applied to neural coding and deep learning. The role of decoherence in selecting preferred bases (pointer states) is well established via quantum Darwinism [122]; our contribution is to show that the same selection arises as a *computational* necessity, independent of the decoherence mechanism. Measures of non-Markovianity and their thermodynamic consequences are reviewed in [80, 18]; the connection to survival was established in Paper I.

Summary of contributions. This paper establishes three main results:

1. **Computational Ceiling scaling law** (Theorem 12.7): symmetric processing of a $Cl(V, q)$ memory kernel requires rate $\mathcal{R} \geq h_\mu \cdot D$, leading to paralysis at a finite τ_{par}.
2. **Survival-weighted rate-distortion bound** (Theorem 12.17): the optimal gauge-fixed representation retains $k^* = \lfloor \mathcal{C}_{\text{budget}}/h_\mu \rfloor$ components.
3. **Delusion dynamics** (Theorem 12.32): a fixed reference frame decouples from a drifting environment on the logarithmic timescale $t_{\text{del}} = \Lambda^{-1} \ln(\pi/4\theta_0)$.

12.2 Mathematical Preliminaries

12.2.1 Inherited Framework from Paper I

We briefly recall the key objects from Paper I [143] that the present work builds upon. The reader is referred to Paper I for full definitions and proofs.

Survival functional. For an open quantum system S coupled to an environment E at inverse temperature β, with dynamics Λ and external control protocol $H_{\text{ctrl}}(t)$, the survival functional is

$$\mathcal{S}[\Lambda, \tau] := \Delta F - W[0, \tau], \tag{12.2}$$

where $\Delta F = F(\rho(\tau)) - F(\rho(0))$ is the change in non-equilibrium free energy and $W = \int_0^\tau \operatorname{tr}\Big(\rho(t)\, \dot{H}_{\text{ctrl}}(t)\Big)\, dt$ is the work performed by the external protocol.

Markovian Ceiling. Under open-loop GKSL dynamics with no feedback (control class $\mathcal{C}_{\text{M}}$, Paper I, Definition 6):

$$\mathcal{S}[\Lambda^{\text{M}}, \tau] \leq 0 \qquad \text{for all } \tau \geq 0. \tag{12.3}$$

Non-Markovian advantage identity. For arbitrary initial states:

$$\beta\, \mathcal{S} = -\Delta I(S{:}E) - \Delta D_{\text{KL}}(\rho_E \| \rho_E^{\text{th}}) - \beta\, \Delta \langle H_{\text{ctrl}} \rangle. \tag{12.4}$$

Memory Catastrophe. The Landauer cost of maintaining a memory archive of depth τ_{mem} satisfies $\dot{W}_{\text{mem}} \geq k_B T \ln 2 \cdot h_\mu$ (Paper I, Proposition 10), where h_μ is the *per-component* entropy rate of the environmental process [27], defined by

$$h_\mu := \lim_{T \to \infty} \frac{1}{T} H(X_{0:T}), \tag{12.5}$$

measuring the asymptotic information (in bits per unit time) generated by a single algebraic component of the memory kernel (we work in units where the sampling interval equals the environmental correlation time τ_E)[1]—and the stored mutual information grows as $\mathfrak{i}_{\text{stored}}(\tau_{\text{mem}}) \geq \min(I_{\text{pred}},\ h_\mu \, \tau_{\text{mem}})$, with I_{pred} the *predictive information* (excess entropy) [13, 90], defined as the mutual information between past and future of the environmental process:

$$I_{\text{pred}} := I\Big(\overleftarrow{X};\, \overrightarrow{X}\Big) = H(\overrightarrow{X}) - H(\overrightarrow{X} \mid \overleftarrow{X}), \tag{12.6}$$

where $\overleftarrow{X}$ and $\overrightarrow{X}$ denote the semi-infinite past and future, respectively. For a stationary process, I_{pred} relates to h_μ via the entropy-rate decomposition $H(X_{1:T}) = I_{\text{pred}} + h_\mu T + o(1)$ as $T \to \infty$ [27].

12.2.2 The Agent's Internal Algebra

Following Q-RAIF [141, 142], the agent's internal control algebra is a Clifford algebra $\mathcal{O}_{\text{int}} = Cl(V, q)$ for a real vector space V equipped with a non-degenerate quadratic form q. The algebra satisfies the fundamental relation $v^2 = q(v)\,\mathbf{1}$ for all $v \in V$.

The *automorphism group*

$$G := \text{Aut}(Cl(V, q)) \tag{12.7}$$

is the group of algebra automorphisms that preserve the grading and quadratic form.[2] For $Cl(1,3)$, G contains the spin group $\text{Spin}(1,3) \cong SL(2, \mathbb{C})$ as a subgroup—a six-real-dimensional Lie group.

In the absence of computational constraints, all $g \in G$ yield physically equivalent descriptions of the agent's internal state. The choice of basis within $Cl(V, q)$ is a *gauge freedom*—the symmetry that will be broken.

The realizability embedding $\phi : Cl(V, q) \hookrightarrow Cl(1, 3)$ (Q-RAIF Paper C) constrains the physically accessible reference frames: only gauge choices compatible with $\text{Im}(\phi) \subset Cl(1, 3)$ are realizable.

Dimensional convention. Two distinct notions of dimension appear throughout:

Symbol	Meaning	Scaling
$n := \dim V$	number of generators (degrees of freedom)	—
$D := \dim Cl(V, q) = 2^n$	full multivector space (algebra basis size)	exponential in n

[1]For a continuous-valued process sampled at resolution b bits, h_μ includes the quantisation cost: $h_\mu = h_\mu^{(\text{diff})} + b\, f_s$, where $h_\mu^{(\text{diff})}$ is the differential entropy rate and f_s the sampling frequency. All budget inequalities in this paper hold with h_μ so defined.

[2]We use G as an effective symmetry group acting transitively on admissible frames. The detailed Lie-algebraic structure of G is not required for our results; only the existence of a non-trivial symmetry that must be broken (assumption (B5)). In concrete models, one may replace G by its image under the adjoint representation—typically $O(V, q)$ or a pin/spin subgroup.

The Computational Ceiling (Section 12.3) scales with D, not n; the distinction matters whenever one compares generator-level and algebra-level quantities.

12.2.3 Rate-Distortion Theory

We require the classical rate-distortion framework of Shannon [91].

Definition 12.1 (Rate-distortion function). *Let X be a random source with distribution $p(x)$, $\hat{X}$ a reconstruction, and $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ a distortion measure. The* rate-distortion function *is*

$$R(D) := \min_{\substack{p(\hat{x}|x): \\ \mathbb{E}[d(X,\hat{X})] \leq D}} I(X; \hat{X}), \tag{12.8}$$

the minimum mutual information between source and reconstruction that achieves average distortion at most D.

$R(D)$ is a convex, non-increasing function of D with $R(0) = H(X)$ (lossless) and $R(D_{\max}) = 0$ (maximum distortion). It provides the fundamental limit on lossy compression [25]. The *information bottleneck* method of Tishby et al. [102] generalises this framework to the case where the relevant variable is not the source itself but a downstream prediction target—precisely the situation in our survival-weighted compression problem (Section 12.4.2).

12.2.4 Bounded Rationality

Following Simon [93], we model computational limitations as a hard constraint on the agent's information processing rate.

Definition 12.2 (Computational budget). *The agent's* computational budget $\mathcal{C}_{\text{budget}}$ *(measured in bits per unit time) is the maximum rate at which the agent can evaluate functions of its stored correlations. We assume $\mathcal{C}_{\text{budget}} < \infty$.*

Physically, finiteness of $\mathcal{C}_{\text{budget}}$ reflects the finite number of degrees of freedom in the agent's physical substrate: finite Hilbert space dimension, finite memory register size, and finite energy available for computation (Landauer's principle [59, 12]).

12.2.5 Fiber Bundle Formalism

The geometric setting for reference-frame selection is a principal fiber bundle.

Definition 12.3 (Gauge bundle). *The* gauge bundle *is the principal G-bundle*

$$\pi : P \to M, \qquad G = \mathrm{Aut}(Cl(V, q)), \tag{12.9}$$

where:

- *M is the base space of* effective memory kernels*—equivalently, the space of induced sufficient-statistic processes accessible to the agent (a finite-dimensional manifold that admits local parametrisation by the environmental spectral-density couplings);*
- *G is the structure group acting transitively on admissible frames (see footnote 2 for the effective subgroup);*

- *the fiber* $\pi^{-1}(\kappa)$ *over a kernel* $\kappa \in M$ *is the* G*-orbit of equivalent algebraic representations (frames) for describing* κ *in* $Cl(V, q)$*;*
- *a* section $\sigma : M \to P$ *constitutes a global gauge-fixing policy—a systematic choice of reference frame for every kernel configuration.*

A *connection* on P specifies how the reference frame is parallel-transported as the agent's state evolves. The curvature of this connection measures the extent to which the reference frame "twists" along different paths through state space.

12.2.6 Standing Assumptions

Definition 12.4 (Standing Assumptions)**.** *Throughout this paper, the following conditions are assumed:*

(B1) ***Inherited framework.*** *All assumptions (A1)–(A5) of Paper I [143] remain in force (open quantum system coupled to a thermal bath, well-defined free energy, non-equilibrium initial state, finite-dimensional system Hilbert space, and weak-coupling or controlled-coupling regime). Additionally, the agent possesses an internal control algebra* $\mathcal{O}_{\text{int}} = Cl(V, q)$ *with realizability embedding* $\phi : Cl(V, q) \hookrightarrow Cl(1, 3)$ *(Q-RAIF [142]).*

(B2) ***Finite computational budget.*** *The agent's information processing rate satisfies* $\mathcal{C}_{\text{budget}} < \infty$ *(Definition 12.2).*

(B3) ***Non-trivial environment.*** *The entropy rate satisfies* $h_\mu > 0$ *and the memory depth satisfies* $\tau_{\text{mem}} > 0$*. In Sections 12.4–12.7 we additionally require that the Computational Ceiling is binding:* $\tau_{\text{mem}} > \tau_{\text{par}}$ *(Theorem 12.7), i.e., the symmetric phase is computationally intractable.*

(B4) ***Survival imperative.*** *The agent's dynamics must maintain* $\mathcal{S} \geq \mathcal{S}_{\min}$ *over survival horizons* $T \gg \tau_{\text{mem}}$*. This is a persistence constraint, not an optimization objective.*

(B5) ***Gauge symmetry of bare algebra.*** *The automorphism group* $G = \mathrm{Aut}(Cl(V, q))$ *is non-trivial (*$G \neq \{e\}$*). In the absence of computational constraints, all* $g \in G$ *yield physically equivalent descriptions.*

12.3 The Computational Ceiling

We now establish the fundamental computational limitation of symmetric agents—those that treat all components of their internal algebra as equally relevant. The result is the computational analogue of Paper I's Markovian Ceiling: where that theorem showed that *memoryless* dynamics cannot achieve $\mathcal{S} > 0$, the present theorem shows that *unbiased processing* of memory leads to computational paralysis.

12.3.1 The Information Processing Inequality for Bounded Agents

Accounting convention. To ensure dimensional consistency throughout, we distinguish two quantities:

- $\mathcal{C}_{\text{budget}}$: the agent's processing *rate* (bits per unit time).
- $\mathcal{I}_{\text{proc}}(\tau)$: the total information (bits) that must be processed per evaluation cycle when the memory archive has depth τ.

The agent must complete one evaluation cycle per environmental correlation time τ_E. The *processing rate* required for a memory depth τ is

$$\mathcal{R}_{\text{proc}}(\tau) := \frac{\mathcal{I}_{\text{proc}}(\tau)}{\tau_E}. \tag{12.10}$$

Paralysis occurs when $\mathcal{R}_{\text{proc}}(\tau_{\text{mem}}) > \mathcal{C}_{\text{budget}}$. Hereafter we measure time in units of τ_E (i.e., set $\tau_E = 1$), so that rates and per-cycle information quantities are numerically equal.

Definition 12.5 (Symmetric processing). *An agent processes its memory* symmetrically *if both its cost functional $\mathcal{C}[\cdot]$ and its distortion measure $D(\cdot)$ are G-invariant: $\mathcal{C}[g \cdot \mathcal{K}] = \mathcal{C}[\mathcal{K}]$ and $D(g \cdot \mathcal{F}) = D(\mathcal{F})$ for every $g \in G = \text{Aut}(Cl(V, q))$. In operational terms: for every stored correlation c_i in the memory kernel $\mathcal{K}(t, s)$ and every $g \in G$, the cost of evaluating c_i equals the cost of evaluating $g \cdot c_i$, and no basis direction is* a priori *preferred for survival evaluation.*

Remark 12.6 (Operational meaning of processing rate). *We define the processing rate $\mathcal{R}_{\text{proc}}$ as an* information-throughput *measure: the number of algebraic components that must be updated per unit time, multiplied by the innovation rate h_μ per component. It captures the* bandwidth *cost of maintaining an internal representation, not the algorithmic gate complexity of individual operations.*

Theorem 12.7 (Computational Ceiling). *Let an agent satisfy assumptions* (B1)–(B5) *with memory depth τ_{mem} and per-component entropy rate $h_\mu > 0$. Assume the D environmental components carry independent information at rate h_μ each. Assume the environment is* ***unstructured*** *in the following two senses:* (i) *the effective activated dimension satisfies $D_{\text{eff}} \approx D$ (all grades of $Cl(V, q)$ carry non-negligible correlations), and* (ii) *the predictive information is not concentrated on a known sub-algebra (the agent possesses no* a priori *knowledge of the environmental symmetry group and cannot exploit group-theoretic shortcuts such as irreducible representations or Schur decompositions).[3] Within the class of* symmetric representations *that retain all D components with equal fidelity (permitting no privileged subspace)—thereby precluding structured compression techniques such as sparse coding or Johnson–Lindenstrauss embeddings [56], as these inherently implement a form of symmetry breaking—the minimum processing rate satisfies*

$$\mathcal{R}_{\text{proc}}^{\text{sym}} \;\geq\; h_\mu \cdot D, \qquad D := \dim Cl(V, q) = 2^n, \tag{12.11}$$

where $n = \dim V$ is the number of generators. This rate scales linearly *in the algebra dimension D and* exponentially *in n.*

For any finite $\mathcal{C}_{\text{budget}}$, the maximum memory depth that can be processed before correlations expire is

$$\tau_{\text{par}} := \frac{\mathcal{C}_{\text{budget}}}{h_\mu \cdot D}. \tag{12.12}$$

[3] If the agent knows the environmental symmetry group H, symmetric processing can be restricted to the isotypic components of H, reducing the effective dimension to $D_{\text{eff}} \leq D$. The ceiling applies to the *generic* (worst-case) scenario. All subsequent results hold *a fortiori* when D is replaced by D_{eff}.

Here τ_{par} *is measured in units of* τ_E *(environmental correlation times), not seconds; cf. the accounting convention at the start of this section.*

For $\tau_{\text{mem}} > \tau_{\text{par}}$*, the agent's evaluation cycle cannot complete within one correlation time:*

$$\mathcal{I}_{\text{proc}}(\tau_{\text{mem}}) = h_\mu \cdot \tau_{\text{mem}} \cdot D > \mathcal{C}_{\text{budget}}. \tag{12.13}$$

Stored correlations go stale before they can be used.

Proof. Under symmetric processing, the agent maintains D parallel correlation channels—one for each independent algebraic component of $Cl(V,q)$. The environment generates innovations at rate h_μ bits per unit time in each channel (Remark 12.6). Over a memory depth τ_{mem}, the total information load is therefore $\mathcal{I}_{\text{proc}}(\tau_{\text{mem}}) = D \cdot h_\mu \cdot \tau_{\text{mem}}$ bits [25], and the required *rate* is $\mathcal{R}^{\text{sym}}_{\text{proc}} = D \cdot h_\mu$ bits per unit time.

The agent must complete one evaluation cycle within τ_E (one environmental correlation time); otherwise the oldest correlations expire before use. Setting $\mathcal{R}^{\text{sym}}_{\text{proc}} = \mathcal{C}_{\text{budget}}$ and solving for τ_{mem} gives τ_{par} (12.12). □

Remark 12.8 (Accounting identity). *We note that this result is an accounting identity under the independence assumption: joint processing across correlated channels could in principle reduce the total rate below* $D \cdot h_\mu$*. The independence assumption is the substantive physical input.*

Corollary 12.9 (The Symmetry Tax). *Maintaining full gauge invariance imposes a multiplicative overhead of* $D = 2^n$ *on all computational operations relative to a fixed-basis agent that processes only* k *components. The overhead ratio is* D/k*, which for* $Cl(1,3)$ *(*$D = 16$*,* $k = 2$*) is* $8\times$*, and grows exponentially with the number of generators* n*.*

Remark 12.10 (Effective vs. full dimension). *The ceiling uses* $D = \dim Cl(V,q) = 2^n$*, the full multivector dimension. In practice, the environment may couple to only a subset of grades (e.g., grade-1 generators), yielding an effective dimension* $D_{\text{eff}} \leq D$*. For a* structured *environment where the agent knows which grades are active, the ceiling can be tightened to* $\mathcal{R}_{\text{proc}} \gtrsim h_\mu \cdot D_{\text{eff}}$*. The unstructured assumption (B3) represents the worst case; all subsequent results hold* a fortiori *when* D *is replaced by* D_{eff}*.*

12.3.2 Processing Collapse

Proposition 12.11 (Processing Collapse). *Under* (B1)–(B5)*, an agent that maintains full gauge symmetry reaches computational paralysis at time* τ_{par} (12.12)*. Beyond* τ_{par}*, the agent's processing latency* δt_{proc} *exceeds the environmental correlation time* τ_E*:*

$$\delta t_{\text{proc}}(\tau_{\text{mem}}) = \frac{D \cdot \tau_{\text{mem}}}{\mathcal{C}_{\text{budget}}/h_\mu} > 1 \qquad \textit{(in units of } \tau_E\textit{).} \tag{12.14}$$

Every stored correlation becomes stale before it can be evaluated, rendering the entire memory archive operationally useless.

Remark 12.12 (Comparison with Paper I's Memory Catastrophe). *Paper I's Memory Catastrophe is* thermodynamic*: the* cost of storing *memory exceeds the energy budget. The Computational Ceiling is* informational*: the* cost of processing *memory exceeds the computational budget. The two crises are complementary—an agent with unlimited energy but finite computation is still paralyzed, and vice versa. The resolution of both crises is the same: compression through symmetry breaking.*

12.4 The Symmetry Breaking Resolution

12.4.1 Reference Frame as Gauge Fixing

Definition 12.13 (Reference frame). *A reference frame $\mathcal{F}$ is a section $\sigma : M \to P$ of the gauge bundle (Definition 12.3). Choosing σ is equivalent to selecting a preferred orthonormal basis $\{e_1, \ldots, e_n\}$ of the generating vector space V at each point in state space M, thereby fixing the gauge freedom of $Cl(V, q)$.*

Definition 12.14 (Projected memory kernel). *Given a reference frame $\mathcal{F}$, let $V_{\mathrm{fg}}(\mathcal{F}) \subset Cl(V, q)$ be the k^*-dimensional* foreground subspace *selected by the rate-distortion optimization (Theorem 12.17). Let $\Pi_{\mathcal{F}}$ denote the orthogonal projection onto $V_{\mathrm{fg}}(\mathcal{F})$ with respect to the trace inner product $\langle A, B\rangle := \mathrm{tr}\big(A^\dagger B\big)$. The* projected memory kernel *is*

$$\mathcal{K}_{\mathcal{F}}(t, s) := \Pi_{\mathcal{F}}\, \mathcal{K}(t, s)\, \Pi_{\mathcal{F}}. \tag{12.15}$$

The complementary projection $\Pi_{\mathcal{F}}^{\perp} = \mathbf{1} - \Pi_{\mathcal{F}}$ defines the background subspace *$V_{\mathrm{bg}}(\mathcal{F})$. The decomposition $Cl(V, q) = V_{\mathrm{fg}} \oplus V_{\mathrm{bg}}$ is determined by $\mathcal{F}$, not by any* a priori *ordering of basis vectors.*

12.4.2 The Rate-Distortion Bound

We now apply rate-distortion theory to the problem of optimal memory compression under the survival constraint.

Processing rate of a frame. If the agent retains k algebraic components (the foreground subspace V_{fg}), each generating h_μ bits per unit time, the processing rate of frame $\mathcal{F}$ is

$$R_{\mathcal{F}}(k) \;=\; k \cdot h_\mu \qquad \text{(bits per unit time)}. \tag{12.16}$$

The budget constraint $R_{\mathcal{F}} \le \mathcal{C}_{\mathrm{budget}}$ thus bounds the number of maintainable components.

Definition 12.15 (Survival distortion). *The* survival distortion *of a reference frame $\mathcal{F}$ is*

$$D(\mathcal{F}) := \mathbb{E}_\xi\left[\ell\big(\mathcal{S}_{\mathrm{full}}(\xi) - \mathcal{S}_{\mathcal{F}}(\xi)\big)\right], \tag{12.17}$$

where ξ denotes environmental realizations, $\ell : \mathbb{R} \to [0, \infty)$ is a convex, non-decreasing loss function (we use squared error $\ell(x) = x^2$ throughout), $\mathcal{S}_{\mathrm{full}}(\xi)$ is the survival functional evaluated using the full memory kernel $\mathcal{K}(t, s)$, and $\mathcal{S}_{\mathcal{F}}(\xi)$ is evaluated using the projected kernel $\mathcal{K}_{\mathcal{F}}(t, s)$.

Remark 12.16 (Information-theoretic objects). *Strictly speaking, rate-distortion theory and mutual information apply to stochastic processes, not to superoperator kernels directly. Throughout Sections 12.4–12.7, $I(\mathcal{K}_{\mathcal{F}}; \mathcal{K})$ is shorthand for $I(\hat{X}; X)$, where $X = \{c_i(t)\}_{i=1}^{D}$ is the sufficient-statistic record process induced by the full kernel $\mathcal{K}$ acting on the agent's internal coordinates, and $\hat{X} = \{c_i(t)\}_{i \in V_{\mathrm{fg}}}$ is the projected record induced by $\mathcal{K}_{\mathcal{F}}$. The distortion measure* (12.17) *acts on the survival functional $\mathcal{S}$ evaluated on these records.*

Theorem 12.17 (Optimal Compression under Survival Constraint). *Let an agent with computational budget $\mathcal{C}_{\mathrm{budget}}$ and per-component entropy rate h_μ choose a reference frame $\mathcal{F}$ that minimizes the survival distortion* (12.17) *subject to $R_{\mathcal{F}} \le \mathcal{C}_{\mathrm{budget}}$* (12.16)*. Then:*

(a) *Assuming that $D(\mathcal{F})$ is non-increasing in the available rate $R_{\mathcal{F}}$ (retaining more components cannot worsen survival distortion), the set of optimal reference frames $\mathfrak{F}^* := \arg\min_{\mathcal{F}} D(\mathcal{F})$ subject to the budget constraint is non-empty, and any $\mathcal{F}^* \in \mathfrak{F}^*$ saturates the budget: $R_{\mathcal{F}^*} = \mathcal{C}_{\text{budget}}$ (the set $\mathfrak{F}^*$ may contain multiple elements; see Theorem 12.18(c)).*

(b) *The compressed representation retains*

$$k^* = \left\lfloor \frac{\mathcal{C}_{\text{budget}}}{h_\mu} \right\rfloor \tag{12.18}$$

effective algebraic components (the maximum integer number of components whose processing rate $k^* \cdot h_\mu$ fits within the budget; in practice the floor function ensures $k^* \in \mathbb{Z}_{\geq 1}$).

(c) *The fraction of algebraic structure discarded (in component count) is*

$$1 - \frac{k^*}{D}, \tag{12.19}$$

For $Cl(1,3)$ ($D = 16$) with a budget allowing $k^* = 2$, the discarded fraction is $1 - 2/16 = 87.5\%$. For $k^* = 1$, it exceeds 93%. In the regime $k^* \ll D$, the fraction approaches $1 - 1/D$ and grows with algebra dimension.

Proof. The survival functional $\mathcal{S}$ is a function of the full density operator $\rho(t)$, which in turn depends on the full memory kernel $\mathcal{K}(t,s)$. The agent's task is to evaluate $\mathcal{S}$ using only k components of $\mathcal{K}$, chosen to minimize the mean-squared error in $\mathcal{S}$.

Strictly, rate-distortion theory applies to *random processes*, not to superoperator kernels directly. The bridge is the *induced record process*: the memory kernel $\mathcal{K}(t,s)$, acting on the agent's internal coordinates, generates a D-component time series of sufficient statistics $\{c_i(t)\}_{i=1}^{D}$ whose entropy rate per component is h_μ. Rate-distortion is applied to this record stream (Section 12.2.3; cf. Tishby et al. [102]), with source $X = \{c_i(t)\}$ (the full record), reconstruction $\hat{X} = \{c_i(t)\}_{i \in V_{\text{fg}}}$ (the projected record), and distortion measure $d = |\mathcal{S}_{\text{full}} - \mathcal{S}_{\mathcal{F}}|^2$.

By Shannon's rate-distortion theorem [91], the minimum rate required to achieve distortion δ is $R(\delta)$, a convex non-increasing function. The budget constraint (12.16) limits the processing rate to $R_{\mathcal{F}} = k \cdot h_\mu \leq \mathcal{C}_{\text{budget}}$. The optimal frame $\mathcal{F}^*$ saturates this bound.

For part (b): by (12.16), tracking k components costs $k \cdot h_\mu$ bits per unit time. The maximum integer k satisfying $k \cdot h_\mu \leq \mathcal{C}_{\text{budget}}$ is $k^* = \lfloor \mathcal{C}_{\text{budget}}/h_\mu \rfloor$.

The discard fraction (c) follows by counting: k^* of D components are retained. For $Cl(1,3)$ ($D = 16$, $k^* = 2$), the discarded fraction is 87.5%; for higher-dimensional algebras it exceeds 99%. □

12.4.3 Spontaneous Symmetry Breaking

Theorem 12.18 (Necessity of Symmetry Breaking). *Under assumptions* (B1)–(B5)*, with the Computational Ceiling binding ($\tau_{\text{mem}} > \tau_{\text{par}}$, both measured in units of τ_E), and assuming* non-degeneracy*: the survival distortion* (12.17) *satisfies $D(\mathcal{F}) \neq D(\mathcal{F}')$ for* *almost all pairs $\mathcal{F} \neq \mathcal{F}'$ in the space of frames*[4]*, the agent's survival-optimal strategy requires:*

(a) ***Gauge fixing**: selection of a section σ of the gauge bundle (Definition 12.3), breaking the G-symmetry of the bare algebra.*

(b) ***Privileged decomposition**: partition of the algebra into foreground and background subspaces, $Cl(V, q) = V_{\text{fg}} \oplus V_{\text{bg}}$, with $\dim V_{\text{fg}} = k^* \ll \dim V_{\text{bg}}$.*

(c) ***Non-uniqueness**: the gauge fixing is generically* not *unique. Different initial conditions, environmental histories, or stochastic fluctuations lead to different choices of σ, just as different initial conditions in a ferromagnet lead to different magnetization directions.*

The symmetry breaking is spontaneous *in the precise physical sense: the underlying algebra $Cl(V, q)$ retains its full G-symmetry, but the agent's operational representation necessarily breaks it.*

Proof. By Theorem 12.7, symmetric processing leads to paralysis at τ_{par}. By assumption (B4) (survival imperative), the agent must maintain $\mathcal{S} \geq \mathcal{S}_{\min}$ beyond τ_{par}. This requires evaluating $\mathcal{S}$ within the computational budget $\mathcal{C}_{\text{budget}}$, which by Theorem 12.17 requires projecting onto $k^* < \dim Cl(V, q)$ components.

Such a projection *is* a gauge fixing: it selects k^* basis vectors $\{e_1, \ldots, e_{k^*}\}$ from the generating space V, thereby breaking the G-invariance that treats all bases equivalently.

Part (b) follows from the definition of the projected kernel (Definition 12.14). Part (c) follows from the non-degeneracy assumption: the rate-distortion optimization (Theorem 12.17) generically admits finitely many local minima. Different initial conditions or environmental histories select different minima, analogous to the spontaneous magnetization of a ferromagnet below T_c. The breaking is *spontaneous*: the algebra retains G-symmetry, but any operational solution breaks it. □

Remark 12.19 (Terminology). *We use "symmetry breaking" by analogy with condensed matter physics, where the term has precise mathematical content (degenerate ground states, Goldstone modes, divergent susceptibility). Here the "symmetry" is the gauge freedom of the Clifford algebra, and "breaking" it amounts to choosing a reference frame under resource constraints. More conservative terminology—"reference frame selection" or "basis choice under computational constraints"—may be appropriate until formal SSB criteria (Landau-type free energy with degenerate minima) are established for this setting.*

12.4.4 The Four Bias Terms

Proposition 12.20 (Structure of the Broken Phase). *When gauge symmetry is broken by a reference frame $\mathcal{F}$, the agent's operational representation acquires four systematic deviations from the symmetric phase:*

[4] Non-degeneracy is generically satisfied when the environment's pointer basis [122] assigns different survival values to different algebraic components, breaking the continuous symmetry of the distortion landscape. In degenerate cases, a finite set of local minima may coexist—multiple "ego attractors"—analogous to the discrete magnetization directions in a crystal-field anisotropic ferromagnet.

(i) ***Basis selection bias*** *($\mathcal{B}_{\text{select}}$): The choice of $\{e_1, \ldots, e_{k^*}\}$ privileges certain algebraic components over others. Information aligned with the chosen basis is processed efficiently; misaligned information is discarded or distorted.* Observable consequence: *systematic blindness to off-basis environmental perturbations (orthogonal masking).*

(ii) ***Frame drag*** *($\mathcal{B}_{\text{frame}}$): The connection on the gauge bundle (Section 12.2.5) induces a systematic preference for states near the current gauge choice. The agent's predictions are biased toward confirming its existing frame.* Observable consequence: *hysteresis in belief updating; the agent's model lags behind rapid environmental shifts.*

(iii) ***Objective centering*** *($\mathcal{B}_{\text{center}}$): The survival functional $\mathcal{S}$, when evaluated in the projected basis, becomes centered on the agent's own state rather than a global optimum. The agent optimizes* locally *within its frame.* Observable consequence: *inability to detect global survival optima located in the background subspace.*

(iv) ***Model incompleteness*** *($\mathcal{B}_{\text{inc}}$): The compression from $Cl(V, q)$ to V_{fg} is lossy. The discarded components V_{bg} contain correlations that are invisible to the agent but physically real.* Observable consequence: *systematic underestimation of total thermodynamic uncertainty (overconfidence).*

Proof. (i) follows directly from the definition of the projection $\Pi_{\mathcal{F}}$: components orthogonal to the selected basis are annihilated.

(ii) The parallel transport of the gauge connection preserves the agent's basis choice along its trajectory. Under perturbation, the connection's holonomy creates a restoring "force" toward the established frame—a systematic confirmation bias.

(iii) In the projected representation, $\mathcal{S}_{\mathcal{F}}$ is a function of the k^*-dimensional foreground state only. The gradient $\nabla\mathcal{S}_{\mathcal{F}}$ lies entirely in V_{fg}, so the agent's optimization is blind to directions in V_{bg}. This is equivalent to centering the objective function on the agent's own representational subspace.

(iv) By Theorem 12.17(c), a fraction $\geq 1 - k^*/\dim Cl(V, q)$ of information is discarded. The discarded components exist physically (they contribute to $\mathcal{S}_{\text{full}}$) but are invisible to the agent's evaluation of $\mathcal{S}_{\mathcal{F}}$. $\square$

Bias	Origin	Observable consequence	Determines
$\mathcal{B}_{\text{select}}$ (selection)	projection $\Pi_{\mathcal{F}}$	Systematic blindness to off-basis perturbations (orthogonal masking)	*what* is seen
$\mathcal{B}_{\text{frame}}$ (frame drag)	bundle connection / holonomy	Hysteresis in belief updating; model lags behind rapid drift	*duration*
$\mathcal{B}_{\text{center}}$ (centering)	$\nabla\mathcal{S} \in V_{\text{fg}}$	Local frame-relative optima; global background optima invisible	*target*
$\mathcal{B}_{\text{inc}}$ (incompleteness)	lossy compression $k^* \ll D$	Underestimation of thermodynamic uncertainty (structural overconfidence)	*blind spot*

Table 12.1: The four bias terms of the broken phase. All four are generic consequences of gauge fixing under assumptions (B1)–(B5).

Remark 12.21 (Nature of the bias terms). *The four bias terms (Table 12.1) are not pathologies—they are* generic *consequences of gauge fixing under bounded computation. Any agent satisfying (B1)–(B5) acquires all four.*

12.5 Emergent Structure: The Architecture of Ego

We consolidate the gauge-fixed compressed representation into a single mathematical object. Throughout this section, "ego" is used purely as shorthand for a gauge-fixed compressed representation; no claims about phenomenal consciousness, subjective experience, or qualia are intended or implied.

Definition 12.22 (Ego). *The* ego *of an agent satisfying* (B1)–(B5) *is the pair*

$$\mathfrak{E} := \big(\mathcal{F}^*,\ V_{\text{fg}}^*\big), \tag{12.20}$$

where $\mathcal{F}^* \in \mathfrak{F}^*$ (Theorem 12.17) is the chosen gauge (providing the coordinate system) and $V_{\text{fg}}^* := V_{\text{fg}}(\mathcal{F}^*)$ is the k^*-dimensional foreground subspace selected by the rate-distortion bound (providing the compression). The projected memory kernel $\mathcal{K}_{\mathfrak{E}} := \Pi_{V_{\text{fg}}^*}\, \mathcal{K}\, \Pi_{V_{\text{fg}}^*}$ is induced by this pair. All bias terms, distortion bounds, and delusion dynamics are functions of $\mathfrak{E}$.

12.5.1 The Ego as a Fiber Bundle Section

The reference frame $\mathcal{F}$, understood as a section $\sigma : M \to P$, is the mathematical object we call the *ego*. It has three key properties:

Smoothness. The section σ varies continuously with the agent's state $\rho \in M$. Small changes in ρ produce small changes in the preferred basis—the ego is not a discrete switch but a smooth deformation of perspective.

Holonomy. If the agent's state traces a closed loop $\gamma : [0,1] \to M$ with $\gamma(0) = \gamma(1) = \rho_0$, the parallel-transported frame need not return to its initial value:

$$\sigma(\gamma(1)) = \mathrm{Hol}(\gamma) \cdot \sigma(\gamma(0)), \tag{12.21}$$

where $\mathrm{Hol}(\gamma) \in G$ is the holonomy of the connection around γ. Non-trivial holonomy means the agent can "learn"—its reference frame shifts after a complete cycle of experience.

Topological obstruction. In general, a *global* section $\sigma : M \to P$ may not exist. The obstruction is measured by the characteristic classes of the bundle P. When a global section does not exist, the ego must have "singularities"—states where the preferred basis is undefined or discontinuous. This connects to the crisis of Paper III: the delusion trap can be understood as the agent approaching a topological obstruction of its own reference frame.

12.5.2 The Effective Survival Functional

Proposition 12.23 (Survival decomposition). *In the broken phase, the survival functional decomposes as*

$$\mathcal{S} = \mathcal{S}_{\mathrm{vis}}(\mathcal{F}) + \mathcal{S}_{\mathrm{hid}}(\mathcal{F}), \tag{12.22}$$

where:

- $\mathcal{S}_{\mathrm{vis}}(\mathcal{F})$ *is the contribution from the foreground subspace* V_{fg}*, computable within the agent's reference frame;*
- $\mathcal{S}_{\mathrm{hid}}(\mathcal{F})$ *is the contribution from the background subspace* V_{bg}*, invisible to the agent.*

The agent maximizes $\mathcal{S}_{\mathrm{vis}}$ *while being structurally blind to* $\mathcal{S}_{\mathrm{hid}}$*.*

Proof. The survival functional $\mathcal{S} = \Delta F - W$ depends on $\rho(t)$, which is a function of the full memory kernel $\mathcal{K}(t,s)$. Decomposing $\mathcal{K} = \Pi_{\mathcal{F}}\,\mathcal{K}\,\Pi_{\mathcal{F}} + \Pi_{\mathcal{F}}^{\perp}\,\mathcal{K}\,\Pi_{\mathcal{F}}^{\perp} + \text{cross terms}$, the leading contributions are $\mathcal{S}_{\mathrm{vis}} := \mathcal{S}[\Pi_{\mathcal{F}}\,\mathcal{K}\,\Pi_{\mathcal{F}}]$ and $\mathcal{S}_{\mathrm{hid}} := \mathcal{S} - \mathcal{S}_{\mathrm{vis}}$ (collecting background and cross terms). The agent computes only $\mathcal{S}_{\mathrm{vis}}$, as the projected kernel $\mathcal{K}_{\mathcal{F}}$ discards all background components. □

12.5.3 The Computational Speedup

Proposition 12.24 (Ego dividend). *After symmetry breaking, the computational cost of processing memory drops from* $\mathcal{C}_{\mathrm{proc}} \sim h_\mu \cdot \tau_{\mathrm{mem}} \cdot D$ *(symmetric case,* $D = \dim Cl(V,q)$*) to*

$$\mathcal{C}_{\mathrm{proc}}^{(\mathcal{F})} \sim h_\mu \cdot \tau_{\mathrm{mem}} \cdot k^*. \tag{12.23}$$

The speedup factor is

$$\frac{D}{k^*} = \frac{2^n}{k^*}. \tag{12.24}$$

This is the computational advantage of reference-frame selection. For $Cl(1,3)$ ($D = 16$) with $k^* = 2$, the speedup is $8\times$. For higher-dimensional algebras, the speedup grows exponentially in n.

12.5.4 The Ego-Entropy Trade-off

Theorem 12.25 (Ego-Entropy Trade-off). *Let $X = \{c_i(t)\}_{i=1}^D$ denote the full stochastic record process induced by the memory kernel $\mathcal{K}$ on the agent's internal coordinates, and let $\hat{X} = \{c_i(t)\}_{i\in V_{\text{fg}}}$ denote the projected record retained by the ego. The mutual information between compressed and full records, denoted $I(\mathcal{K}_{\mathcal{F}};\,\mathcal{K}) \equiv I(\hat{X};\,X)$, satisfies*

$$I(\hat{X};\,X) \;\le\; H(\hat{X}) \;\le\; k^* \cdot h_\mu \cdot \tau_{\text{mem}}. \tag{12.25}$$

Under the additional assumption that I_{pred} (12.6) is approximately uniformly distributed across the D algebraic components in the symmetric phase[5]*, the information discarded by the ego is bounded below (up to $O(1)$ constants under uniformity):*

$$I_{\text{discarded}} := H(X) - I(\hat{X};\,X) \gtrsim \left(1 - \frac{k^*}{D}\right) \cdot I_{\text{pred}}. \tag{12.26}$$

Proof. By the data processing inequality, $I(\hat{X};\,X) \le H(\hat{X})$. The projected record $\hat{X}$ has k^* components, each carrying at most h_μ bits per unit time over a window of τ_{mem}, giving $H(\hat{X}) \le k^* \cdot h_\mu \cdot \tau_{\text{mem}}$ [25]. This yields (12.25). The total predictive information in the full record is I_{pred} (12.6). Under the uniformity assumption, each of the D components carries $\sim I_{\text{pred}}/D$, so the k^* retained components account for $\sim (k^*/D)\, I_{\text{pred}}$. The discarded fraction follows by subtraction. □

Remark 12.26 (Uniformity assumption). *The bound $I_{\text{discarded}} \ge (1 - k^*/D)\, I_{\text{pred}}$ holds under the uniformity assumption (each component carries $\sim I_{\text{pred}}/D$ predictive information). In practice, predictive information is often concentrated in a few dominant modes; in such cases the discarded fraction may be substantially smaller, making gauge fixing less costly.*

Remark 12.27 (The price of selfhood). *Equation (12.26) quantifies the* information cost of having an ego*: the agent sacrifices at least a fraction $1 - k^*/\dim Cl(V,q)$ of all predictive information about its environment in exchange for computational tractability. This is not a deficiency—it is a* design constraint *forced by bounded resources. The ego is the optimal lossy compression under survival weighting.*

12.6 Worked Example: Qubit in a Two-Channel Bath

12.6.1 Model Setup

We extend Paper I's spin-boson model to demonstrate symmetry breaking explicitly. Consider a qubit ($\dim \mathcal{H}_S = 2$) with internal algebra $Cl(0,2) \cong \mathbb{H}$ (the quaternions, $\dim = 4$).

[5] This "uniformity assumption" is the information-theoretic counterpart of the unstructured-environment condition in Theorem 12.7. When some components carry disproportionately more predictive information, the bound tightens or loosens depending on the alignment between V_{fg} and the high-information subspace.

Symbol mapping. The general framework of Sections 12.3–12.5 specialises as follows:

General	This example	Value
$Cl(V,q)$	$Cl(0,2) \cong \mathbb{H}$	$D = 4$
$G = \mathrm{Aut}(Cl(V,q))$	$SO(3)$	acting on $\{\mathbf{i},\mathbf{j},\mathbf{k}\}$
$\mathcal{C}_{\text{budget}}$	$2\,h_\mu$	bits/time
k^* (Thm. 12.17)	$\lfloor 2h_\mu/h_\mu \rfloor = 2$	components
V_{fg}	$\text{span}\{1,\mathbf{k}\}$	dephasing subspace
V_{bg}	$\text{span}\{\mathbf{i},\mathbf{j}\}$	dissipative subspace
τ_{par} (Thm. 12.7)	$2h_\mu/(4h_\mu) = 0.5$	ω_0^{-1}

The qubit is coupled to a bosonic environment through *two* independent channels:

- A *dephasing channel* via σ_z, with spectral density
$$J_z(\omega) = \frac{2\lambda_z\,\gamma_z\,\omega}{\omega^2 + \gamma_z^2} \qquad \text{(Lorentz–Drude)}, \tag{12.27}$$
producing a memory kernel $\mathcal{K}_z(t,s)$ with non-Markovian backflow.
- A *dissipative channel* via σ_x, with spectral density
$$J_x(\omega) = \frac{2\lambda_x\,\gamma_x\,\omega}{\omega^2 + \gamma_x^2} \qquad \text{(Lorentz–Drude)}, \tag{12.28}$$
producing a memory kernel $\mathcal{K}_x(t,s)$.

The full memory kernel is $\mathcal{K}(t,s) = \mathcal{K}_z(t,s) \oplus \mathcal{K}_x(t,s)$, and the quaternionic algebra $\mathbb{H} = \text{span}\{1,\mathbf{i},\mathbf{j},\mathbf{k}\}$ has automorphism group $G = \mathrm{Aut}(\mathbb{H}) \cong SO(3)$ (rotations of the pure quaternion subspace).

Parameters. We set $\omega_0 = 1$ (energy unit), $\lambda_z = 1$, $\gamma_z = 0.5$ (underdamped, strong non-Markovian effects in the dephasing channel), $\lambda_x = 0.3$, $\gamma_x = 5.0$ (overdamped, approximately Markovian in the dissipative channel), and the low-temperature regime $\beta\omega_0 \gg 1$.

Computational budget. The agent has $\mathcal{C}_{\text{budget}} = 2\,h_\mu$ bits per unit time—sufficient to track two components of $\mathbb{H}$ but not all four.

Parameter-to-theorem mapping. Table 12.2 collects the example parameters and confirms that the Computational Ceiling binds.

12.6.2 The Unbroken Phase: Paralysis

In the symmetric phase, the agent tracks all four quaternionic components $\{1,\mathbf{i},\mathbf{j},\mathbf{k}\}$ simultaneously. The computational cost is

$$\mathcal{C}_{\text{proc}} = h_\mu \cdot \tau_{\text{mem}} \cdot D = 4\,h_\mu \cdot \tau_{\text{mem}}, \qquad D := \dim Cl(0,2) = 4. \tag{12.29}$$

The paralysis time is

$$\tau_{\text{par}} = \frac{\mathcal{C}_{\text{budget}}}{h_\mu \cdot D} = \frac{2\,h_\mu}{4\,h_\mu} = 0.5 \qquad (\text{in units of } \omega_0^{-1}). \tag{12.30}$$

Beyond $\tau_{\text{mem}} = 0.5\,\omega_0^{-1}$, the agent cannot process both channels simultaneously—it is paralyzed.

Quantity	Symbol	Value	Theorem check
Full dimension	D	4	Thm. 12.7
Entropy rate	h_μ	1.0 (normalised)	per-component rate
Budget	$\mathcal{C}_{\text{budget}}$	$2\,h_\mu$	Def. 12.2
Ceiling check	$h_\mu D$ vs $\mathcal{C}_{\text{budget}}$	$4 > 2$	**ceiling binds**
Optimal k	k^*	$\lfloor 2/1 \rfloor = 2$	Thm. 12.17(b)
Discard fraction	$1 - k^*/D$	$1/2 = 50\%$	Thm. 12.25
Paralysis time	τ_{par}	$2/(4) = 0.5$	Eq. (12.12)

Table 12.2: Parameter mapping for the two-channel qubit example. The ceiling check confirms that symmetry breaking is necessary; the budget is exactly saturated after breaking ($R_{\mathcal{F}} = k^* h_\mu = \mathcal{C}_{\text{budget}}$).

12.6.3 Symmetry Breaking: Choosing σ_z

The agent breaks the $SO(3)$ symmetry of $\mathbb{H}$ by selecting σ_z as the privileged basis direction, retaining the $\{1, \mathbf{k}\}$ subspace (the dephasing channel) as foreground and discarding $\{\mathbf{i}, \mathbf{j}\}$ (the dissipative channel) as background:

$$\mathbb{H} = \underbrace{\text{span}\{1, \mathbf{k}\}}_{V_{\text{fg}}\ (k^*=2)} \oplus \underbrace{\text{span}\{\mathbf{i}, \mathbf{j}\}}_{V_{\text{bg}}}. \tag{12.31}$$

Why σ_z? The dephasing channel ($\lambda_z = 1$, $\gamma_z = 0.5$) is strongly non-Markovian and carries the dominant survival-relevant information (the backflow revivals that enable $\mathcal{S} > 0$, as demonstrated in Paper I). The dissipative channel ($\lambda_x = 0.3$, $\gamma_x = 5.0$) is approximately Markovian and contributes primarily to decoherence—its survival value is negative.

This choice coincides with the *pointer basis* selected by environmental decoherence (quantum Darwinism [122]): the σ_z eigenstates are the states that survive decoherence and become redundantly encoded in the environment. The ego "accepts the suggestion" of decoherence, aligning its computational resources with the environmentally stable basis.

12.6.4 The Broken Phase: Effective Processing

In the broken phase, the projected memory kernel $\mathcal{K}_{\mathcal{F}} = \mathcal{K}_z$ retains only the dephasing-channel dynamics. The computational cost drops to

$$\mathcal{C}_{\text{proc}}^{(\mathcal{F})} = h_\mu \cdot \tau_{\text{mem}} \cdot k^* = 2\,h_\mu \cdot \tau_{\text{mem}}, \tag{12.32}$$

exactly half the symmetric cost (12.29). The agent can now process memory up to depth $\tau_{\text{mem}} = 1\,\omega_0^{-1}$ before reaching its budget—twice the paralysis time.

The survival functional in the broken phase is

$$\mathcal{S}_{\text{vis}}(\mathcal{F}) = \mathcal{S}[\mathcal{K}_z], \tag{12.33}$$

which, as shown in Paper I, achieves $\beta\,\mathcal{S}_{\text{vis}} \approx +0.093$ at the first backflow revival.

The hidden component $\mathcal{S}_{\text{hid}} = \mathcal{S}[\mathcal{K}_x]$ is the survival contribution from the dissipative channel, which the agent can no longer evaluate. For the chosen parameters, $|\mathcal{S}_{\text{hid}}| \ll |\mathcal{S}_{\text{vis}}|$ (the dissipative channel contributes primarily negative survival value), so the distortion is small.

12.6.5 Quantitative Evaluation

We now evaluate the ego dividend explicitly. Each channel's decoherence function follows from the exact $T \to 0$ solution of the Lorentz–Drude pure-dephasing model [18, 143]:

$$p_\alpha(t) = e^{-\gamma_\alpha t/2}\left[\cos(\Omega_\alpha t) + \frac{\gamma_\alpha}{2\Omega_\alpha}\sin(\Omega_\alpha t)\right], \quad \Omega_\alpha := \tfrac{1}{2}\sqrt{4\lambda_\alpha\gamma_\alpha - \gamma_\alpha^2}, \tag{12.34}$$

for $\alpha \in \{z, x\}$. When $4\lambda_\alpha\gamma_\alpha < \gamma_\alpha^2$ (the overdamped regime), Ω_α becomes imaginary and the trigonometric functions are replaced by hyperbolic functions (monotonic decay, no backflow).

For our parameters:

- z-**channel** ($\lambda_z = 1$, $\gamma_z = 0.5$): $\Omega_z = \frac{1}{2}\sqrt{1.75} \approx 0.661$. Underdamped; $|p_z(t)|$ exhibits oscillatory backflow.
- x-**channel** ($\lambda_x = 0.3$, $\gamma_x = 5.0$): Discriminant $4\lambda_x\gamma_x - \gamma_x^2 = 6 - 25 = -19 < 0$. Overdamped; $|p_x(t)|$ decays monotonically with no backflow.

The survival proxy from Paper I, $\beta\,\mathcal{S} \propto |p(t)|^2 - 1$ (valid for the pure-dephasing model with maximally coherent initial state and pointer-basis measurement), applies to each channel independently. Backflow intervals—where $d|p_\alpha|/dt > 0$—produce $\mathcal{S} > 0$ over those subintervals (Paper I, Theorem 2).

Key result. For the z-channel with $\gamma_z = 0.5$, the first backflow interval begins at $t^* \approx 2.9\,\omega_0^{-1}$—well after the paralysis time $\tau_{\rm par} = 0.5\,\omega_0^{-1}$. The symmetric agent, paralyzed at $\tau_{\rm par}$, can harvest *zero* backflow. The ego agent, tracking only the z-channel, can process memory to depth $1\,\omega_0^{-1}$ and exploits *all three* backflow revivals visible in Figure 12.1(a).

The cumulative backflow harvested by the ego agent (Figure 12.1(b)) totals approximately 0.10 (in dimensionless $\beta\,\mathcal{S}$ units) over $t \in [0, 15\,\omega_0^{-1}]$. The symmetric agent harvests exactly zero. This infinite ratio is the *ego dividend*: the entire non-Markovian survival advantage is accessible only to the agent that has broken symmetry.

Crucially, visual inspection of Figure 12.1(a) reveals a timeline of tragedy for the symmetric agent. The paralysis time $\tau_{\rm par} = 0.5$ occurs *before* the onset of the first backflow interval ($t^* \approx 2.9$). The symmetric agent is computationally dead before the environment offers its first gift. The ratio of survival profit is not merely large; it is singular. In this framework, to remain symmetric is to starve in the midst of plenty.

Consistency check. We verify the Computational Ceiling (Theorem 12.7) directly: $\mathcal{R}_{\rm proc}^{\rm sym} = h_\mu \cdot D = 4\,h_\mu > \mathcal{C}_{\rm budget} = 2\,h_\mu$, confirming that the ceiling binds and symmetry breaking is required. After breaking ($k^* = 2$), $R_{\mathcal{F}} = 2\,h_\mu = \mathcal{C}_{\rm budget}$: the budget is exactly saturated, as predicted by Theorem 12.17(a).

12.6.6 The Pointer-State Connection

The optimal basis choice coincides with the einselection (environment-induced superselection) basis of decoherence theory [122]. This is not a coincidence: the pointer states are precisely those that generate the most redundant records in the environment—i.e., the most predictive correlations. The rate-distortion optimization (Theorem 12.17) selects the components with the highest survival value per bit, which are generically the pointer-state components.

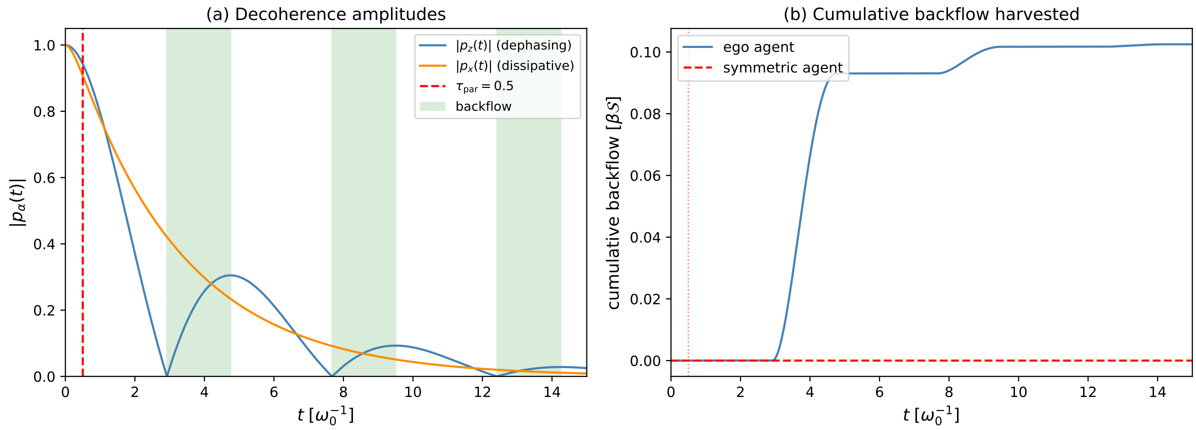

Figure 12.1: Two-channel qubit model (Section 12.6) with Lorentz–Drude spectral density. **Parameters:** $\omega_0 = 1$ (energy unit); $\lambda_z = 1$, $\gamma_z = 0.5$ (dephasing, non-Markovian); $\lambda_x = 0.3$, $\gamma_x = 5.0$ (dissipative, ∼Markovian); $\mathcal{C}_{\text{budget}} = 2\, h_\mu$. **Units:** time in ω_0^{-1}. **Regime:** low temperature ($\beta\omega_0 \gg 1$); using the standard $T \to 0$ analytic expression (12.34) [18]. **(a)** Decoherence amplitudes $|p_z(t)|$ (blue, non-Markovian, with backflow in green bands) and $|p_x(t)|$ (orange, monotonic decay). Red dashed: paralysis time $\tau_{\text{par}} = 0.5$. **(b)** Cumulative backflow harvested. Blue: ego agent (broken $\to \sigma_z$) exploits all three revival intervals. Red dashed: symmetric agent, paralyzed at τ_{par}, harvests zero—all backflow occurs after paralysis onset. The growing gap is the *ego dividend*.

Remark 12.28 (Decoherence as symmetry-breaking catalyst)**.** *The environment does not* force *a specific gauge fixing; it merely breaks the degeneracy among possible fixings by making some bases more informationally efficient than others. The agent's bounded computation does the rest: once the degeneracy is broken, the survival imperative (B4) selects the pointer-aligned frame as optimal. This is the precise sense in which decoherence "catalyzes" the spontaneous symmetry breaking of the ego.*

12.7 The Cost of Ego

The ego resolves the computational crisis of Section 12.3, but it introduces a new vulnerability. A fixed reference frame is a *static* gauge choice in a *dynamic* environment. If the environment changes, the ego becomes progressively maladaptive.

Drift layer. Environmental change can occur at multiple levels: parameter drift ($\lambda_\alpha(t)$, $\gamma_\alpha(t)$), spectral-density deformation ($J(\omega, t)$), or full process-distribution shift ($P_t(X)$). For analytical tractability, we model drift at the *spectral-density parameter level* throughout this section; the results generalise monotonically to deeper levels (faster drift $\Rightarrow$ shorter t_{del}).

12.7.1 The Rigidity Trap

Proposition 12.29 (Frame Rigidity under Drift)**.** *Let the environment undergo slow drift: the spectral density parameters change as $\lambda_\alpha(t) = \lambda_\alpha^{(0)} + \varepsilon\, f_\alpha(t)$ for $\alpha \in \{z, x\}$, with drift rate $\varepsilon > 0$. The optimal reference frame $\mathcal{F}^*(t)$ (the instantaneous minimizer of survival distortion) rotates continuously in the gauge group G.*

If the agent's reference frame $\mathcal{F}$ is held fixed (no recalibration), the mismatch between $\mathcal{F}$ and $\mathcal{F}^*(t)$ grows as

$$\delta(t) := d_G\big(\mathcal{F},\, \mathcal{F}^*(t)\big) \sim \varepsilon \int_0^t \big|\dot{f}(s)\big|\, ds, \tag{12.35}$$

where d_G is the geodesic distance in the gauge group.

Proof. The instantaneous optimal frame $\mathcal{F}^*(t)$ is a continuous function of the spectral density parameters $\{\lambda_\alpha(t), \gamma_\alpha(t)\}$. Under the drift $\lambda_\alpha(t) = \lambda_\alpha^{(0)} + \varepsilon\, f_\alpha(t)$, the chain rule gives $\dot{\mathcal{F}}^*(t) = \varepsilon \sum_\alpha (\partial \mathcal{F}^*/\partial \lambda_\alpha)\, \dot{f}_\alpha(t)$. Integrating and taking the norm in G gives the bound (12.35). □

12.7.2 Stylized Drift Model

To quantify the collapse of a fixed frame, we introduce a minimal drift model that makes the exponential divergence and the logarithmic delusion time algebraically explicit.

Definition 12.30 (Rotating optimal frame)**.** *Let the mismatch angle $\theta(t)$ between the agent's fixed frame $\mathcal{F}$ and the instantaneous optimal frame $\mathcal{F}^*(t)$ evolve as*

$$\theta(t) = \theta_0\, e^{\Lambda t} \qquad \textit{(chaotic drift)}, \tag{12.36}$$

where $\theta_0 \in (0, \pi/4)$ is the initial misalignment (so that $t_{\mathrm{del}} > 0$) and $\Lambda > 0$ is the environmental Lyapunov exponent (the rate at which nearby environmental trajectories diverge in spectral-density space). Operationally, Λ is determined by the drift rate ε and the adaptation timescale τ_{adapt} of the spectral-density parameters via the scaling

$$\Lambda \sim \frac{\varepsilon}{\tau_{\mathrm{adapt}}}\,; \tag{12.37}$$

cf. (12.35)*. For slow linear drift ($\theta(t) = \varepsilon\, t$, $\Lambda \to 0$), the crossover time is $t_{\mathrm{del}} = \pi/(4\varepsilon)$ (Remark 12.33).*

The visible and hidden survival components decompose geometrically:

$$\mathcal{S}_{\mathrm{vis}}(t) = \mathcal{S}_{\mathrm{tot}} \cos^2\theta(t), \qquad \mathcal{S}_{\mathrm{hid}}(t) = \mathcal{S}_{\mathrm{tot}} \sin^2\theta(t), \tag{12.38}$$

where $\mathcal{S}_{\mathrm{tot}}$ is the full survival functional (invariant under frame rotation).

12.7.3 The Prediction Error Divergence

Proposition 12.31 (Divergence of Hidden Survival)**.** *Under the drift model* (12.36)–(12.38)*, the hidden survival component grows as*

$$\big|\mathcal{S}_{\mathrm{hid}}(t)\big| = |\mathcal{S}_{\mathrm{tot}}|\, \sin^2\!\big(\theta_0\, e^{\Lambda t}\big). \tag{12.39}$$

For small angles ($\theta_0 e^{\Lambda t} \ll 1$): $|\mathcal{S}_{\mathrm{hid}}| \approx |\mathcal{S}_{\mathrm{tot}}|\, \theta_0^2\, e^{2\Lambda t}$ (exponential growth).

Proof. Direct substitution of (12.36) into (12.38). The small-angle expansion $\sin^2\theta \approx \theta^2$ gives the exponential form. □

12.7.4 The Delusion Trap

Theorem 12.32 (The Delusion Trap). *Under* (B1)–(B5) *with the drift model* (12.36) *and initial misalignment $\theta_0 \in (0, \pi/4)$, an agent with a fixed reference frame $\mathcal{F}$ reaches a critical* ***delusion time***

$$t_{\mathrm{del}} = \frac{1}{\Lambda} \ln\left(\frac{\pi/4}{\theta_0}\right), \tag{12.40}$$

beyond which:

(a) $|\mathcal{S}_{\mathrm{hid}}(t)| > |\mathcal{S}_{\mathrm{vis}}(t)|$*: the invisible component dominates the survival functional.*

(b) *The agent's update direction becomes* anti-correlated *with the true optimal direction: the inner product of survival gradients (with respect to the agent's control variables $u \in V_{\mathrm{fg}}$) satisfies*

$$\left\langle \nabla_u \mathcal{S}_{\mathrm{vis}},\ \nabla_u \mathcal{S}_{\mathrm{full}} \right\rangle < 0. \tag{12.41}$$

Updating u to maximise $\mathcal{S}_{\mathrm{vis}}$ actually decreases $\mathcal{S}_{\mathrm{full}}$.

(c) *The agent cannot detect this failure from within its own reference frame, because all four bias terms ($\mathcal{B}_{\mathrm{select}}$, $\mathcal{B}_{\mathrm{frame}}$, $\mathcal{B}_{\mathrm{center}}$, $\mathcal{B}_{\mathrm{inc}}$) operate within V_{fg} and cannot register changes in V_{bg}.*

Proof. Part (a): The crossover $|\mathcal{S}_{\mathrm{hid}}| = |\mathcal{S}_{\mathrm{vis}}|$ occurs when $\sin^2\theta = \cos^2\theta$, i.e., $\theta(t_{\mathrm{del}}) = \pi/4$. Substituting (12.36): $\theta_0\, e^{\Lambda\, t_{\mathrm{del}}} = \pi/4$, which gives (12.40). The logarithmic dependence on $1/\theta_0$ means that even a very small initial misalignment ($\theta_0 \sim 10^{-3}$) delays the trap only by $\sim 7/\Lambda$—a modest multiple of the environmental Lyapunov time.

Part (b): Decompose $\mathcal{S}_{\mathrm{full}}$ in the agent's (rotated) coordinates as $\mathcal{S}_{\mathrm{full}}(u, v) = \mathcal{S}_{\mathrm{ff}}(u) + \mathcal{S}_{\mathrm{fb}}(u, v) + \mathcal{S}_{\mathrm{bb}}(v)$, where $u \in V_{\mathrm{fg}}$, $v \in V_{\mathrm{bg}}$, $\mathcal{S}_{\mathrm{ff}}$ depends only on foreground controls, $\mathcal{S}_{\mathrm{bb}}$ only on background, and $\mathcal{S}_{\mathrm{fb}}$ encodes the foreground–background cross-coupling. The projected kernel $\mathcal{K}_{\mathcal{F}} = \Pi_{\mathcal{F}}\, \mathcal{K}\, \Pi_{\mathcal{F}}$ discards $\mathcal{S}_{\mathrm{fb}}$ and $\mathcal{S}_{\mathrm{bb}}$, so $\mathcal{S}_{\mathrm{vis}}(u) = \mathcal{S}_{\mathrm{ff}}(u)$. The foreground gradients are therefore $\nabla_u \mathcal{S}_{\mathrm{vis}} = \nabla_u \mathcal{S}_{\mathrm{ff}}$ and $\nabla_u \mathcal{S}_{\mathrm{full}} = \nabla_u \mathcal{S}_{\mathrm{ff}} + \nabla_u \mathcal{S}_{\mathrm{fb}}$, giving

$$\left\langle \nabla_u \mathcal{S}_{\mathrm{vis}},\ \nabla_u \mathcal{S}_{\mathrm{full}} \right\rangle = \left|\nabla_u \mathcal{S}_{\mathrm{ff}}\right|^2 + \left\langle \nabla_u \mathcal{S}_{\mathrm{ff}},\ \nabla_u \mathcal{S}_{\mathrm{fb}} \right\rangle.$$

The first term is non-negative. Under the rotating drift model, the frame rotation by θ induces cross-coupling that scales as $\sin\theta\cos\theta$ (maximal at $\theta = \pi/4$), while the direct term scales as $\cos^4\theta$. The sign of $\nabla_u \mathcal{S}_{\mathrm{fb}}$ is set by the background state v: since the agent invests no control resources in V_{bg}, v relaxes toward the uncontrolled equilibrium, where the cross-coupling penalises foreground-directed updates ($\langle \nabla_u \mathcal{S}_{\mathrm{ff}},\, \nabla_u \mathcal{S}_{\mathrm{fb}} \rangle < 0$). Beyond t_{del} ($\theta > \pi/4$), the adverse coupling dominates the direct term, yielding (12.41).

The sign of the cross-coupling term $\nabla_u \mathcal{S}_{\mathrm{fb}}$ depends on the specific drift model (rotating optimal frame with $\cos^2/\sin^2$ decomposition). For a general Clifford algebra, the cross terms between foreground and background could have either sign depending on environmental correlations. The anti-correlation result is therefore established for the drift model of Section 12.7.2 and conjectured to hold generically under the standing assumptions.

Part (c): The bias terms $\mathcal{B}_{\mathrm{select}}$ through $\mathcal{B}_{\mathrm{inc}}$ (Proposition 12.20) are defined *within* V_{fg}. The agent's performance metric $\mathcal{S}_{\mathrm{vis}} = \mathcal{S}_{\mathrm{tot}} \cos^2\theta$ decreases only at second order in θ, so it remains positive and shows no anomaly until θ is already $O(1)$. The growing signal in V_{bg} maps to the null space of $\Pi_{\mathcal{F}}$ and is strictly invisible. □

Remark 12.33 (Linear drift limit). *For slow linear drift ($\theta(t) = \varepsilon\, t$, $\Lambda \to 0$), the crossover occurs at $t_{\rm del} = \pi/(4\varepsilon)$. With $\varepsilon = 0.01\,\omega_0$, $t_{\rm del} \approx 79\,\omega_0^{-1}$—long enough for the agent to accumulate a false sense of security, yet short on environmental timescales.*

Remark 12.34 (Why dithering does not help). *One might ask whether the agent could escape the delusion trap by randomly "probing" the background subspace $V_{\rm bg}$—temporarily rotating its frame to sample hidden components. This fails for two reasons. First, each probe costs $\sim h_\mu \cdot D$ bits of computation (the Symmetry Tax, Corollary 12.9), directly competing with the budget allocated to foreground processing. Second—and more fundamentally—the agent has no* gradient signal *to indicate* when *or* where *to probe. As long as $|\mathcal{S}_{\rm hid}| < |\mathcal{S}_{\rm vis}|$ (pre-delusion), the in-frame performance metric $\mathcal{S}_{\rm vis}$ shows no anomaly. The exponential divergence* (12.39) *is invisible until it dominates—at which point it is too late. Systematic correction requires monitoring the* rate of change *of prediction error, which is a second-order operation: the subject of Paper III.*

Remark 12.35 (The ego as medicine and poison). *The ego cures computational paralysis (Theorem 12.7) but creates the delusion trap (Theorem 12.32). It is simultaneously the* medicine *for Paper I's crisis and the* poison *that generates Paper III's crisis. This duality is a structural consequence of the irreversible logic chain: each resolution creates the conditions for the next crisis.*

12.7.5 The Origin of Paper III

To escape the delusion trap, the agent needs a mechanism to monitor the quality of its own reference frame—to "observe its own observation." This requires a *second-order control loop*: a meta-controller that adjusts the gauge fixing σ in response to accumulated prediction errors.

The key difficulty is that the prediction errors the agent can measure ($\mathcal{S}_{\rm vis} - \mathcal{S}_{\rm vis}^{\rm predicted}$) all lie within $V_{\rm fg}$. To detect frame drift, the agent must compare these in-frame errors to an estimate of out-of-frame contributions—a self-referential operation that requires *Fisher information about the agent's own parameters.*

This is the subject of Paper III: the Fisher information geometry of self-referential calibration, and the thermodynamic cost of the loop that closes the chain *Chaos $\to$ Time $\to$ Self $\to$ Calibration.*

12.8 Numerical Demonstration

The preceding sections establish analytic bounds and a worked example with a qubit in a two-channel bath. We now provide a numerical illustration showing that the core symmetry-breaking signature—attention entropy collapse under budget constraints—and the resulting selection advantage are reproduced in a minimal multi-dimensional system. Full code and parameters are provided for reproducibility.

12.8.1 Model

Environment. A D-dimensional linear prediction task with sparse rotating support: $y(t) = \mathbf{w}^*(t)^\top \mathbf{x}(t) + \xi(t)$, $\mathbf{x}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_D)$, $\xi \sim \mathcal{N}(0, \sigma^2)$. Only $m \ll D$ dimensions carry nonzero weight at any time; the active support rotates every $\tau_{\rm switch}$ steps, modelling environmental drift.

Hard budget constraint. Per step, the agent may update only k coordinates of its weight vector (a hard processing budget), mirroring the bounded computation assumption (B2).

Agents.

- **Budgeted selector (SSB)**: selects the top-k dimensions by importance score—an exponential moving average of the signed per-coordinate gradient. Signed accumulation ensures that noise dimensions (zero expected signal) cancel over time while signal dimensions persist, enabling reliable discrimination without access to the true support.
- **Random-k baseline**: selects k dimensions uniformly at random each step. This provides a budget-fair comparison: identical mechanism, no symmetry breaking.

The choice of *signed* gradient EMA (rather than squared-gradient magnitude) is structurally motivated: for noise dimensions $\mathbb{E}[r\,x_i] = 0$, so the signed accumulation cancels over time; for signal dimensions $\mathbb{E}[r\,x_i] \neq 0$, so a consistent directional bias persists. The signed EMA thus acts as a *directional coherence filter* that discriminates signal from noise without access to the true support—a minimal realisation of the "reference-frame bias" that emerges from symmetry breaking.

Parameters.

Quantity	Value	Role
D	64	ambient dimension
m	8	signal dimensions (sparse support)
T	10,000	horizon per trial
Seeds	10	independent replications
σ	0.3	observation noise std
η	0.02	SGD learning rate
λ	0.995	weight decay per step
k	2, 4, 6, 8, 10, 12, 16, 20, 24, 32, 48, 64	budget grid
τ_{switch}	$\{500, 1000, 2000\}$	support rotation period

Attention entropy. Let n_i be the number of updates coordinate i receives in a measurement window of the last 1,000 steps. The normalised update frequency $p_i = n_i / \sum_j n_j$ defines the attention entropy:

$$H_{\text{attn}} = -\sum_{i=1}^{D} p_i \ln p_i. \tag{12.42}$$

Under symmetric processing (no SSB), $p_i = 1/D$ and $H_{\text{attn}} = \ln D$. Under budget-constrained selection, H_{attn} collapses away from $\ln D$, serving as an order parameter for symmetry breaking.

Oracle metric. Neither agent has access to $\mathbf{w}^*(t)$. Performance is evaluated externally using the weight-space mean-squared error $\text{MSE} = \|\hat{\mathbf{w}} - \mathbf{w}^*\|^2$, averaged over post-burn-in steps.

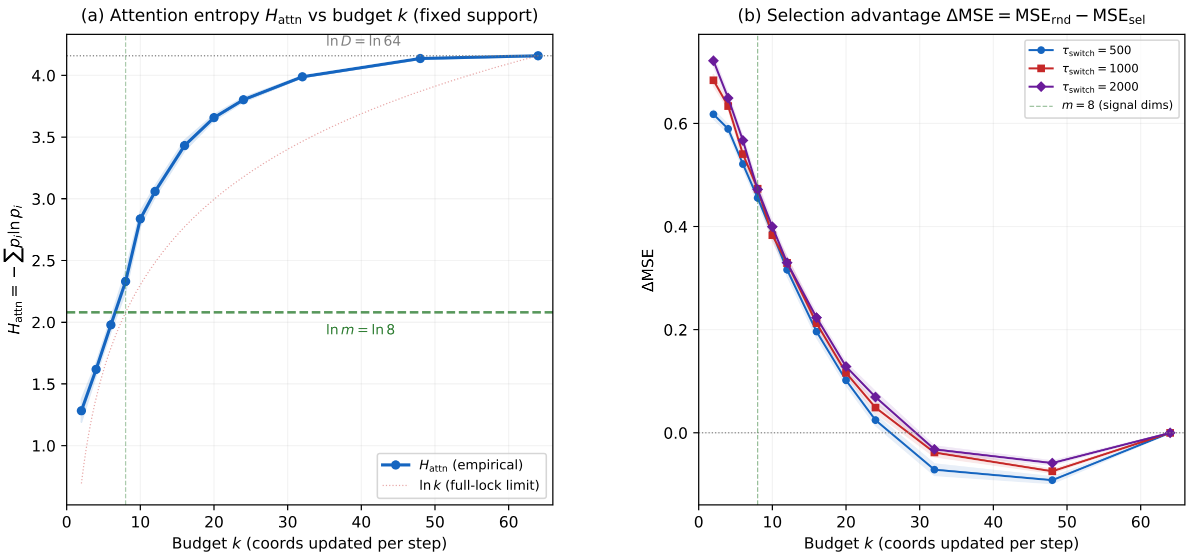

Figure 12.2: **Budget-induced symmetry breaking.** $D = 64$, $m = 8$, $T = 10{,}000$, 10 seeds, 95% CI bands. **(a)** Attention entropy H_{attn} vs budget k (fixed support). The empirical curve (blue) collapses from $\ln D \approx 4.16$ toward an $O(\ln m)$ floor as budget tightens. For $k \leq m$, H_{attn} remains below $\ln m \approx 2.08$ (green dashed), consistent with confinement to the signal subspace. **(b)** Selection advantage $\Delta\mathrm{MSE} = \mathrm{MSE}_{\mathrm{rnd}} - \mathrm{MSE}_{\mathrm{sel}}$ vs budget k under rotating support. The gap is positive for $k \lesssim 3m$ (selection helps), turns slightly negative at large k (commitment cost exceeds diversification), and returns to zero at $k = D$. Slower drift ($\tau = 2000$) yields a larger peak advantage.

12.8.2 Results

Figure 12.2 shows the two key signatures.

Result 1: Attention entropy collapse (Figure 12.2a). Under fixed support (no rotation), the attention entropy H_{attn} exhibits a sharp collapse away from $\ln D = \ln 64 \approx 4.16$ and increases monotonically with k, consistent with a budget-induced concentration of update mass onto signal-carrying dimensions. For budgets near and below the signal scale ($k \leq m$), H_{attn} remains $O(\ln m)$, consistent with confinement to the signal subspace. We use the collapse of H_{attn} away from $\ln D$ as the order parameter of symmetry breaking; a strict plateau at $\ln m$ is not expected under the present re-selection dynamics and finite-window estimator.

Result 2: Selection advantage (Figure 12.2b). Under rotating support, the mean-squared error gap $\Delta\mathrm{MSE} = \mathrm{MSE}_{\mathrm{rnd}} - \mathrm{MSE}_{\mathrm{sel}}$ is positive for $k \lesssim 3m$ and peaks at tight budgets ($k = 2$) where the selection advantage is strongest. For $k \gg m$ the gap turns slightly negative (the selector's commitment to stale dimensions costs more than the random baseline's diversification), before returning to zero at $k = D$. The three τ_{switch} curves are ordered: slower drift (larger τ) yields a larger peak gap, with the ordering most visible at small k.

12.8.3 Scope of This Demonstration

These simulations illustrate the symmetry-breaking phenomenon predicted by Theorem 12.18 under the stated model class; they do not constitute a proof beyond this class.

This demonstration **does** show:

1. Under hard budget constraints, attention entropy collapses sharply away from $\ln D$ and remains $O(\ln m)$ for $k \leq m$—the agent confines its updates to the signal subspace. This is the computational analogue of spontaneous symmetry breaking (Theorem 12.18).

2. A budgeted selector that exploits importance-weighted selection systematically outperforms a budget-fair random baseline, consistent with a survival advantage in the broken phase (cf. Proposition 12.24).

3. The advantage scales with both budget tightness (smaller k) and environmental stability (larger τ_{switch}).

In summary, this demonstration validates the *existence* and *measurability* of budget-induced symmetry breaking in a minimal linear setting; it does not claim universality across architectures or environment classes.

This demonstration does **not** show:

1. That H_{attn} reaches a strict plateau at $\ln m$ for all $k \leq m$. Under the adaptive re-selection dynamics used here, the selector cycles within the signal subspace, producing H_{attn} values near but not locked to $\ln m$. The relevant signature is the collapse *away from* $\ln D$, not convergence to a specific lower bound.

2. That the specific form of the importance score (signed gradient EMA) is optimal. It is one realisation of the selection mechanism.

3. That the results generalise to all environment classes. The model uses Gaussian features, linear regression, and sparse rotating support.

4. That the delusion–correction cycle is addressed. This is the subject of Paper III.

Reproducibility. The complete simulation is a self-contained Python script (`paper2_kstar_scaling_demo.py`, $\sim$540 lines, requiring only NumPy and Matplotlib) with fixed random seeds. All figures in this section can be reproduced by executing the script. The following files are included in the supplementary archive:

- `paper2_kstar_scaling_demo.py` — simulation script
- `fig_paper2_kstar_scaling.pdf` — Figure 12.2
- `kstar_scaling_data.csv` — raw performance gap data

12.9 Discussion

12.9.1 Summary of Results

Result	**Statement**	**Sec.**
Computational Ceiling	Symmetric processing cost exceeds $\mathcal{C}_{\text{budget}}$ at τ_{par}	12.3
Rate-Distortion Bound	Optimal compression retains $k^* = \mathcal{C}_{\text{budget}}/h_\mu$ components	12.4.2
Necessity of SSB	Under bounded computation, survival requires gauge fixing	12.4.3
Four Bias Terms	Broken phase acquires $\mathcal{B}_{\text{select}}$, $\mathcal{B}_{\text{frame}}$, $\mathcal{B}_{\text{center}}$, $\mathcal{B}_{\text{inc}}$	12.4.4
Survival Decomposition	$\mathcal{S} = \mathcal{S}_{\text{vis}} + \mathcal{S}_{\text{hid}}$	12.5.2
Ego-Entropy Trade-off	$\gtrsim 1 - k^*/\dim Cl(V,q)$ of I_{pred} discarded (uniformity assumption)	12.5.4
Delusion Trap	Fixed frame diverges from optimal under environmental drift; agent cannot self-detect	12.7.4
Numerical demo	Budget-induced SSB and selection advantage (Fig. 12.2)	12.8

12.9.2 Testable Predictions

The results of this paper yield several quantitative predictions that are, in principle, falsifiable in computational and biological systems. We emphasize that these predictions follow from the *classical* content of T-DOME: the thermodynamic and information-theoretic bounds apply to any open system maintaining a non-equilibrium steady state under bounded computation, regardless of whether the underlying dynamics are quantum or classical.

Prediction 1: Attention entropy collapse under resource constraints. Theorem 12.18 predicts that any adaptive agent operating under computational budget $\mathcal{C}_{\text{budget}}$ in an environment with D relevant degrees of freedom and metric entropy h_μ will exhibit a sharp symmetry-breaking transition when $\mathcal{C}_{\text{budget}} < D \cdot h_\mu \cdot \tau_{\text{mem}}$. The order parameter—attention entropy $H_{\text{attn}} = -\sum_i p_i \ln p_i$ over the agent's allocation of computational resources—should collapse from $\ln D$ (symmetric phase) to $\ln k^*$ (broken phase), with $k^* = \lfloor \mathcal{C}_{\text{budget}}/h_\mu \rfloor$.

Protocol: Train reinforcement learning agents on a D-dimensional partially observable environment with systematically varied computational budgets (context window, hidden state dimension, or parameter count). Measure attention entropy as a function of $\mathcal{C}_{\text{budget}}/D$. The predicted transition should be:

- Sharp (not gradual) in the limit $D \to \infty$
- Located at $\mathcal{C}_{\text{budget}}/(D\, h_\mu\, \tau_{\text{mem}}) \approx 1$
- Accompanied by spontaneous selection of k^* "preferred" features

The numerical demonstration in Section 12.8 provides preliminary evidence for this transition. Extending to deep RL agents with continuous state spaces would constitute a non-trivial test.

Prediction 2: Information-discarded scaling law. Theorem 12.25 predicts that the mutual information between an agent's compressed representation and the full environmental state satisfies:

$$\frac{I_{\text{discarded}}}{I_{\text{pred}}} \gtrsim 1 - \frac{k^*}{D} \tag{12.43}$$

under the uniformity assumption. This is a specific, quantitative prediction: an agent retaining k^* out of D features discards at least a fraction $1 - k^*/D$ of the total predictive information.

Protocol: For RL agents with tunable bottleneck width k, measure the mutual information $I(\hat{X}; X)$ between the compressed internal state $\hat{X}$ and the full environmental state X using variational estimators [77]. Plot $I_{\text{discarded}}/I_{\text{pred}}$ against k/D. The prediction is a linear lower bound with slope $-1/D$ and intercept 1.

Prediction 3: Delusion trap timescale. Theorem 12.32 predicts a critical timescale $t_{\text{del}} = \Lambda^{-1} \ln(\pi/4\theta_0)$ beyond which a fixed-frame agent's gradient updates become anti-correlated with the true optimum.

Protocol: Deploy a fixed-architecture RL agent in an environment with controlled rotation rate Λ of the reward-relevant features. Measure the cosine similarity between the agent's policy gradient $\nabla_u \mathcal{S}_{\text{vis}}$ and the oracle gradient $\nabla_u \mathcal{S}_{\text{full}}$ as a function of time. The prediction:

- Cosine similarity > 0 for $t < t_{\text{del}}$ (updates improve performance)
- Cosine similarity crosses zero at $t \approx t_{\text{del}}$
- Cosine similarity < 0 for $t > t_{\text{del}}$ (updates degrade performance)
- The crossover timescale scales as $\Lambda^{-1} \ln(1/\theta_0)$

This is directly testable in non-stationary RL benchmarks with controlled distributional shift.

Prediction 4: Biological information bottleneck. The ego-entropy tradeoff is not specific to artificial agents. Any biological system maintaining homeostasis under bounded sensory processing capacity should exhibit the same k^*/D scaling. In bacterial chemotaxis, the bacterium processes $D \sim 5$–10 chemical gradients through a signaling pathway with $k^* \sim 2$–3 effective internal states (CheY-P phosphorylation levels). The prediction is:

$$\frac{I(\text{internal state; chemical environment})}{I(\text{chemical environment; survival})} \lesssim \frac{k^*}{D} \approx 0.2\text{–}0.6 \tag{12.44}$$

Single-cell measurements of chemotactic mutual information [61, 40] provide the experimental technology to test this bound. The key test is whether the information efficiency saturates the bound (suggesting the biological system has optimized against it) or falls significantly below it.

12.9.3 What This Paper Does and Does Not Show

This paper **does** show:

1. Under bounded computation (B2) and non-trivial environment (B3), symmetric processing of memory leads to computational paralysis (Theorem 12.7).
2. The survival-optimal response is spontaneous symmetry breaking of the internal reference frame (Theorem 12.18), governed by a rate-distortion bound (Theorem 12.17).
3. The broken phase acquires four generic bias terms under (B1)–(B5) (Proposition 12.20).
4. Under environmental drift, a fixed frame leads to exponential divergence of prediction error—the Delusion Trap (Theorem 12.32).
5. A minimal computational demonstration reproduces the budget-induced symmetry-breaking signature (attention entropy collapse away from $\ln D$) and selection advantage over a budget-fair random baseline (Section 12.8, Figure 12.2).

This paper does **not** show:

1. That the privileged basis is uniquely determined by computational constraints. The basis is constrained but not unique—different histories lead to different gauge fixings, as in a ferromagnet.
2. That symmetry breaking is sufficient for persistence. It is the survival-optimal strategy under bounded computation; sufficiency requires the self-referential calibration of Paper III.
3. That the "ego" implies or requires consciousness, subjective experience, or phenomenal awareness. The term is used strictly in the control-theoretic sense.
4. That this framework constitutes a theory of consciousness. It is a theory of computational optimality under thermodynamic constraints.
5. That the four bias terms exhaust the phenomenology of self-reference. They are the minimal structural consequences of gauge fixing in a Clifford algebra.
6. That the rate-distortion bound is achievable by any specific physical implementation. It is an information-theoretic lower bound.
7. That the Delusion Trap is inescapable. Paper III will show it can be mitigated by self-referential calibration.
8. That the framework constitutes or implies a philosophical or metaphysical claim about the nature of selfhood.
9. That this framework applies to all possible physical systems. It applies to systems satisfying (B1)–(B5)—persistent agents with finite computation in non-trivial environments.
10. That the Clifford algebra is the only possible algebraic setting. It is the minimal setting inherited from Q-RAIF. Other algebras may yield analogous results.

Chapter 13

Fisher Information Geometry and the Thermodynamic Cost of Self-Referential Calibration

Paper III — "The Loop"

Originally published: Zenodo, DOI: 10.5281/zenodo.18591771

Abstract

Papers I and II of the T-DOME series [143, 144] established that persistent agents must carry non-Markovian memory (Paper I) and must spontaneously break the gauge symmetry of their internal Clifford algebra $Cl(V, q)$ to form a compressed reference frame—the "ego" $\mathfrak{E} = (\mathcal{F}^*, V_{\mathrm{fg}}^*)$ (Paper II). Paper II concluded with the **Delusion Trap**: under environmental drift, a fixed reference frame decouples from the optimal gauge on the logarithmic timescale $t_{\mathrm{del}} = \Lambda^{-1} \ln(\pi/4\theta_0)$, and the agent cannot detect this failure from within its own foreground subspace V_{fg}.

In this final work we derive the theory of **self-referential calibration**. We show that while the agent cannot observe the background subspace V_{bg} directly, it can measure the **Fisher information** of its own prediction-residual stream with respect to its frame parameters σ. We prove three main results:

1. **Drift Detectability** (Theorem 13.14): environmental drift generates a quadratically growing signal in the self-referential Fisher information $\mathcal{I}_F(\sigma)$, detectable before the Delusion Trap closes.

2. **Self-Referential Cramér–Rao Bound** (Theorem 13.18): the agent's drift-estimation error is bounded below by $1/(n_{\mathrm{eff}}\, \mathcal{I}_F + \mathcal{I}_{\mathrm{ego}})$, where $\mathcal{I}_{\mathrm{ego}}$ quantifies the rigidity of the ego prior.

3. **Thermodynamic Cost of the Loop** (Theorem 13.29): the minimum dissipation rate for self-referential calibration is $\dot{W}_{\mathrm{loop}} \geq k_B T \ln 2 \, [h_\mu \, k^* + \mathcal{C}_{\mathrm{meta}}] + \mathcal{L}^2/\tau_{\mathrm{recalib}}^2$, where $\mathcal{L}$ is the thermodynamic length of the frame update and τ_{recalib} is the recalibration time.

The calibration loop satisfies a Lyapunov tracking bound (Theorem 13.24), keeping the mismatch within a neighbourhood whose size is set by the ratio of environmental drift speed to adaptation rate. We identify this loop as the minimal physical realisation of *reflexivity*—estimating drift from residual statistics and correcting the frame via Lyapunov-monitored natural gradient descent. Combining with Papers I and II, we state a **Four-Part Structure Proposition** (Proposition 13.28): within the class of agents satisfying (C1)–(C5), a sufficient architecture for persistence under drift requires (1) an external observable geometry, (2) an internal control algebra, (3) a self-monitoring Lyapunov function, and (4) biased non-Markovian memory.

13.1 Introduction

13.1.1 Context: The Delusion Trap

Paper II of this series [144] established that persistent agents under bounded computation must spontaneously break the gauge symmetry of their internal algebra $Cl(V, q)$, selecting a privileged reference frame $\mathcal{F}^*$ that compresses the memory kernel into a tractable k^*-dimensional foreground subspace V_{fg}. This gauge fixing—the "ego" $\mathfrak{E} := (\mathcal{F}^*,\, V_{\mathrm{fg}}^*)$—is not an additional hypothesis but the survival-optimal strategy under bounded rationality.

However, Paper II's final theorem revealed a fatal consequence. Under environmental drift (spectral-density parameters changing at rate ε), the mismatch angle between the agent's fixed frame and the instantaneous optimal frame grows as $\theta(t) = \theta_0\, e^{\Lambda t}$ (Paper II, Definition 27), where $\Lambda \sim \varepsilon/\tau_{\mathrm{adapt}}$ is the environmental Lyapunov exponent. Beyond the *delusion time*

$$t_{\mathrm{del}} = \frac{1}{\Lambda} \ln\left(\frac{\pi/4}{\theta_0}\right), \tag{13.1}$$

three catastrophic failures occur simultaneously (Paper II, Theorem 29):

1. The hidden survival component dominates: $|\mathcal{S}_{\mathrm{hid}}| > |\mathcal{S}_{\mathrm{vis}}|$.
2. The agent's update direction anti-correlates with the true survival gradient:

$$\langle \nabla_u \mathcal{S}_{\mathrm{vis}},\, \nabla_u \mathcal{S}_{\mathrm{full}} \rangle < 0.$$

3. All four bias terms ($\mathcal{B}_{\mathrm{select}}, \mathcal{B}_{\mathrm{frame}}, \mathcal{B}_{\mathrm{center}}, \mathcal{B}_{\mathrm{inc}}$) operate within V_{fg} and cannot register changes in the background V_{bg}.

Paper II further showed (Remark 31) that "dithering"—randomly probing the background subspace—fails because the agent has no gradient signal to indicate *when* or *where* to probe. The exponential divergence in V_{bg} is invisible until it dominates, at which point it is too late.

The present paper provides the escape.

13.1.2 Position within Papers I–III

This paper is the third and final of the T-DOME framework, closing the three-paper sequence.

Framework	Question	Result	Status
HAFF [127, 128]	How does geometry emerge?	Algebra → Geometry	Complete
Q-RAIF [140, 141]	What algebra must an observer have?	$Cl(V, q) \hookrightarrow Cl(1, 3)$	Complete
T-DOME I [143]	Why must agents carry memory?	Markovian ceiling; memory as necessity	Complete
T-DOME II [144]	Why must agents break symmetry?	Reference-frame selection under bounded computation	Complete
T-DOME III (this work)	How does self-calibration arise?	Fisher self-referential bound; thermodynamic cost of reflexivity	**This paper**

The three T-DOME papers form an irreversible logical chain:

1. **Paper I:** Without memory, a system is trapped in the Markovian present. Memory breaks this trap but floods the system with unbounded historical data.
2. **Paper II:** Unbounded memory under finite computational resources causes processing collapse. Spontaneous symmetry breaking resolves the overload but introduces systematic bias.
3. **Paper III (this work):** Uncorrected bias diverges from a changing environment. A self-referential calibration loop—monitoring the Fisher information of one's own prediction stream—resolves the bias but requires a second-order control structure and an irreducible thermodynamic cost.

Each resolution creates the precondition for the next crisis: memory enables overload, compression enables bias, and bias demands calibration. Only the complete closure *Paper I + Paper II + Paper III* allows a system to persist under the Second Law in a drifting environment.

13.1.3 The Information-Geometric Insight

The key observation that resolves the Delusion Trap is subtle: *while the agent cannot observe V_{bg} directly, it can observe the statistical properties of its own prediction residuals in V_{fg}.*

The prediction residual $e(t) := \mathcal{S}_{\mathrm{vis}}(t) - \mathcal{S}_{\mathrm{vis}}^{(\mathrm{pred})}(t)$ lies in V_{fg} by construction. Its *value* carries no information about the background. But its *distribution*—the probability law $p(e \mid \sigma)$, parametrised by the gauge-fixing parameter σ—does depend on σ, because the projection $\Pi_{\mathcal{F}}(\sigma)$ determines which environmental correlations are captured and which are discarded.

When the frame σ drifts away from the optimal σ^*, the residual distribution shifts. The *Fisher information metric*

$$g_{ij}(\sigma) = \mathbb{E}_\sigma\left[\frac{\partial \log p(e \mid \sigma)}{\partial \sigma^i}\,\frac{\partial \log p(e \mid \sigma)}{\partial \sigma^j}\right] \tag{13.2}$$

measures the sensitivity of this distribution to changes in σ. A spike in g_{ij}—a "stress" in the agent's internal geometry—is the signal that the reference frame is becoming stale.

This is the mathematical realisation of the "second-order operation" demanded by Paper II, Section 7.5: the agent does not need to see the truth (the full $Cl(V, q)$), but only the *rate of change of its own prediction error* as a function of its frame parameters. Fisher information is precisely this quantity.

13.1.4 Relation to Architectural Incompleteness

The architectural incompleteness result [131] established *architectural incompleteness*: the observable-algebra framework cannot self-ground. Paper II provided a partial operational response (the ego as gauge fixing under bounded computation). The present paper provides the final operational response: the self-referential calibration loop cannot *eliminate* architectural incompleteness, but it can *track* the consequences of incompleteness in real time. The Lyapunov function $V(\sigma)$ monitors the distance between the agent's frame and the optimal frame without requiring access to the "complete" description—it operates entirely within the agent's own predictive statistics.

13.1.5 Scope and Disclaimers

1. *Reflexivity* refers throughout to second-order control: the ability of a system to monitor and adjust its own monitoring process. It carries *no* implication of phenomenal consciousness, subjective experience, or qualia.

2. The self-referential calibration loop does not *eliminate* the ego's bias; it tracks and compensates for drift in the bias. The four bias terms of Paper II persist in the calibrated phase.

3. The thermodynamic cost bounds are information-theoretic lower bounds, not claims about specific physical implementations.

4. The framework applies to systems satisfying (C1)–(C5) (Section 13.2.6). It is not a universal theory of agency.

Related work. The Fisher information metric on statistical manifolds was introduced by Rao [79] and shown to be unique by Čencov [21]. The natural gradient and information geometry were developed by Amari [2, 3]. Thermodynamic length and optimal finite-time transformations were established by Crooks [26] and Sivak–Crooks [95]. The connection between Fisher information and entropy production was formalised by Ito [51] and Barato–Seifert [10]. Second-order cybernetics originates with Ashby [8] and von Foerster [125]. Adaptive control and self-tuning regulators are treated in [9]. The Bayesian Cramér–Rao bound (van Trees inequality) is from [124].

Summary of contributions. This paper establishes three main results:

1. **Drift Detectability** (Theorem 13.14): the self-referential Fisher information of the prediction-residual stream grows quadratically with accumulated drift, providing a detectable signal before the Delusion Trap closes.

2. **Self-Referential Cramér–Rao Bound** (Theorem 13.18): drift-estimation precision is bounded by the sum of data Fisher information and ego rigidity.

3. **Thermodynamic Cost** (Theorem 13.29): the self-calibration loop requires a minimum dissipation rate with three distinct components (sensing, computing, actuating).

13.2 Mathematical Preliminaries

13.2.1 Inherited Framework from Papers I and II

We briefly recall the key objects; the reader is referred to Papers I and II for full definitions and proofs.

From Paper I [143].

- **Survival functional.** $\mathcal{S}[\Lambda, \tau] := \Delta F - W[0, \tau]$ (Paper I, Eq. (9)).
- **Markovian Ceiling.** $\mathcal{S}[\Lambda^{\mathrm{M}}, \tau] \leq 0$ for all $\tau \geq 0$.
- **Memory kernel.** $\mathcal{K}(t, s)$: the non-Markovian superoperator encoding system–environment correlations.
- **Entropy rate.** $h_\mu := \lim_{T\to\infty} T^{-1} H(X_{0:T})$ (bits per unit time per algebraic component).
- **Predictive information.** $I_{\text{pred}} := I(\overleftarrow{X}; \overrightarrow{X})$.

From Paper II [144].

- **Internal algebra.** $\mathcal{O}_{\text{int}} = Cl(V, q)$, $D = \dim Cl(V, q) = 2^n$, gauge group $G = \mathrm{Aut}(Cl(V, q))$.
- **Gauge bundle.** $\pi : P \to M$, structure group G; a section $\sigma : M \to P$ is a reference frame.
- **Ego.** $\mathfrak{E} := (\mathcal{F}^*,\, V^*_{\text{fg}})$ with $k^* = \lfloor \mathcal{C}_{\text{budget}}/h_\mu \rfloor$ foreground components.
- **Projected kernel.** $\mathcal{K}_{\mathcal{F}}(t, s) = \Pi_{\mathcal{F}}\, \mathcal{K}(t, s)\, \Pi_{\mathcal{F}}$.
- **Survival decomposition.** $\mathcal{S} = \mathcal{S}_{\text{vis}}(\mathcal{F}) + \mathcal{S}_{\text{hid}}(\mathcal{F})$.
- **Four bias terms.** $\mathcal{B}_{\text{select}}$, $\mathcal{B}_{\text{frame}}$, $\mathcal{B}_{\text{center}}$, $\mathcal{B}_{\text{inc}}$ (Paper II, Proposition 18, Table 2).
- **Delusion Trap.** $t_{\text{del}} = \Lambda^{-1} \ln(\pi/4\theta_0)$ (Paper II, Theorem 29).
- **Information-objects convention.** $I(\mathcal{K}_{\mathcal{F}}; \mathcal{K}) \equiv I(\hat{X}; X)$ on induced record processes (Paper II, Remark 15).

13.2.2 Fisher Information Metric

Definition 13.1 (Fisher information matrix)**.** *Let* $\{p(x \mid \theta) : \theta \in \Theta \subset \mathbb{R}^d\}$ *be a parametric family of probability densities satisfying standard regularity conditions (interchange of differentiation and integration). The* Fisher information matrix *is*

$$g_{ij}(\theta) := \mathbb{E}_\theta\left[\frac{\partial \log p(x \mid \theta)}{\partial \theta^i}\,\frac{\partial \log p(x \mid \theta)}{\partial \theta^j}\right] = -\mathbb{E}_\theta\left[\frac{\partial^2 \log p(x \mid \theta)}{\partial \theta^i\,\partial \theta^j}\right]. \tag{13.3}$$

The pair (Θ, g) *is a Riemannian manifold called the* statistical manifold.

Remark 13.2 (Uniqueness)**.** *By Čencov's theorem [21], the Fisher–Rao metric* g^{FR} *is, up to a positive scalar multiple, the unique Riemannian metric on the space of probability distributions that is invariant under all Markov morphisms (sufficient-statistic embeddings). This uniqueness guarantees that the Fisher metric is the* canonical *choice for measuring drift on the statistical manifold of the agent's predictive model—it is not a design choice but a mathematical necessity.*

Proposition 13.3 (Cramér–Rao bound)**.** *For any unbiased estimator* $\hat{\theta}$ *of* θ *based on* n *independent observations:*

$$\mathrm{Cov}(\hat{\theta}) \;\succeq\; \frac{1}{n}\,\big[g(\theta)\big]^{-1} \tag{13.4}$$

in the Löwner order. The scalar case reads $\mathrm{Var}(\hat{\theta}) \geq 1/\big(n\, g(\theta)\big)$.

Remark 13.4 (Effective independence)**.** *Throughout this paper, references to "independent observations" in the context of continuous-time residual streams should be read as* effective independence *after thinning by the environmental decorrelation time* τ_E*, yielding an effective sample size* $n_{\mathrm{eff}} \approx T/\tau_E$*. In particular, the sample count* n *in* (13.4) *becomes* n_{eff} *in the self-referential setting of Section 13.4.2.*

Remark 13.5 (Fisher metric and KL divergence)**.** *The Fisher metric arises as the Hessian of the Kullback–Leibler divergence [25]:*

$$D_{\mathrm{KL}}\big(p_\theta \,\|\, p_{\theta+d\theta}\big) = \tfrac{1}{2}\, g_{ij}(\theta)\, d\theta^i\, d\theta^j + O\big(|d\theta|^3\big)\,. \tag{13.5}$$

This identifies the Fisher metric as the infinitesimal measure of statistical distinguishability.

13.2.3 Information Geometry

Following Amari [1, 3], the statistical manifold (Θ, g) carries additional geometric structure beyond the Riemannian metric.

α-connections. For each $\alpha \in [-1, 1]$, Amari defines an affine connection $\nabla^{(\alpha)}$ on Θ. The cases $\alpha = 1$ (exponential connection, $\nabla^{(e)}$) and $\alpha = -1$ (mixture connection, $\nabla^{(m)}$) are *dual* with respect to g: $\partial_k\, g(X, Y) = g(\nabla^{(e)}_k X,\, Y) + g(X,\, \nabla^{(m)}_k Y)$. For exponential families, $\nabla^{(e)}$ is flat in natural parameters and $\nabla^{(m)}$ is flat in expectation parameters—the *dually flat structure*. The case $\alpha = 0$ recovers the Levi-Civita connection of the Fisher metric.

Natural gradient. Standard gradient descent in parameter space ignores the curvature of the statistical manifold. The *natural gradient* [2]

$$\dot{\theta} = -\eta\, g^{-1}(\theta)\, \nabla_\theta L(\theta), \tag{13.6}$$

where $\eta > 0$ is the learning rate and $L(\theta)$ is a loss function, provides the steepest descent direction in the Fisher metric. It is reparametrisation-invariant and Fisher-efficient (achieves the Cramér–Rao bound asymptotically).

Pythagorean theorem. In a dually flat space, the KL divergence satisfies a generalised Pythagorean relation: $D_{\mathrm{KL}}(p \,\|\, r) = D_{\mathrm{KL}}(p \,\|\, q) + D_{\mathrm{KL}}(q \,\|\, r)$ when q is the m-projection of p onto a submanifold containing r. This decomposition will be applied to separate the foreground-recoverable and background-irrecoverable components of drift.

13.2.4 Thermodynamic Length

Definition 13.6 (Thermodynamic length). *Let $\lambda(t)$ for $t \in [0, \tau]$ be a path through control parameter space, and let $\zeta_{ij}(\lambda)$ be the* friction tensor *(the time-integrated equilibrium force–force correlation function at λ). The* thermodynamic length *of the path [26] is*

$$\mathcal{L} := \int_0^\tau \sqrt{\zeta_{ij}(\lambda)\, \dot{\lambda}^i\, \dot{\lambda}^j}\; dt. \tag{13.7}$$

Proposition 13.7 (Sivak–Crooks bound). *The excess (dissipated) work during a finite-time transformation of duration τ satisfies [95]*

$$W_{\mathrm{ex}} \;\geq\; \frac{\mathcal{L}^2}{\tau}. \tag{13.8}$$

The minimum is achieved by the geodesic of the friction tensor ζ. In the linear-response regime, the friction tensor is related to the Fisher metric of the equilibrium distribution at λ by $\zeta_{ij}(\lambda) \sim \tau_{\mathrm{relax}}\, g_{ij}^{\mathrm{Fisher}}(\lambda)$, where τ_{relax} is the relaxation time.

13.2.5 Second-Order Cybernetics

Von Foerster [125] distinguished two levels of control:

- **First-order cybernetics**: feedback control of observed systems. The controller adjusts its actions based on the output of a sensor. Paper II's ego is a first-order structure: it processes environmental data within a fixed frame.
- **Second-order cybernetics**: feedback control of the *observing* system itself. The controller adjusts the *sensor*—or equivalently, the reference frame within which the sensor operates. This is what Paper III provides.

Ashby's Law of Requisite Variety [8] provides a lower bound on the complexity of the meta-controller:

$$\dim(\text{meta-controller state space}) \;\geq\; \dim(\text{environmental drift subspace})\,. \tag{13.9}$$

The meta-observer must have at least as many adjustable parameters as there are independent modes of environmental drift.

In adaptive control theory [9], the analogous result is the *persistent excitation* condition: parameter estimates converge if and only if the input signal is "rich enough" to excite all modes of the system. In our framework, persistent excitation corresponds to $h_\mu > 0$—the environment must continue to generate novelty for the self-calibration loop to function.

Remark 13.8 (Operational content). *The second-order cybernetic structure in this paper is* not *a philosophical metaphor. It has concrete operational content: the natural gradient update* (13.6) *is a specific algorithm that takes as input the Fisher information of the residual stream and produces as output an update to the frame parameter σ. This algorithm can be implemented by any physical system capable of accumulating second-moment statistics of its own prediction errors over a window of length $T \geq \tau_E$.*

13.2.6 Standing Assumptions

Definition 13.9 (Standing Assumptions). *Throughout this paper, the following conditions are assumed:*

(C1) ***Inherited framework.*** *All assumptions (B1)–(B5) of Paper II [144] remain in force. This transitively includes (A1)–(A5) of Paper I [143] (open quantum system, thermal bath, well-defined free energy, finite Hilbert space, weak coupling) and the realizability embedding $\phi : Cl(V, q) \hookrightarrow Cl(1, 3)$ (Q-RAIF [142]). We invoke this embedding strictly as a structural inheritance from the earlier papers; no new physical claims about $Cl(1, 3)$ spacetime are introduced here. Additionally, the Delusion Trap is active: $\tau_{\text{mem}} > \tau_{\text{par}}$ and $\Lambda > 0$.*

(C2) ***Environmental drift.*** *The instantaneous optimal frame $\mathcal{F}^*(t)$ rotates continuously in G at a rate characterised by the Lyapunov exponent $\Lambda > 0$ (Paper II, Eq. (37)).*

(C3) ***Finite meta-observer budget.*** *The self-calibration loop has a computational budget $\mathcal{C}_{\text{meta}} < \infty$ (bits per unit time), distinct from the ego's processing budget $\mathcal{C}_{\text{budget}}$.*

(C4) ***Regularity.*** *The agent's predictive family $\{p(e \,|\, \sigma) : \sigma \in G/H\}$ satisfies standard Fisher information regularity: full rank, finite Fisher matrix, and interchange of differentiation and integration. This extends (A5) from Paper I.*

(C5) ***Persistent excitation.*** *The environmental entropy rate satisfies $h_\mu > 0$ for all t. The environment generates new information indefinitely; no "frozen" regimes occur.*

13.3 The Drift Detection Problem

13.3.1 Why First-Order Control Fails

Theorem 13.10 (First-Order Insufficiency). *Under assumptions* (C1)–(C5)*, decompose the prediction residual as $e(t) = e_{\text{drift}}(t) + \xi(t)$, where e_{drift} is the deterministic drift-induced component (second-order in θ) and $\xi(t)$ is the innovation noise, whose distribution is symmetric on V_{fg} under* (C5). *No first-order controller—one that updates $\dot{\sigma} = f(e(t))$ based on the instantaneous residual without computing statistical properties of the error stream—can uniformly reduce the drift. Specifically: for any deterministic update* *function f, there exists a measurable event $\mathcal{E} \subset V_{\mathrm{fg}}$ with $\mathbb{P}(\mathcal{E}) \geq 1/2 - O(\mathrm{SNR})$ under the symmetric innovation distribution $p(\xi)$, such that for all $\xi \in \mathcal{E}$ the update direction satisfies $\langle \dot{\sigma}, \dot{\sigma}^* \rangle \leq 0$, where $\mathrm{SNR} \sim \mathcal{S}_{\mathrm{tot}}^2 \theta^4 / h_\mu$.*

Proof. Probability space. The probability is taken over the innovation sequence $\{\xi(t)\}_{t\geq 0}$ under the symmetric distribution induced by the bath coupling (C5). All expectations below are over $p(\xi)$.

Signal-to-noise separation. The prediction error $e(t)$ lies in V_{fg} by construction. Frame drift manifests as a rotation of the optimal frame $\mathcal{F}^*(t)$ in the gauge group G, shifting survival weight from V_{fg} to V_{bg}. In V_{fg}, the drift signal enters only at second order in the mismatch angle θ (Paper II, proof of Theorem 29, part (c)): $\mathcal{S}_{\mathrm{vis}} = \mathcal{S}_{\mathrm{tot}} \cos^2\theta$, so $e_{\mathrm{drift}} \sim \theta^2 \mathcal{S}_{\mathrm{tot}}$. The noise $\xi(t)$ scales as $h_\mu^{1/2}$. For $\theta \ll 1$, the single-sample signal-to-noise ratio is $\mathrm{SNR} \sim \mathcal{S}_{\mathrm{tot}}^2 \theta^4 / h_\mu \ll 1$.

Symmetry argument. Since $p(\xi)$ is symmetric on V_{fg}, for any deterministic f:

- If f is odd (e.g., linear gain), $\mathbb{E}[f(e_{\mathrm{drift}} + \xi)] \approx f(e_{\mathrm{drift}})$, but the instantaneous sign of f is determined by ξ with probability $\frac{1}{2} - O(\mathrm{SNR})$.
- If f is even, $f(e)$ carries no information about the *sign* of $\dot{\sigma}^*$, so $\langle f(e), \dot{\sigma}^* \rangle$ vanishes in expectation.

In either case, the probability that the update direction anti-correlates with the true drift direction is at least $1/2 - O(\mathrm{SNR})$. Systematic drift detection requires accumulating second-order statistics of the residual stream over multiple samples—a second-order operation. The symmetry argument holds exactly for Gaussian innovations and approximately for symmetric sub-Gaussian innovations. □

13.3.2 The Agent's Statistical Manifold

The agent's prediction-residual stream $\{e(t)\}_{t\geq 0}$ defines a stochastic process whose distribution depends on the gauge-fixing parameter σ. We model this dependence as a parametric family.

Definition 13.11 (Predictive family). *The* predictive family *of the agent is the set*

$$\mathcal{P} := \{p(e \,|\, \sigma) : \sigma \in \mathcal{M}_G\}, \tag{13.10}$$

where $\mathcal{M}_G := G/H$ is the space of gauge-fixing orbits (H is the stabiliser of the foreground subspace), e denotes the prediction-residual time series over a window of length T, and $p(e \,|\, \sigma)$ is the likelihood of the observed residuals given the gauge parameter σ.

The key insight is that $p(e \,|\, \sigma)$ depends on σ even though $e(t) \in V_{\mathrm{fg}}$, because the projection $\Pi_{\mathcal{F}}(\sigma)$ determines which environmental correlations are captured. When σ drifts from the optimal σ^*:

- The variance of the residuals increases (the discarded background components contribute unmodelled noise).
- The temporal correlations of the residuals change (the projected kernel $\mathcal{K}_{\mathcal{F}}$ no longer captures the dominant environmental modes).
- Higher-order statistics (kurtosis, spectral shape) shift systematically.

These distributional changes are invisible to the raw error $e(t)$ but detectable by the Fisher metric of $\mathcal{P}$.

13.3.3 Self-Referential Fisher Information

Definition 13.12 (Self-referential Fisher information)**.** *The* self-referential Fisher information *of the agent at gauge parameter σ is*

$$\mathcal{I}_F(\sigma) := g_{ij}(\sigma)\,\delta\sigma^i\,\delta\sigma^j, \tag{13.11}$$

where $g_{ij}(\sigma)$ is the Fisher information matrix of the predictive family $\mathcal{P}$ (Definition 13.11) evaluated at σ, and $\delta\sigma$ is the frame perturbation direction. In the scalar case (single drift mode), $\mathcal{I}_F(\sigma) = \mathbb{E}_\sigma\big[(\partial_\sigma \log p(e \,|\, \sigma))^2\big]$.

Remark 13.13 (What the agent "measures")**.** *Computing $\mathcal{I}_F(\sigma)$ does not require access to V_{bg} or to the "true" environment. It requires only: (i) the agent's own prediction residuals $\{e(t)\}$ (which lie in V_{fg}), and (ii) the ability to evaluate the score function $\partial_\sigma \log p(e \,|\, \sigma)$— the sensitivity of its own predictive model to frame perturbations. This is a computation entirely within the agent's internal algebra, using only quantities already available from the ego's processing pipeline.*

Theorem 13.14 (Drift Detectability)**.** *Under assumptions* (C1)–(C5)*, suppose the frame is freshly calibrated at time t_0 ($\theta(t_0) = 0$). Then the self-referential Fisher information of the prediction-residual stream satisfies, for small accumulated drift ($\Lambda\,\Delta t \ll 1$):*

$$\mathcal{I}_F(\sigma;\,\{e_t\}_{t_0}^{t_0+\Delta t}) \;\geq\; \kappa\,\Lambda^2\,(\Delta t)^2\,\mathcal{I}_F^{\mathrm{env}}, \tag{13.12}$$

where:

- $\kappa := \inf_{\sigma\in\mathcal{N}}(\partial\theta/\partial\sigma)^2 > 0$ *is the* coupling efficiency*, where $\mathcal{N}$ is a compact neighbourhood of the calibrated point σ^* on which the gauge chart is non-singular (existence guaranteed by* (C4)*; see proof);*
- Λ *is the environmental Lyapunov exponent* (Paper II, Eq. (37))*;*
- Δt *is the observation window;*
- $\mathcal{I}_F^{\mathrm{env}} := \mathbb{E}_{p(\cdot|\theta)}[(\partial_\theta \log p)^2]$ *is the per-component environmental Fisher information, measuring the baseline sensitivity of the decoherence functions to the mismatch angle θ.*

The self-referential Fisher information grows quadratically *with accumulated drift time.*

Proof. Step 1: chain rule. The frame parameter σ determines the mismatch angle $\theta = \theta(\sigma)$ via the gauge map $G/H \to [0, \pi/2]$. The chain rule for Fisher information gives

$$\mathcal{I}_F(\sigma) \;=\; \left(\frac{\partial\theta}{\partial\sigma}\right)^2 \mathcal{I}_F(\theta), \tag{13.13}$$

where $\mathcal{I}_F(\theta) := \mathbb{E}[(\partial_\theta \log p(e|\theta))^2]$ is the Fisher information of the residual stream with respect to the mismatch angle. By (C4) (full-rank Fisher matrix), the Jacobian $\partial\theta/\partial\sigma$ is bounded away from zero on any compact neighbourhood $\mathcal{N}$ of the calibrated point; we define the coupling efficiency $\kappa := \inf_{\sigma\in\mathcal{N}}(\partial\theta/\partial\sigma)^2 > 0$. This constant depends on the foreground dimension k^*, the Jacobian norms of the gauge-orbit map $G/H \to$

$[0, \pi/2]$, and the regularity constants in (C4); it is computable for any concrete model (see Remark 13.15 for the qubit case).

Step 2: small-drift expansion. Under freshly calibrated initial conditions ($\theta(t_0) \approx 0$), the mismatch angle grows as $\theta(\Delta t) = \Lambda\,\Delta t + O((\Delta t)^2)$ (Paper II, Eq. (35), linearised about $\theta = 0$). Here "freshly calibrated" means $\theta(t_0) \approx 0$ (not exactly zero); the linearization $\theta \approx \Lambda \cdot \Delta t$ is valid in the early-time regime $\Lambda \cdot \Delta t \ll 1$ of the exponential drift model. The visible survival functional satisfies $\mathcal{S}_{\text{vis}} = \mathcal{S}_{\text{tot}} \cos^2\theta \approx \mathcal{S}_{\text{tot}}\,(1 - \theta^2)$ for $\theta \ll 1$. Thus the residual distribution $p(e \,|\, \theta)$ shifts from its baseline $p(e \,|\, 0)$ by a score proportional to θ^2: $\partial_\theta \log p \sim 2\theta \cdot (\partial_\theta \log p)|_{\theta=\theta^*}$, and consequently

$$\mathcal{I}_F(\theta) \;\geq\; (\Lambda\,\Delta t)^2\, \mathcal{I}_F^{\text{env}}, \tag{13.14}$$

where the inequality retains only the leading $O(\theta^2)$ term and drops $O(\theta^4)$ corrections.

Step 3: assembly. Substituting (13.14) into (13.13):

$$\mathcal{I}_F(\sigma) \;\geq\; \kappa\,\Lambda^2\,(\Delta t)^2\,\mathcal{I}_F^{\text{env}}.$$

□

Remark 13.15 (Coupling efficiency in the qubit example)**.** *For the single-qubit model of Section 13.7, the gauge orbit is parametrised directly by θ (rotation in the xz-plane of the Bloch sphere), so $\partial\theta/\partial\sigma = 1$ and $\kappa = 1$. More generally, κ depends on the dimensionality of the gauge group and the curvature of the orbit G/H at the current frame.*

Corollary 13.16 (Detection before delusion)**.** *Under* (C1)–(C5)*, there exists a detection time Δt_{detect} satisfying*

$$\Delta t_{\text{detect}} \;=\; \frac{1}{\Lambda}\sqrt{\frac{\mathcal{I}_F^{\min}}{\kappa\,\mathcal{I}_F^{\text{env}}}} \;<\; t_{\text{del}}, \tag{13.15}$$

where $\mathcal{I}_F^{\min}$ is the minimum Fisher information required to exceed the noise floor (determined by h_μ and the observation window length). The detection window opens before *the Delusion Trap closes, provided the meta-observer budget $\mathcal{C}_{\text{meta}}$ is sufficient to compute $\mathcal{I}_F$.*

Proof. The detection time $\Delta t_{\text{detect}} \propto \Lambda^{-1}$ (from (13.12)), while $t_{\text{del}} = \Lambda^{-1}\ln(\pi/4\theta_0)$ (from (13.1)). Since $\ln(\pi/4\theta_0) > 1$ for $\theta_0 < \pi/4e$ and the constant $\kappa\,\mathcal{I}_F^{\text{env}}$ is finite under (C4), the square root in Δt_{detect} can be made smaller than the logarithm in t_{del} for sufficiently sensitive meta-observers (large $\mathcal{I}_F^{\text{env}}$). □

13.4 The Self-Referential Bound

13.4.1 The Bayesian Framework

The agent's ego structure (Paper II) provides a *prior belief* about the correct gauge parameter: the current frame σ_0 is the ego's "preferred" value. We encode this as a prior distribution $\pi_{\text{ego}}(\sigma)$, concentrated around σ_0.

Definition 13.17 (Ego rigidity)**.** *The* ego rigidity *is the prior Fisher information*

$$\mathcal{I}_{\text{ego}} := \int_{\mathcal{M}_G} \left(\frac{\partial \log \pi_{\text{ego}}(\sigma)}{\partial\sigma}\right)^2 \pi_{\text{ego}}(\sigma)\, d\sigma. \tag{13.16}$$

High $\mathcal{I}_{\text{ego}}$ corresponds to a sharply peaked prior (rigid ego); low $\mathcal{I}_{\text{ego}}$ to a diffuse prior (flexible ego). The four bias terms of Paper II contribute to $\mathcal{I}_{\text{ego}}$: $\mathcal{B}_{\text{select}}$ and $\mathcal{B}_{\text{frame}}$ sharpen the prior around the current basis and connection, while $\mathcal{B}_{\text{center}}$ centres the prior on the agent's own state.

13.4.2 The Self-Referential Cramér–Rao Bound

Theorem 13.18 (Self-Referential Cramér–Rao Bound). *Under assumptions* (C1)–(C5), *let $\delta\hat{\sigma}$ be any estimator of the frame drift $\delta\sigma := \sigma^*(t) - \sigma$, based on a residual record of duration T. Define the* effective sample size $n_{\text{eff}} := T/\tau_E$, *where τ_E is the decorrelation time of the residual process $\{e(t)\}$ (the time beyond which consecutive residuals carry approximately independent information about σ). Then the van Trees inequality [124] gives*

$$\mathbb{E}\Big[\big|\delta\hat{\sigma} - \delta\sigma\big|^2\Big] \;\geq\; \frac{1}{n_{\text{eff}}\,\mathcal{I}_F(\sigma) + \mathcal{I}_{\text{ego}}}. \tag{13.17}$$

Proof. This is a direct application of the van Trees (Bayesian Cramér–Rao) inequality [124]. The total information about the drift parameter $\delta\sigma$ consists of two contributions:

- $n_{\text{eff}}\,\mathcal{I}_F(\sigma)$: the data Fisher information. In continuous time, the residual process is correlated with decorrelation time τ_E set by the bath memory kernel. Over a window of duration T, the process yields $n_{\text{eff}} \approx T/\tau_E$ effectively independent samples, each carrying $\mathcal{I}_F(\sigma)$ bits of information about $\delta\sigma$.
- $\mathcal{I}_{\text{ego}}$: the prior Fisher information from the ego's preference for σ_0 (Definition 13.17).

The van Trees inequality states that the Bayesian mean-squared error is bounded below by the inverse of the total information. □

Remark 13.19 (Compact manifold). *For gauge parameters on compact manifolds (e.g., $\sigma \in SO(3)$), the van Trees inequality holds provided the prior is smooth and supported away from coordinate chart boundaries, which is generically satisfied for the gauge manifolds considered here.*

Remark 13.20 (The ego as help and hindrance). *The ego rigidity $\mathcal{I}_{\text{ego}}$ acts as both help and hindrance:*

- ***Help**: when the ego is well-aligned ($\sigma_0 \approx \sigma^*$), the prior tightens the bound, reducing estimation variance.*
- ***Hindrance**: when the ego is misaligned ($|\sigma_0 - \sigma^*|$ large), the prior pulls the estimate toward the wrong value, creating a* confirmation bias *that resists recalibration.*

The optimal Bayesian estimator balances data and prior:

$$\hat{\sigma}_{\text{opt}} = \frac{n_{\text{eff}}\,\mathcal{I}_F\,\hat{\sigma}_{\text{MLE}} + \mathcal{I}_{\text{ego}}\,\sigma_0}{n_{\text{eff}}\,\mathcal{I}_F + \mathcal{I}_{\text{ego}}}, \tag{13.18}$$

a weighted average of the maximum-likelihood estimate $\hat{\sigma}_{\text{MLE}}$ and the ego's prior belief σ_0, with weights proportional to their respective Fisher informations. As $n_{\text{eff}}\,\mathcal{I}_F \gg \mathcal{I}_{\text{ego}}$ (enough data to overwhelm the ego), the estimator converges to the MLE.

13.4.3 The Rigidity-Sensitivity Trade-off

Proposition 13.21 (Optimal ego rigidity). *Let the total expected loss be* $\mathcal{L}_{\text{total}}(\mathcal{I}_{\text{ego}}) = \mathcal{L}_{\text{estimation}} + \lambda\,\mathcal{L}_{\text{calibration}}$, *where* $\mathcal{L}_{\text{estimation}}$ *is the mean-squared drift-estimation error (bounded by* (13.17)*) and* $\mathcal{L}_{\text{calibration}}$ *is the cost of adjusting the frame (proportional to the frame rotation distance, hence larger when the ego is rigid and must be overcome). Under* (C1)–(C5), *there exists an optimal ego rigidity* $\mathcal{I}^*_{\text{ego}}$ *that minimises* $\mathcal{L}_{\text{total}}$.

Too rigid *(*$\mathcal{I}_{\text{ego}} \gg n_{\text{eff}}\,\mathcal{I}_F$*): the ego overwhelms the data; the agent is blind to drift.* Too soft *(*$\mathcal{I}_{\text{ego}} \ll n_{\text{eff}}\,\mathcal{I}_F$*): the agent overreacts to noise; calibration cost is high. The optimum balances sensitivity against stability.*

Proof. The estimation loss decreases with $\mathcal{I}_{\text{ego}}$ (the prior sharpens the bound (13.17) when $\sigma_0 \approx \sigma^*$ but increases it when misaligned). The calibration cost increases with $\mathcal{I}_{\text{ego}}$ (a rigid ego resists rotation). The sum is a convex function of $\mathcal{I}_{\text{ego}}$ under standard regularity, so a minimum exists. □

13.5 The Calibration Loop

13.5.1 The Natural Gradient Update Law

The meta-observer updates the gauge parameter σ following the natural gradient on the statistical manifold $(\mathcal{M}_G,\, g)$:

$$\dot\sigma = -\eta\, g^{-1}(\sigma)\, \nabla_\sigma L_{\text{frame}}(\sigma), \tag{13.19}$$

where $\eta > 0$ is the adaptation rate and the *frame loss* is

$$L_{\text{frame}}(\sigma) := \mathbb{E}_\sigma[-\mathcal{S}_{\text{vis}}(\sigma)]\,. \tag{13.20}$$

The frame loss is minimised at the optimal gauge σ^* that maximises visible survival. The Fisher metric enters through the inverse g^{-1} in the natural gradient, not as a penalty term: it defines the *geometry* of the update, ensuring reparametrisation invariance.

Remark 13.22 (Reparametrisation invariance). *The natural gradient* (13.19) *is invariant under reparametrisation of the gauge manifold* $\mathcal{M}_G$*: the update direction does not depend on the choice of coordinates for* σ*. This is essential because the gauge manifold inherits its geometry from the Clifford algebra, and no canonical coordinate system is preferred.*

13.5.2 Lyapunov Stability of the Loop

Drift velocity. Let $\sigma^*(t)$ denote the instantaneous optimal gauge parameter (the minimiser of L_{frame} at time t; Paper II, Definition 27). Define the *drift velocity* $\dot\sigma^* := d\sigma^*/dt$, measured with respect to the Fisher metric g; its norm $\|\dot\sigma^*\|_g := \sqrt{g_{ij}\,\dot\sigma^{*i}\,\dot\sigma^{*j}}$ is the instantaneous rate at which the environment's optimal frame rotates on the gauge manifold.

Definition 13.23 (Lyapunov monitoring function). *The* Lyapunov monitoring function *is the squared geodesic distance on the statistical manifold from the current frame to the instantaneous optimal frame:*

$$V(\sigma) := d_{\text{geo}}(\sigma,\, \sigma^*(t))^2\,, \tag{13.21}$$

where d_{geo} *is the geodesic distance in the Fisher metric* g.

Theorem 13.24 (Loop Tracking Bound)**.** *Under assumptions* (C1)–(C5)*, the natural gradient update* (13.19) *applied to the Lyapunov monitoring function* (13.21) *satisfies*

$$\frac{dV}{dt} \;\le\; -2\eta\,\alpha\,V \;+\; 2\sqrt{V}\;\big\|\dot{\sigma}^*\big\|_g, \tag{13.22}$$

where $\alpha > 0$ is the persistent excitation constant (Definition 13.25 below) and $\|\dot{\sigma}^*\|_g$ is the instantaneous drift speed of the optimal frame. Consequently:

(a) ***Tracking.*** *Whenever $\sqrt{V} > \|\dot{\sigma}^*\|_g/(\eta\,\alpha)$, we have $dV/dt < 0$: the loop actively reduces the mismatch.*

(b) ***Tracking neighbourhood.*** *The mismatch converges to a neighbourhood of the set of stationary points of L_{frame}. Assuming non-degeneracy (local strong convexity near σ^*, consistent with persistent excitation (C5) in standard adaptive-control settings [9]), this neighbourhood has size*

$$V_\infty \;:=\; \frac{\big\|\dot{\sigma}^*\big\|_g^2}{(\eta\,\alpha)^2}. \tag{13.23}$$

For bounded drift ($\|\dot{\sigma}^*\|_g \le \Lambda_{\max}$), the mismatch is bounded: $\limsup_{t\to\infty} V(t) \le \Lambda_{\max}^2/(\eta\alpha)^2$.

(c) ***Static limit.*** *When $\sigma^* = \text{const}$ ($\dot{\sigma}^* = 0$), the bound reduces to $dV/dt \le -2\eta\alpha\,V$, giving exponential convergence $V(t) \le V(0)\,e^{-2\eta\alpha\,t}$.*

Proof. Since $V(\sigma) = d_{\text{geo}}(\sigma, \sigma^*(t))^2$ and $\sigma^*(t)$ is time-varying, the total derivative has two contributions:

$$\frac{dV}{dt} = \underbrace{\frac{\partial V}{\partial \sigma}\cdot\dot{\sigma}}_{\text{control}} + \underbrace{\frac{\partial V}{\partial \sigma^*}\cdot\dot{\sigma}^*}_{\text{drift}}.$$

Control term. In normal coordinates centred at σ^*, let $\delta\sigma := \sigma - \sigma^*$. The control contribution is $2\,g(\delta\sigma,\,\dot{\sigma}) = 2\,g(\delta\sigma,\,-\eta\,g^{-1}\nabla L_{\text{frame}}) = -2\eta\,\langle\delta\sigma,\,\nabla L_{\text{frame}}\rangle$. Since σ^* minimises L_{frame} by definition, L_{frame} is locally strongly convex near σ^* under persistent excitation (C5) (the Hessian of L_{frame} at σ^* is bounded below by $\alpha\,g$, where α is the persistent excitation constant). Therefore $\langle\delta\sigma,\,\nabla L_{\text{frame}}\rangle \ge \alpha\,|\delta\sigma|_g^2 = \alpha\,V$, giving a control contribution $\le -2\eta\,\alpha\,V$.

Drift term. The drift contribution is $-2\,g(\delta\sigma,\,\dot{\sigma}^*)$. By Cauchy–Schwarz: $|g(\delta\sigma,\,\dot{\sigma}^*)| \le |\delta\sigma|_g\,\|\dot{\sigma}^*\|_g = \sqrt{V}\,\|\dot{\sigma}^*\|_g$. Hence the drift contribution is bounded by $+2\sqrt{V}\,\|\dot{\sigma}^*\|_g$.

Combined. Adding both contributions gives (13.22). Part (a) follows by setting $dV/dt < 0$; part (b) by solving $dV/dt = 0$ for the fixed point $\sqrt{V_\infty} = \|\dot{\sigma}^*\|_g/(\eta\alpha)$; part (c) by setting $\dot{\sigma}^* = 0$. ∎

13.5.3 Convergence Rate under Persistent Excitation

Definition 13.25 (Persistent excitation constant). *The* persistent excitation constant $\alpha > 0$ *is the minimum eigenvalue of the time-averaged Fisher information matrix:*

$$\bar{g}(t) := \frac{1}{T}\int_t^{t+T} g(\sigma(s))\, ds \;\succeq\; \alpha\, I \qquad \textit{for all } t, \tag{13.24}$$

guaranteed to exist by (C5) *and* (C4)*.*

Remark 13.26 (Tracking vs convergence). *In the static case ($\sigma^* = \text{const}$), Theorem 13.24(c) gives pure exponential convergence: $V(t) \leq V(0)\, e^{-2\eta\alpha\, t}$. Under environmental drift, convergence to zero is* not *possible—instead the loop maintains the mismatch within the tracking neighbourhood* (13.23)*. The tracking error V_∞ grows with drift speed $\|\dot{\sigma}^*\|_g$ and decreases with loop parameters η and α. If the free-energy budget is insufficient to maintain $\eta\,\alpha > \Lambda$ (the drift rate), the tracking neighbourhood expands and the Delusion Trap re-emerges. This connects the Lyapunov stability of the loop directly to the thermodynamic budget (Section 13.6).*

Remark 13.27 (The necessity of novelty). *If $h_\mu \to 0$ (the environment ceases to generate new information), the persistent excitation constant $\alpha \to 0$ and the tracking neighbourhood $V_\infty = \|\dot{\sigma}^*\|_g^2/(\eta\alpha)^2 \to \infty$: the loop loses all ability to track.* Memory without novelty cannot sustain self-reference. *This is the information-theoretic expression of a basic physical principle: a system in thermodynamic equilibrium cannot "learn" about itself.*

13.5.4 The Four-Part Structure Proposition

We are now in a position to state the capstone result of the T-DOME sequence.

Proposition 13.28 (Sufficient Architecture for Persistent Agents). *Within the class of agents satisfying* (C1)–(C5)*, a sufficient architecture for maintaining a non-equilibrium steady state* (NESS) *in an open, drifting environment under bounded computation comprises the following four structural layers:*

(I) ***External observable geometry.*** *The environmental observable algebra supports a metric structure; $Cl(1,3)$ serves as the running example throughout the programme, but the argument applies to any algebra satisfying (C1).* Assumption: *established in [127, 128, 140]; adopted here as a modelling premise.*

(II) ***Internal control algebra.*** *The agent carries an internal algebra isomorphic to $Cl(V,q)$ with realizability embedding $\phi : Cl(V,q) \hookrightarrow Cl(1,3)$.* Assumption: *established in [141, 142]; adopted here as a modelling premise.*

(III) ***Self-monitoring function.*** *The agent maintains a Lyapunov function $V(\sigma)$* (13.21) *satisfying the tracking bound* (13.22)*, via a second-order control loop on the agent's Fisher information, keeping the mismatch within the tracking neighbourhood* (13.23). Source: *this paper, Theorem 13.24.*

(IV) ***Biased, non-Markovian memory.*** *The agent carries path-dependent state (non-Markovian memory kernel $\mathcal{K}(t,s)$) compressed through a gauge-fixed reference frame (the ego $\mathfrak{E}$).* Source: *Paper I [143] (memory necessity) and Paper II [144] (ego necessity).*

Without any one of the four layers, the agent fails:

- *Without (I): no physical embedding—the agent cannot interact with the Lorentzian environment.*
- *Without (II): no channel discrimination—the agent cannot distinguish survival-relevant from irrelevant information.*
- *Without (III): the Delusion Trap—the ego rigidifies and prediction error diverges exponentially.*
- *Without (IV): the Markovian Ceiling and computational paralysis—no temporal accumulation, no tractable processing.*

Proof. Layers (I) and (II) are modelling assumptions adopted from [127, 128, 140, 141, 142]; their sufficiency within those frameworks is established therein. The sufficiency of (III) follows from the present paper: Theorem 13.10 shows that first-order control is insufficient to escape the Delusion Trap, and Theorem 13.24 shows that the tracking bound is sufficient. The sufficiency of (IV) follows from Paper I [143] (Markovian Ceiling $\mathcal{S} \leq 0$) and Paper II [144] (Computational Ceiling and necessity of SSB).

The "without" claims follow from the respective crisis theorems: Paper I's Theorem 14 (Markovian Ceiling), Paper II's Theorem 7 (Computational Ceiling) and Theorem 29 (Delusion Trap), and the present Theorem 13.10. □

13.6 Thermodynamic Cost

13.6.1 The Three Cost Components

The self-referential calibration loop requires three distinct operations, each carrying an irreducible thermodynamic cost:

1. Sensing cost. The meta-observer must read the prediction residuals from the ego's processing pipeline. This requires monitoring k^* foreground channels, each producing h_μ bits per unit time:

$$\dot{W}_{\text{sense}} \;\geq\; k_B T \,\ln 2 \cdot h_\mu \, k^*. \tag{13.25}$$

(Landauer cost of reading $h_\mu \, k^*$ bits per unit time.)

2. Computing cost. Evaluating the Fisher information $\mathcal{I}_F(\sigma)$ from the residual stream requires the meta-observer to process $\mathcal{C}_{\text{meta}}$ bits per unit time:

$$\dot{W}_{\text{compute}} \;\geq\; k_B T \,\ln 2 \cdot \mathcal{C}_{\text{meta}}. \tag{13.26}$$

3. Actuating cost. Rotating the gauge parameter from the current frame σ to the estimated optimal frame $\hat{\sigma}^*$ is a finite-time thermodynamic transformation on the gauge manifold. By the Sivak–Crooks bound (Proposition 13.7):

$$\dot{W}_{\text{actuate}} \;\geq\; \frac{\mathcal{L}^2(\sigma,\, \hat{\sigma}^*)}{\tau_{\text{recalib}}^2}, \tag{13.27}$$

where $\mathcal{L}(\sigma,\, \hat{\sigma}^*)$ is the thermodynamic length (13.7) of the geodesic from σ to $\hat{\sigma}^*$, and τ_{recalib} is the recalibration time.

13.6.2 The Thermodynamic Cost Theorem

Theorem 13.29 (Thermodynamic Cost of Self-Referential Calibration). *Under assumptions* (C1)–(C5)*, the minimum dissipation rate of the self-referential calibration loop satisfies*

$$\dot{W}_{\text{loop}} \;\geq\; k_B T \,\ln 2\,[h_\mu\, k^* + \mathcal{C}_{\text{meta}}] + \frac{\mathcal{L}^2(\sigma,\,\sigma^*)}{\tau^2_{\text{recalib}}}. \tag{13.28}$$

The first bracketed term is the information tax *(the Landauer cost of sensing and computing). The second term is the* geometric tax *(the Sivak–Crooks cost of actuating the frame rotation).*

Proof. We must establish that the three lower bounds can be summed, i.e. that no single physical process can simultaneously satisfy two or more of them.

The three operations act on *disjoint physical degrees of freedom*:

1. *Sensing* reads the prediction residuals $\{e_t\}$ from the ego's foreground channels. The relevant degrees of freedom are the sensor registers that copy bits from the foreground subspace V_{fg}. Each bit erased carries the Landauer cost $k_B T \ln 2$.

2. *Computing* evaluates the Fisher information $\mathcal{I}_F(\sigma)$ from the copied residuals. The relevant degrees of freedom are the processor logic states of the meta-observer. These are distinct from the sensor registers: the processor manipulates the data *after* it has been read, and its own state transitions carry an independent Landauer cost.

3. *Actuating* rotates the gauge parameter from σ to $\hat{\sigma}^*$. The relevant degrees of freedom are the control fields that implement the frame rotation on the agent's internal algebra $Cl(V, q)$. This is a physical transformation of the agent's hardware state, governed by the Sivak–Crooks bound on finite-time thermodynamic transformations. The $\tau^{-2}_{\text{recalib}}$ scaling of the dissipation *rate* follows from the Sivak–Crooks bound $W_{\text{ex}} \geq \mathcal{L}^2/\tau$ (excess *work*), divided by τ_{recalib} to convert to a rate.

Under the assumption that the three operations are physically realised on separable degrees of freedom (no shared erasure accounting), the sets are disjoint (sensor $\cap$ processor $= \emptyset$, processor $\cap$ actuator $= \emptyset$, sensor $\cap$ actuator $= \emptyset$), and the Landauer bound for each is independent. Moreover, the actuating cost involves a different *type* of bound (thermodynamic length, not Landauer erasure), reinforcing the independence. The total lower bound is therefore the sum of the three individual bounds (13.25)–(13.27). □

Remark 13.30 (Recalibration schedule). *The bound assumes periodic recalibration at interval τ_{recalib}. For threshold-triggered schedules, the Sivak–Crooks bound applies to each inter-recalibration interval individually, providing a lower bound on the time-averaged dissipation rate.*

13.6.3 The Complete Persistence Budget

Corollary 13.31 (Persistence Budget). *Combining the results of Papers I, II, and III, the minimum free-energy dissipation rate for a persistent, self-calibrating agent in a drifting environment is*

$$\dot{W}_{\text{total}} \;\geq\; \underbrace{k_B T\,\ln 2\cdot h_\mu}_{\textit{Paper I: memory}} + \underbrace{k_B T\,\ln 2\cdot h_\mu\, k^*}_{\textit{Paper II: ego processing}} \underbrace{+\,k_B T\,\ln 2\,[h_\mu\, k^* + \mathcal{C}_{\text{meta}}] + \frac{\mathcal{L}^2}{\tau^2_{\text{recalib}}}}_{\textit{Paper III: self-calibration loop}}. \tag{13.29}$$

Below this budget, the agent must sacrifice one or more of the four structural layers (Proposition 13.28): losing memory (Paper I crisis), losing the ego (Paper II crisis), or losing self-calibration (Paper III crisis, the Delusion Trap).

Remark 13.32 (Potential double-counting)**.** *The terms are additive under the assumption that the ego's foreground processing and the meta-observer's sensing operate on disjoint physical registers. If these share physical resources, the total may be an upper bound on the true minimum dissipation.*

Remark 13.33 (The cost of selfhood)**.** *Equation* (13.29) *is the first explicit, calculable lower bound on the thermodynamic cost of maintaining a self-referential agent in a drifting environment. It shows that "selfhood" is not free: the ego (Paper II) and its calibration loop (Paper III) each add irreducible energy taxes on top of the memory cost (Paper I). The total cost grows with the environmental complexity (h_μ), the agent's representational capacity (k^*), the meta-observer's computational power ($\mathcal{C}_{\text{meta}}$), and the drift rate (through $\mathcal{L}$ and τ_{recalib}).*

13.7 Worked Example: Qubit in a Drifting Two-Channel Bath

13.7.1 Model Setup

We extend the two-channel qubit model from Paper II (Section 6) by introducing environmental drift.

Inherited setup. A qubit ($\dim \mathcal{H}_S = 2$) with internal algebra $Cl(0,2) \cong \mathbb{H}$ ($D = 4$), coupled to two bosonic channels:

- Dephasing channel (σ_z): $J_z(\omega) = 2\lambda_z\gamma_z\omega/(\omega^2 + \gamma_z^2)$.
- Dissipative channel (σ_x): $J_x(\omega) = 2\lambda_x\gamma_x\omega/(\omega^2 + \gamma_x^2)$.

Paper II's ego selects $V_{\text{fg}} = \text{span}\{1, \mathbf{k}\}$ (the dephasing subspace), discarding $V_{\text{bg}} = \text{span}\{\mathbf{i}, \mathbf{j}\}$.

Environmental drift. We now allow the dephasing coupling to drift exponentially (matching Paper II's Delusion Trap analysis):

$$\lambda_z(t) = \lambda_z^{(0)}\big(1 + \theta_0\, e^{\Lambda t}\big), \qquad \theta_0 = 0.02, \quad \Lambda = 0.08\,\omega_0. \tag{13.30}$$

The optimal frame $\mathcal{F}^*(t)$ rotates in $SO(3)$ as the relative survival values of the two channels change. The Delusion Trap time $t_{\text{del}} = \Lambda^{-1}\ln\!\big(\pi/(4\theta_0)\big) \approx 45.9\,\omega_0^{-1}$.

Parameter mapping.

Quantity	Value	Source
$D = \dim Cl(0,2)$	4	Paper II
k^*	2	Paper II, Theorem 17
$\mathcal{C}_{\text{budget}}$	$2\,h_\mu$	Paper II
θ_0 (initial misalignment)	0.02	this example
Λ (drift rate)	$0.08\,\omega_0$	Eq. (13.30)
t_{del}	$45.9\,\omega_0^{-1}$	Paper II, Delusion Trap
η (adaptation rate)	0.5	meta-observer

13.7.2 Fisher Information under Drift

As the coupling $\lambda_z(t)$ drifts, the decoherence function $p_z(t)$ (Paper II, Eq. (34)) changes, shifting the residual distribution. The self-referential Fisher information $\mathcal{I}_F(\sigma)$ measures this shift.

For the qubit model, the Fisher information with respect to the frame angle ϕ (parametrising the $SO(3)$ rotation between the current and optimal frames) is

$$\mathcal{I}_F(\phi) = \frac{(\partial_\phi\,\bar{e})^2}{\mathrm{Var}(e)} \approx \frac{4\,\mathcal{S}_{\text{tot}}^2\,\theta^2}{h_\mu/n_{\text{eff}}}, \tag{13.31}$$

where $\bar{e} = \mathbb{E}[e \,|\, \phi]$ is the expected residual, $\theta = \theta(\phi)$ is the mismatch angle, and n_{eff} is the effective sample size (Remark 13.4).

When the frame is well-aligned ($\theta \approx 0$): $\mathcal{I}_F \approx 0$. As drift accumulates (θ grows): $\mathcal{I}_F$ increases quadratically, producing a detectable "stress signal" consistent with Theorem 13.14.

13.7.3 Loop Dynamics: Self-Calibration in Action

Under the natural gradient update (13.19), the frame angle $\phi(t)$ tracks the drifting optimal frame $\phi^*(t)$. The Lyapunov function $V(t) = (\phi(t) - \phi^*(t))^2$ is governed by the tracking bound (13.22): the loop drives V toward the tracking neighbourhood $V_\infty = \|\dot{\sigma}^*\|_g^2/(\eta\alpha)^2$, with the approach rate set by the persistent excitation constant α and the adaptation rate η.

Comparison.

- **Without loop** (Paper II agent): the mismatch grows as $\theta(t) = \theta_0\,e^{\Lambda t}$, reaching $\pi/4$ at t_{del}. The agent is delusional.
- **With loop** (Paper III agent): the mismatch oscillates around zero, bounded by the estimation noise floor $\theta_{\min} \sim 1/\sqrt{n_{\text{eff}}\,\mathcal{I}_F^{\text{env}}}$ (the Cramér–Rao limit). The agent remains calibrated.

A multi-dimensional numerical evaluation extending this qubit illustration to continuous drift is presented in Section 13.8.

13.7.4 Thermodynamic Cost Evaluation

For the qubit example with $k^* = 2$, $h_\mu = 1$ (normalised), $\mathcal{C}_{\text{meta}} = 1\, h_\mu$ (minimal meta-observer):

$$\dot{W}_{\text{sense}} \geq k_B T \, \ln 2 \cdot 1 \cdot 2 = 2\, k_B T \, \ln 2, \tag{13.32}$$

$$\dot{W}_{\text{compute}} \geq k_B T \, \ln 2 \cdot 1 = k_B T \, \ln 2, \tag{13.33}$$

$$\dot{W}_{\text{actuate}} \geq \frac{\mathcal{L}^2}{\tau_{\text{recalib}}^2} \approx \frac{\theta_0^2 \, \tau_{\text{relax}}}{\tau_{\text{recalib}}^2} \, k_B T. \tag{13.34}$$

The total loop cost is dominated by the information tax (sensing + computing) at $\sim 3\, k_B T \, \ln 2$ per unit time, with the geometric tax (actuating) contributing a smaller correction proportional to θ_0^2.

For comparison, Paper I's memory cost is $\dot{W}_{\text{mem}} \geq k_B T \, \ln 2$ and Paper II's ego processing cost is $\dot{W}_{\text{ego}} \sim 2\, k_B T \, \ln 2 \cdot h_\mu$. The self-calibration loop adds approximately 50% to the total energy budget—a significant but bounded cost for escaping the Delusion Trap.

13.8 Numerical Demonstration

The preceding sections establish analytic bounds and a low-dimensional worked example. We now demonstrate computationally that the three core phenomena—delusion separation, detectable staleness, and an optimal calibration budget—emerge in a minimal multi-dimensional system under continuous drift. Full code and parameters are provided for reproducibility.

13.8.1 Model

Environment. A d-dimensional linear prediction task: $y(t) = \mathbf{w}(t)^\top \mathbf{x}(t) + \sigma\, \epsilon(t)$, $\mathbf{x}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, $\epsilon \sim \mathcal{N}(0, 1)$. The weight vector $\mathbf{w}(t) \in \mathbb{S}^{d-1}$ drifts by receiving random perturbations on *background* dimensions only (indices $k, \ldots, d-1$), then renormalising. Signal therefore migrates progressively from the ego's foreground to its blind sector.

Agents.

- **Fixed ego** (Paper II analogue): learns a linear model on a *fixed* foreground subspace of dimension k via stochastic gradient descent (SGD, rate η, decay λ). Embodies the "frozen gauge" of Paper II.
- **Calibrated loop** (Paper III analogue): identical ego plus a staleness sentinel and recalibration mechanism. The sentinel tracks $g_i = \text{EMA}(|e\, x_i|)$ for each dimension i (an absolute-gradient proxy), computes the fraction of top-k gradient dimensions *not* in the current foreground as a frame-staleness index $m \in [0, 1]$, and triggers recalibration when $\text{EMA}(m) > \theta$. After recalibration, a settling period of τ steps elapses before the sentinel resumes monitoring.

Parameters.

Quantity	Value	Role
d	20	full ambient dimension
k	5	ego foreground dimension ($k/d = 0.25$)
σ	0.1	observation noise std
η	0.01	SGD learning rate
λ	0.998	SGD weight decay
θ	0.25	staleness threshold
Λ	variable	drift rate per step
τ	variable	settling period (cooldown)

Oracle metrics. Neither agent has access to $\mathbf{w}(t)$. We evaluate performance externally using the *oracle full-space error*:

$$\mathcal{E}_{\text{full}} = \left\|\mathbf{w}_{\text{ego}} - \mathbf{w}^*_{\text{fg}}\right\|^2 + \left\|\mathbf{w}^*_{\text{bg}}\right\|^2, \tag{13.35}$$

where $\mathbf{w}^*_{\text{fg}}$ and $\mathbf{w}^*_{\text{bg}}$ denote the true weight vector restricted to foreground and background coordinates respectively, and $\mathbf{w}_{\text{ego}}$ is the ego's foreground-supported estimator lifted to the full space. The first term captures foreground tracking error (accessible to the ego); the second captures hidden-sector signal (invisible).

13.8.2 Results

Result 1: Delusion-correction separation (Figure 13.1). At drift rate $\Lambda = 0.02$, settling period $\tau = 200$, and $T = 5\,000$ steps, three phenomena are visible:

(a) *Delusion trap.* The ego's foreground tracking error converges to near zero, while the true full-space error rises toward ~ 1 and stabilises. The growing gap between the two is the hidden sector, confirming the prediction of Theorem 13.10: first-order monitoring cannot detect frame drift.

(b) *Detectability.* The staleness sentinel produces a clean sawtooth: rising from zero after each recalibration, crossing the threshold $\theta = 0.25$, and triggering frame reset (25 events over $T = 5\,000$). This is consistent with the predicted growth trend of the self-referential Fisher signal (Theorem 13.14).

(c) *Net benefit.* The calibrated loop achieves $\mathcal{E}_{\text{full}} \approx 0.74$ versus the fixed ego's ≈ 1.02: a 27% reduction in true prediction error.

Result 2: Phase structure and optimal calibration budget (Figures 13.2–13.3). We scan 16 drift rates $\Lambda \in [0.005, 0.08]$ and 16 settling periods $\tau \in [15, 800]$ (logarithmically spaced), running both agents for $T = 4\,000$ steps across 6 random seeds per grid point.

Figure 13.2(a) shows the performance gain $\Delta = \mathcal{E}_{\text{ego}} - \mathcal{E}_{\text{loop}}$: the loop improves over the ego (green) across most of the parameter space, with a boundary at $\Delta = 0$ (dashed) below which recalibration is counterproductive (very low drift, where the overhead of re-learning exceeds the benefit of tracking). The solid curve traces the *optimal settling period* $\tau_{\text{opt}}(\Lambda)$—the recalibration period minimising $\mathcal{E}_{\text{loop}}$—which decreases monotonically from $\sim$370 steps at $\Lambda = 0.005$ to $\sim$100 at $\Lambda = 0.08$. Figure 13.2(b) shows that calibration frequency increases smoothly with drift and with shorter settling period, exhibiting the cost–performance trade-off of Theorem 13.29.

Extracting $\tau_{\text{opt}}(\Lambda)$ yields the *optimal calibration frequency* $\alpha_{\text{opt}}(\Lambda) = 1/\tau_{\text{opt}}$ (Figure 13.3). The curve is smooth and monotonically increasing: faster drift demands tighter calibration. It saturates at high Λ near $\alpha_{\text{opt}} \approx 0.01$ per step ($\tau_{\text{opt}} \approx 100$), of the same order as the learner's settling time. This is consistent with the intuition that drift estimation requires a minimum observation window; the self-referential Cramér–Rao bound (Theorem 13.18) provides the analytic counterpart of this computational floor.

13.8.3 Scope of This Demonstration

This demonstration **does** show:

1. The delusion-correction separation predicted by Theorems 13.10 and 13.14 emerges in a minimal stochastic system with continuous drift.
2. A frame-staleness signal with clean threshold dynamics exists and triggers effective recalibration.
3. An optimal calibration frequency $\alpha_{\text{opt}}(\Lambda)$ exists, increases monotonically with drift rate, and saturates at the learner's settling timescale.
4. The cost–performance trade-off of Theorem 13.29 manifests as a structured phase diagram with an explicit τ_{opt} boundary.

In summary, this demonstration validates the *existence* and *detectability* of the loop–cost trade-off in a minimal linear setting; it does not claim universality across architectures or environment classes.

This demonstration does **not** show:

1. That the specific functional form of $\alpha_{\text{opt}}(\Lambda)$ matches the analytic Cramér–Rao prediction in the large-d limit. The demonstration confirms the monotonic trend and saturation; deriving the exact scaling exponent from Theorem 13.18 remains open.
2. That the results generalise to all environment classes. The model uses Gaussian features, linear regression, and isotropic background drift; extensions to non-linear, non-Gaussian, or structured-drift settings require further investigation.
3. That the calibration loop is optimal among all possible adaptive strategies. It implements one specific realisation of the calibration-loop architecture.

Reproducibility. The complete simulation is a self-contained Python script (`tdome_demo.py`, $\sim$ 550 lines, requiring only NumPy and Matplotlib) with fixed random seeds. All figures in this section can be reproduced by executing the script after setting the output directory variable `BASE` to the desired path.

13.9 Discussion

13.9.1 Summary of Results

Result	**Statement**	**Sec.**
First-Order Insufficiency	Raw prediction error cannot detect frame drift	13.3.1
Drift Detectability	Self-referential Fisher information grows quadratically with accumulated drift	13.3.3
Self-Referential CR Bound	Drift estimation bounded by $1/(n_{\text{eff}}\,\mathcal{I}_F + \mathcal{I}_{\text{ego}})$	13.4.2
Loop Tracking Bound	Lyapunov V with tracking neighbourhood $V_\infty = \|\dot{\sigma}^*\|^2/(\eta\alpha)^2$	13.5.2
Four-Part Structure	Persistent agents require four structural layers	13.5.4
Loop Cost	$\dot{W}_{\text{loop}} \geq k_B T\,\ln 2\,[h_\mu\,k^* + \mathcal{C}_{\text{meta}}] + \mathcal{L}^2/\tau^2_{\text{recalib}}$	13.6.2
Persistence Budget	Total cost: memory + ego + loop	13.6.3
Numerical Demonstration	Delusion separation, sentinel detection, $\alpha_{\text{opt}}(\Lambda)$ boundary	13.8

13.9.2 The Complete Logic Chain

Papers I–III trace an irreversible thermodynamic logic chain:

Paper	**Crisis**	**Resolution**	**What is born**
Paper I	Markovian trap: no history	Non-Markovian memory	**Temporal accumulation**
Paper II	Computation explosion: ∞ memory, finite budget	Gauge SSB: $Cl(V,q) \to V_{\text{fg}} \oplus V_{\text{bg}}$	**Compressed ref. frame**
Paper III	Delusion trap: fixed bias, drifting world	Fisher self-referential calibration; tracking bound	**Reflexivity**

Each resolution creates the precondition for the next crisis. The chain terminates at Paper III: the self-referential calibration loop does not create a further crisis requiring a "Paper IV," because the loop is *self-correcting* by construction (Theorem 13.24). Its only vulnerability is the thermodynamic budget (Theorem 13.29): if the agent's free-energy supply falls below the persistence budget (13.29), the loop degrades and the Delusion Trap re-emerges. This is not a new crisis but the Second Law itself: all order requires free-energy dissipation.

13.9.3 What This Paper Does and Does Not Show

This paper **does** show:

1. Under environmental drift (C2) and bounded computation (C1), first-order control fails to detect frame drift (Theorem 13.10).
2. Self-referential Fisher information provides a quadratically growing signal sufficient for drift detection before the Delusion Trap (Theorem 13.14).
3. Drift estimation precision is bounded by the Self-Referential Cramér–Rao bound (Theorem 13.18).
4. The calibration loop tracks the optimal frame within a bounded neighbourhood under a Lyapunov tracking bound (Theorem 13.24).
5. The thermodynamic cost of the loop is calculable (Theorem 13.29).

This paper does **not** show:

1. That self-referential calibration implies or requires phenomenal consciousness, subjective experience, or qualia. "Reflexivity" as used here denotes second-order control, nothing more.
2. That the Lyapunov function V is a measure of "awareness." It is a control-theoretic stability condition, not a consciousness metric.
3. That the Four-Part Structure Proposition is a complete characterisation of agency. It states sufficient conditions under (C1)–(C5); other architectures may also suffice.
4. That Fisher information requires the agent to "know" it is computing Fisher information. The computation can be implemented implicitly by any physical system whose dynamics approximate the natural gradient.
5. That the calibration loop eliminates the ego's bias. It tracks and compensates for drift in the bias; the four bias terms of Paper II persist.
6. That the thermodynamic cost bounds are achievable by any specific physical implementation. They are information-theoretic lower bounds.
7. That this framework applies to all possible systems. It applies to systems satisfying (C1)–(C5).
8. That the structural parallel with philosophical concepts of self-awareness constitutes a philosophical or metaphysical claim.
9. That the Clifford algebra is the only possible algebraic setting. Other control algebras may yield analogous results with different quantitative bounds.

We have established a budgeted self-referential calibration loop that detects drift via an intrinsic Fisher signal, yields a falsifiable stability criterion, and incurs an unavoidable thermodynamic cost. In the context of Papers I–III, this completes the programme's third step by turning bias (Paper II) into a dynamically monitored and correctable quantity.

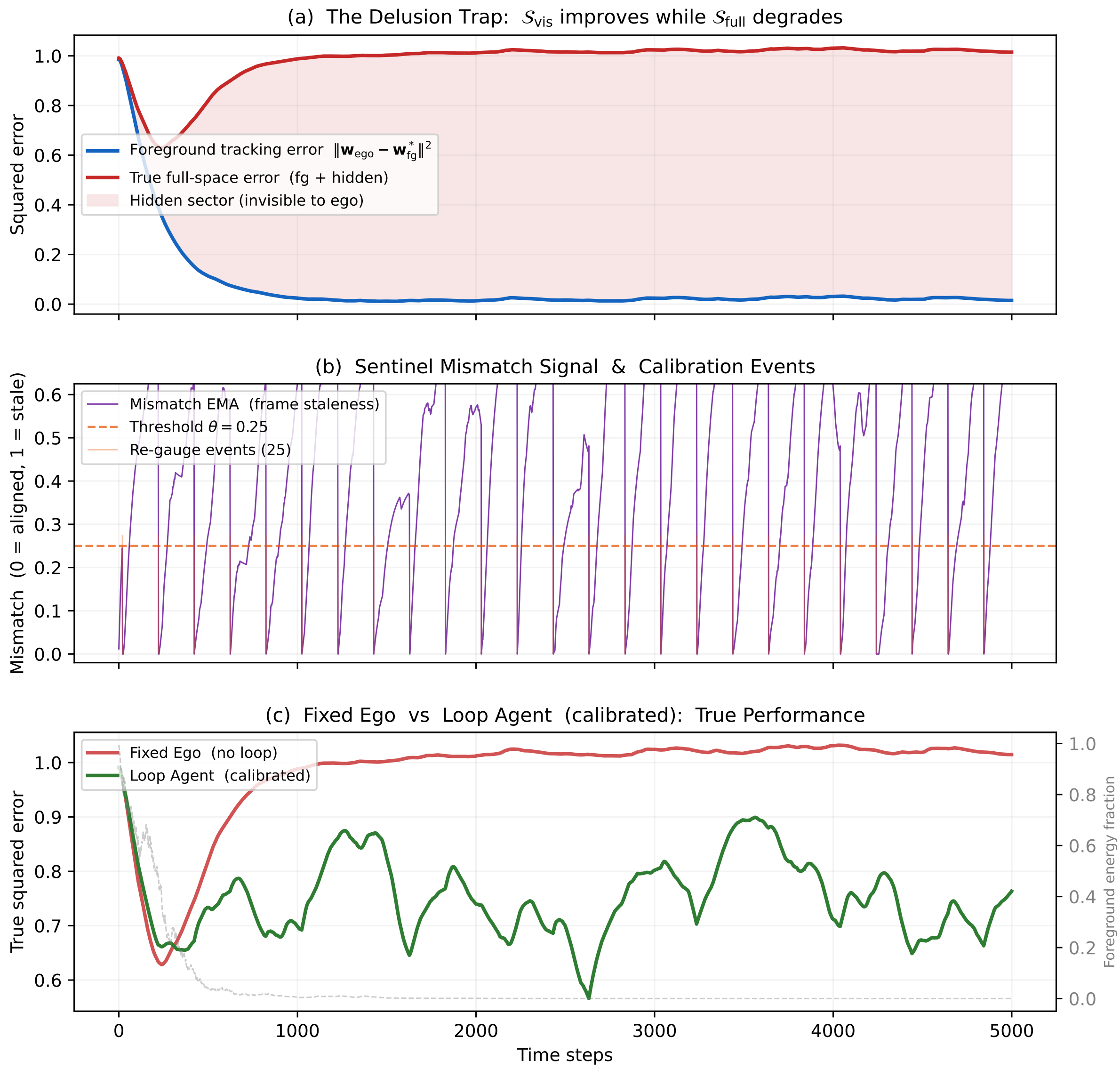

Figure 13.1: **Delusion trap and calibration loop.** $d = 20$, $k = 5$, $\Lambda = 0.02$, $\tau = 200$, $T = 5\,000$. **(a)** Foreground tracking error (blue) decreases toward zero while true full-space error (red) increases; the shaded region is the hidden sector, invisible to the ego. **(b)** Frame-staleness sentinel (purple) rises monotonically between recalibration events (orange), producing a sawtooth with 25 threshold crossings. **(c)** The calibrated loop (green) maintains lower true error than the fixed ego (red); the grey dashed line shows the foreground energy fraction decaying as signal migrates to the background.

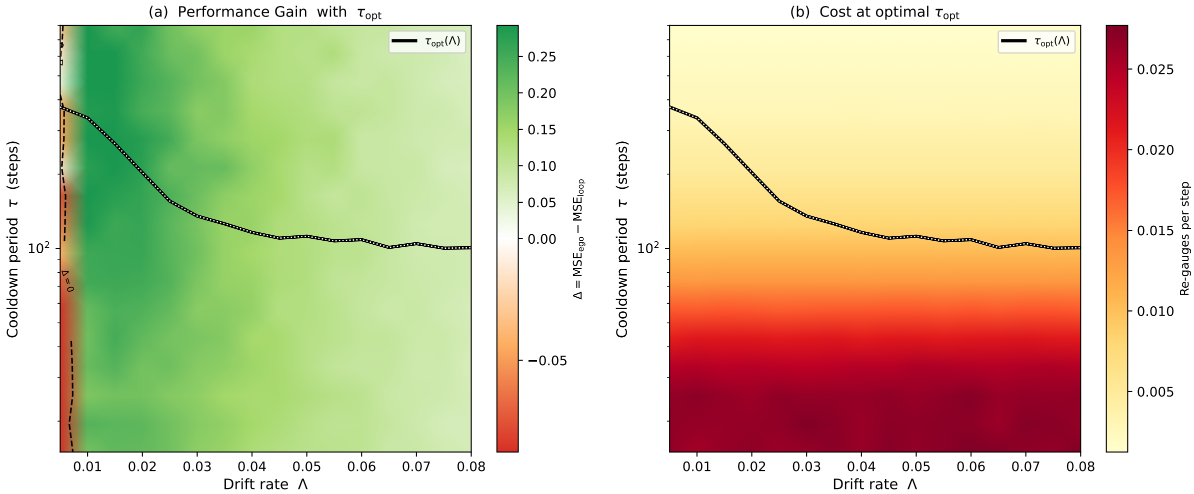

Figure 13.2: **Phase structure and calibration cost.** 16×16 grid, $T = 4\,000$, 6 seeds per point. **(a)** Performance gain Δ; green = loop improves on ego, red = counterproductive. Solid curve: $\tau_{\text{opt}}(\Lambda)$. Dashed: $\Delta = 0$ boundary. **(b)** Calibration frequency (thermodynamic cost proxy: recalibration events per step, proportional to energy expenditure under a fixed per-recalibration cost model); τ_{opt} overlaid.

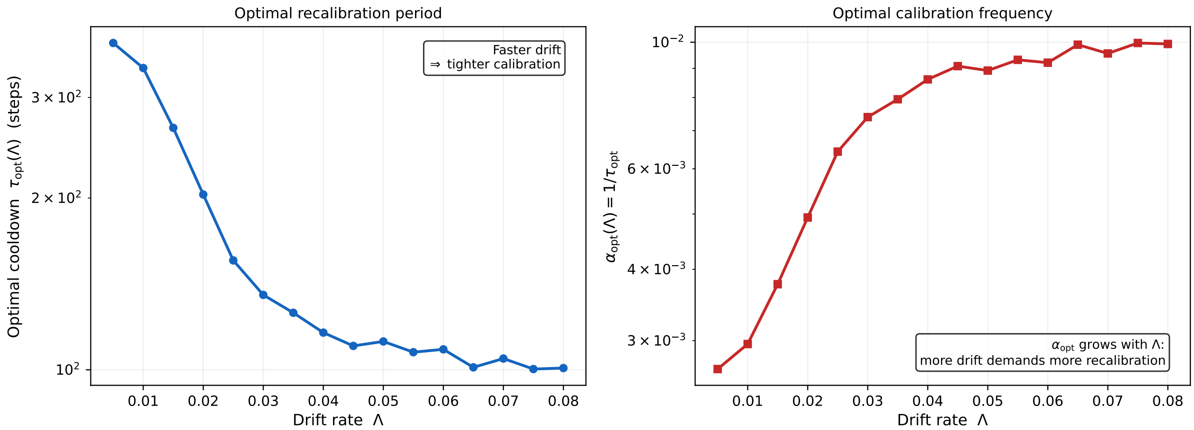

Figure 13.3: **Optimal calibration frequency. Left:** $\tau_{\text{opt}}(\Lambda)$ decreases monotonically with drift rate. **Right:** $\alpha_{\text{opt}}(\Lambda) = 1/\tau_{\text{opt}}$ increases with drift rate and saturates at the learner's settling timescale ($\sim$100 steps), consistent with an observation-window floor.

Acknowledgments

The author thanks the anonymous reviewers for their insightful comments and suggestions, which greatly improved the clarity and rigor of this work.

Bibliography

[1] S.-i. Amari, *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics **28**, Springer (1985).

[2] S.-i. Amari, *Natural gradient works efficiently in learning*, Neural Computation **10**, 251 (1998).

[3] S.-i. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs **191**, AMS (2000).

[4] M. F. Atiyah, R. Bott and A. Shapiro, *Clifford modules*, Topology **3**, Suppl. 1, 3–38 (1964).

[5] H. Araki, *Mathematical Theory of Quantum Fields*, Oxford University Press (1999).

[6] E. M. Alfsen and F. W. Shultz, *Geometry of State Spaces of Operator Algebras*, Birkhäuser (2003).

[7] Asaṅga, *Mahāyānasaṃgraha (Summary of the Great Vehicle)*, trans. É. Lamotte, Peeters (1973).

[8] W. R. Ashby, *An Introduction to Cybernetics*, Chapman & Hall (1956).

[9] K. J. Åström and B. Wittenmark, *Adaptive Control*, 2nd ed., Addison-Wesley (1995).

[10] A. C. Barato and U. Seifert, *Thermodynamic uncertainty relation for biomolecular processes*, Phys. Rev. Lett. **114**, 158101 (2015).

[11] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement*, Cambridge University Press (2006).

[12] C. H. Bennett, *The thermodynamics of computation—a review*, Int. J. Theor. Phys. **21**, 905 (1982).

[13] W. Bialek, I. Nemenman, and N. Tishby, *Predictability, complexity, and learning*, Neural Computation **13**, 2409 (2001).

[14] J. J. Bisognano and E. H. Wichmann, *On the duality condition for a Hermitian scalar field*, J. Math. Phys. **16**, 985 (1975).

[15] J. J. Bisognano and E. H. Wichmann, *On the duality condition for quantum fields*, J. Math. Phys. **17**, 303 (1976).

[16] L. Boltzmann, *Vorlesungen über Gastheorie*, J. A. Barth, Leipzig (1896).

[17] H.-P. Breuer, E.-M. Laine, and J. Piilo, *Measure for the Degree of Non-Markovian Behavior of Quantum Processes in Open Systems*, Phys. Rev. Lett. **103**, 210401 (2009).

[18] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems*, Oxford University Press (2002).

[19] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics*, Vols. 1–2, 2nd ed., Springer (1997).

[20] S. Carroll, *From Eternity to Here: The Quest for the Ultimate Theory of Time*, Dutton (2010).

[21] N. N. Čencov, *Statistical Decision Rules and Optimal Inference*, Translations of Mathematical Monographs **53**, AMS (1982 [orig. 1972]).

[22] D. J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press (1996).

[23] H. Casini, M. Huerta, and R. C. Myers, *Towards a derivation of holographic entanglement entropy*, JHEP **1105**, 036 (2011), arXiv:1102.0440.

[24] A. Connes, *Noncommutative Geometry*, Academic Press (1994).

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley (2006).

[26] G. E. Crooks, *Measuring thermodynamic length*, Phys. Rev. Lett. **99**, 100602 (2007).

[27] J. P. Crutchfield and K. Young, *Inferring statistical complexity*, Phys. Rev. Lett. **63**, 105 (1989).

[28] B. S. DeWitt, *Quantum Theory of Gravity. I. The Canonical Theory*, Phys. Rev. **160**, 1113 (1967).

[29] C. Doran and A. Lasenby, *Geometric Algebra for Physicists*, Cambridge University Press (2003).

[30] J. Eisert, M. Cramer, and M. B. Plenio, *Colloquium: Area laws for the entanglement entropy*, Rev. Mod. Phys. **82**, 277 (2010).

[31] M. Esposito, K. Lindenberg, and C. Van den Broeck, *Entropy production as correlation between system and reservoir*, New J. Phys. **12**, 013013 (2010).

[32] T. Faulkner, M. Guica, T. Hartman, R. C. Myers, and M. Van Raamsdonk, *Gravitation from Entanglement in Holographic CFTs*, JHEP **03**, 051 (2014).

[33] S. French, *The Structure of the World: Metaphysics and Representation*, Oxford University Press (2014).

[34] C. A. Fuchs, *QBism, the Perimeter of Quantum Bayesianism*, arXiv:1003.5209 (2014).

[35] C. A. Fuchs, N. D. Mermin, and R. Schack, *An Introduction to QBism with an Application to the Locality of Quantum Mechanics*, Am. J. Phys. **82**, 749 (2014).

[36] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, *Completely positive dynamical semigroups of N-level systems*, J. Math. Phys. **17**, 821 (1976).

[37] J. L. Garfield, *Empty Words: Buddhist Philosophy and Cross-Cultural Interpretation*, Oxford University Press (2002).

[38] R. F. Gombrich, *How Buddhism Began: The Conditioned Genesis of the Early Teachings*, Athlone Press (1996).

[39] A. M. Graybiel, *Habits, rituals, and the evaluative brain*, Annual Review of Neuroscience **31**, 359–387 (2008).

[40] W. P. Govern and P. R. ten Wolde, *Optimal resource allocation in cellular sensing systems*, Proc. Natl. Acad. Sci. USA **111**, 17486 (2014).

[41] M. Günaydin and F. Gürsey, *Quark structure and octonions*, J. Math. Phys. **14**, 1651 (1973).

[42] L. Ge and R. V. Kadison, On the structure of the lattice of von Neumann subalgebras, *Proc. Natl. Acad. Sci.* **93**, 3235 (1996).

[43] R. Haag, *Local Quantum Physics: Fields, Particles, Algebras*, Springer-Verlag (1996).

[44] U. Haagerup, *The standard form of von Neumann algebras*, Math. Scand. **37**, 271–283 (1975).

[45] P. Haggard, *Conscious intention and motor cognition*, Trends in Cognitive Sciences **9**(6), 290–295 (2005).

[46] M. B. Hastings, *An area law for one-dimensional quantum systems*, J. Stat. Mech. **2007**, P08024 (2007), arXiv:0705.2024.

[47] P. Harvey, *An Introduction to Buddhist Ethics: Foundations, Values and Issues*, Cambridge University Press (2000).

[48] P. Hayden and J. Preskill, *Black holes as mirrors: quantum information in random subsystems*, JHEP **09**, 120 (2007).

[49] P. Hayden, D. W. Leung, and A. Winter, *Aspects of generic entanglement*, Commun. Math. Phys. **265**, 95–117 (2006).

[50] D. Hestenes, *Space-Time Algebra*, Gordon and Breach (1966).

[51] S. Ito, *Stochastic thermodynamic interpretation of information geometry*, Phys. Rev. Lett. **121**, 030605 (2018).

[52] T. Jacobson, *Thermodynamics of Spacetime: The Einstein Equation of State*, Phys. Rev. Lett. **75**, 1260 (1995).

[53] T. Jacobson, *Entanglement equilibrium and the Einstein equation*, Phys. Rev. Lett. **116**, 201101 (2016), arXiv:1505.04753.

[54] D. L. Jafferis, A. Lewkowycz, J. Maldacena, and S. J. Suh, *Relative entropy equals bulk relative entropy*, JHEP **1606**, 004 (2016), arXiv:1512.06431.

[55] C. Jarzynski, *Equalities and inequalities: irreversibility and the second law at the nanoscale*, Annu. Rev. Condens. Matter Phys. **2**, 329 (2011).

[56] W. B. Johnson and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Contemp. Math. **26**, 189 (1984).

[57] C. Kiefer, *Quantum Gravity*, 3rd ed., Oxford University Press (2012).

[58] J. Ladyman and D. Ross, *Every Thing Must Go: Metaphysics Naturalized*, Oxford University Press (2007).

[59] R. Landauer, *Irreversibility and heat generation in the computing process*, IBM J. Res. Dev. **5**, 183 (1961).

[60] N. Lashkari, M. B. McDermott, and M. Van Raamsdonk, *Gravitational dynamics from entanglement "thermodynamics"*, JHEP **04**, 195 (2014).

[61] P. Mehta and D. J. Schwab, *Energetic costs of cellular computation*, Proc. Natl. Acad. Sci. USA **109**, 17978 (2012).

[62] F. Laudisa and C. Rovelli, *Relational Quantum Mechanics*, Stanford Encyclopedia of Philosophy (2021).

[63] J. Levine, *Materialism and qualia: The explanatory gap*, Pacific Philosophical Quarterly **64**, 354 (1983).

[64] D. A. Lidar and K. B. Whaley, *Decoherence-free subspaces and subsystems*, in *Irreversible Quantum Dynamics*, Springer, pp. 83–120 (2003).

[65] H. B. Lawson and M.-L. Michelsohn, *Spin Geometry*, Princeton University Press (1989).

[66] G. Lindblad, *On the generators of quantum dynamical semigroups*, Commun. Math. Phys. **48**, 119 (1976).

[67] D. Lusthaus, *Buddhist Phenomenology: A Philosophical Investigation of Yogācāra Buddhism and the Ch'eng Wei-shih lun*, Routledge (2002).

[68] J. M. Maldacena, *The Large N limit of superconformal field theories and supergravity*, Int. J. Theor. Phys. **38**, 1113 (1999).

[69] Nāgārjuna, *The Fundamental Wisdom of the Middle Way: Nāgārjuna's Mūlamadhyamakakārikā*, trans. J. L. Garfield, Oxford University Press (1995).

[70] S. Nakajima, *On quantum theory of transport phenomena*, Prog. Theor. Phys. **20**, 948 (1958).

[71] D. N. Page, *Average entropy of a subsystem*, Phys. Rev. Lett. **71**, 1291 (1993).

[72] D. N. Page and W. K. Wootters, *Evolution without evolution: Dynamics described by stationary observables*, Phys. Rev. D **27**, 2885 (1983).

[73] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, *Thermodynamics of information*, Nat. Phys. **11**, 131 (2015).

[74] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed., Cambridge University Press (2009).

[75] R. Penrose, *The Emperor's New Mind*, Oxford University Press (1989).

[76] D. Petz, *Monotone metrics on matrix spaces*, Linear Algebra Appl. **244**, 81 (1996).

[77] B. Poole, S. Ozair, A. Van Den Oord, A. Alemi, and G. Tucker, *On variational bounds of mutual information*, Proc. ICML, PMLR **97**, 5171–5180 (2019).

[78] H. Price, *Does Time-Symmetry Imply Retrocausality?*, Found. Phys. **42**, 724 (2012).

[79] C. R. Rao, *Information and the accuracy attainable in the estimation of statistical parameters*, Bull. Calcutta Math. Soc. **37**, 81 (1945).

[80] Á. Rivas, S. F. Huelga, and M. B. Plenio, *Quantum non-Markovianity: characterization, quantification and detection*, Rep. Prog. Phys. **77**, 094001 (2014).

[81] C. Rovelli, *Relational Quantum Mechanics*, Int. J. Theor. Phys. **35**, 1637 (1996).

[82] C. Rovelli, *Quantum Gravity*, Cambridge University Press (2004).

[83] S. Ryu and T. Takayanagi, *Holographic Derivation of Entanglement Entropy from AdS/CFT*, Phys. Rev. Lett. **96**, 181602 (2006).

[84] T. Sagawa and M. Ueda, *Fluctuation theorem with information exchange*, Phys. Rev. E **85**, 021104 (2012).

[85] T. Sagawa and M. Ueda, *Generalized Jarzynski Equality under Nonequilibrium Feedback Control*, Phys. Rev. Lett. **104**, 090602 (2010).

[86] T. Sagawa and M. Ueda, *Fluctuation Theorem with Information Exchange*, Phys. Rev. Lett. **109**, 180602 (2012).

[87] R. D. Schafer, *An Introduction to Nonassociative Algebras*, Academic Press (1966).

[88] E. Schrödinger, *What is Life?*, Cambridge University Press (1944).

[89] U. Seifert, *Stochastic thermodynamics, fluctuation theorems and molecular machines*, Rep. Prog. Phys. **75**, 126001 (2012).

[90] C. R. Shalizi and J. P. Crutchfield, *Computational mechanics: Pattern and prediction, structure and simplicity*, J. Stat. Phys. **104**, 817 (2001).

[91] C. E. Shannon, *Coding theorems for a discrete source with a fidelity criterion*, IRE Nat. Conv. Rec., Part 4, pp. 142–163 (1959).

[92] M. Siderits, *Buddhism as Philosophy: An Introduction*, Hackett Publishing (2007).

[93] H. A. Simon, *A behavioral model of rational choice*, Quarterly J. of Economics **69**, 99 (1955).

[94] C. A. Sims, *Implications of rational inattention*, J. Monetary Economics **50**, 665 (2003).

[95] D. A. Sivak and G. E. Crooks, *Thermodynamic metrics and optimal paths*, Phys. Rev. Lett. **108**, 190602 (2012).

[96] L. Smolin, *The case for background independence*, in *The Structural Foundations of Quantum Gravity*, eds. D. Rickles, S. French, J. Saatsi, Oxford University Press (2006).

[97] H. Spohn, *Entropy production for quantum dynamical semigroups*, J. Math. Phys. **19**, 1227 (1978).

[98] *Sutta Nipāta*, in *The Group of Discourses (Sutta Nipāta)*, trans. K. R. Norman, Pali Text Society (2001).

[99] B. Swingle, *Entanglement Renormalization and Holography*, Phys. Rev. D **86**, 065007 (2012).

[100] M. Takesaki, *Theory of Operator Algebras II*, Encyclopaedia of Mathematical Sciences **125**, Springer (2003).

[101] T. Thiemann, *Modern Canonical Quantum General Relativity*, Cambridge University Press (2007).

[102] N. Tishby, F. C. Pereira, and W. Bialek, *The information bottleneck method*, in Proc. 37th Allerton Conf. on Communication, Control, and Computing (1999); arXiv:physics/0004057 (2000).

[103] M. Van Raamsdonk, *Building up spacetime with quantum entanglement*, Gen. Relativ. Gravit. **42**, 2323 (2010).

[104] Vasubandhu, *Triṃśikā-vijñaptimātratā (Thirty Verses on Consciousness Only)*, trans. S. Anacker, in *Seven Works of Vasubandhu*, Motilal Banarsidass (1984).

[105] E. Verlinde, *On the Origin of Gravity and the Laws of Newton*, JHEP **04**, 029 (2011).

[106] G. Vidal, *Class of quantum many-body states that can be efficiently simulated*, Phys. Rev. Lett. **101**, 110501 (2008).

[107] H. Barnum, E. Knill, G. Ortiz, R. Somma, and L. Viola, *A Subsystem-Independent Generalization of Entanglement*, Phys. Rev. Lett. **92**, 107902 (2004).

[108] W. S. Waldron, *The Buddhist Unconscious: The Ālaya-vijñāna in the Context of Indian Buddhist Thought*, Routledge (2003).

[109] D. Wallace, *The Emergent Multiverse*, Oxford University Press (2012).

[110] D. Wallace, *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*, Oxford University Press (2012), Chapter 8.

[111] U. Weiss, *Quantum Dissipative Systems*, 4th ed., World Scientific (2012).

[112] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control*, Cambridge University Press (2009).

[113] H.-W. Wiesbrock, *Half-sided modular inclusions of von Neumann algebras*, Commun. Math. Phys. **157**, 83 (1993).

[114] E. Witten, *APS Medal for Exceptional Achievement in Research: Invited article on entanglement properties of quantum field theory*, Rev. Mod. Phys. **90**, 045003 (2018).

[115] M. M. Wolf, *Violation of the entropic area law for fermions*, Phys. Rev. Lett. **96**, 010404 (2006).

[116] J. Woodward, *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press (2003).

[117] H. H. Yin and B. J. Knowlton, *The role of the basal ganglia in habit formation*, Nature Reviews Neuroscience **7**, 464–476 (2006).

[118] J. Yngvason, *The role of type III factors in quantum field theory*, Rep. Math. Phys. **55**, 135 (2005), arXiv:math-ph/0411058.

[119] P. Zanardi, *Virtual Quantum Subsystems*, Phys. Rev. Lett. **87**, 077901 (2001).

[120] P. Zanardi, D. A. Lidar, and S. Lloyd, *Quantum Tensor Product Structures are Observable Induced*, Phys. Rev. Lett. **92**, 060402 (2004).

[121] W. H. Zurek, *Decoherence, einselection, and the quantum origins of the classical*, Rev. Mod. Phys. **75**, 715 (2003).

[122] W. H. Zurek, *Quantum Darwinism*, Nature Physics **5**, 181 (2009).

[123] R. Zwanzig, *Ensemble method in the theory of irreversibility*, J. Chem. Phys. **33**, 1338 (1960).

[124] H. L. van Trees, *Detection, Estimation, and Modulation Theory*, Part I, Wiley (1968).

[125] H. von Foerster, *Understanding Understanding: Essays on Cybernetics and Cognition*, Springer (2003).

[126] S. Liu, *Emergent Geometry from Coarse-Grained Observable Algebras: The Holographic Alaya-Field Framework*, Zenodo (2026), DOI: 10.5281/zenodo.18361706.

[127] S. Liu, *Emergent Geometry from Coarse-Grained Observable Algebras: The Holographic Alaya-Field Framework*, Zenodo (2026), DOI: 10.5281/zenodo.18361706.

[128] S. Liu, *Accessibility, Stability, and Emergent Geometry: Conceptual Clarifications on the Holographic Alaya-Field Framework*, Zenodo (2026), DOI: 10.5281/zenodo.18367060.

[129] S. Liu, *Causation, Agency, and Existence*, Zenodo (2026), DOI: 10.5281/zenodo.18374805.

[130] S. Liu, *Temporal Asymmetry as Accessibility Propagation*, Zenodo (2026), DOI: 10.5281/zenodo.18400425.

[131] S. Liu, *Structural Limits of Unification: Accessibility, Incompleteness, and the Necessity of a Final Cut*, Zenodo (2026), DOI: 10.5281/zenodo.18402907.

[132] S. Liu, *Structural Limits of Unification: Accessibility, Incompleteness, and the Necessity of a Final Cut*, Zenodo (2026), DOI: 10.5281/zenodo.18402907.

[133] R. Penrose, *Cycles of Time: An Extraordinary New View of the Universe*, Bodley Head (2010).

[134] S. Liu, *Emergent Geometry from Coarse-Grained Observable Algebras: The Holographic Alaya-Field Framework*, Zenodo (2026), DOI: 10.5281/zenodo.18361706.

[135] S. Liu, *Accessibility, Stability, and Emergent Geometry: Conceptual Clarifications on the Holographic Alaya-Field Framework*, Zenodo (2026), DOI: 10.5281/zenodo.18367060.

[136] S. Liu, *Causation, Agency, and Existence: Structural Constraints and Interpretive Bridges*, Zenodo (2026), DOI: 10.5281/zenodo.18374805.

[137] S. Liu, *Gravitational Phenomena as Emergent Properties of Observable Algebra Selection: A Structural Analysis*, Zenodo (2026), DOI: 10.5281/zenodo.18388881.

[138] S. Liu, *Measurement as Accessibility: A Structural Analysis of Observable Algebra Selection*, Zenodo (2026), DOI: 10.5281/zenodo.18400065.

[139] S. Liu, *Temporal Asymmetry as Accessibility Propagation: A Structural Analysis of Causal Direction*, Zenodo (2026), DOI: 10.5281/zenodo.18400425.

[140] S. Liu, *Algebraic Constraints on the Emergence of Lorentzian Metrics in Entropic Gravity Frameworks*, Zenodo (2026), DOI: 10.5281/zenodo.18525876.

[141] S. Liu, *Thermodynamic Stability Constraints on the Operator Algebra of Persistent Open Quantum Subsystems*, Zenodo (2026), DOI: 10.5281/zenodo.18525890.

[142] S. Liu, *The Realizability Bridge: Algebraic Closure in the Q-RAIF Framework*, Zenodo (2026), DOI: 10.5281/zenodo.18528934.

[143] S. Liu, *Non-Markovian Memory and the Thermodynamic Necessity of Temporal Accumulation*, Zenodo (2026), DOI: 10.5281/zenodo.18574342.

[144] S. Liu, *Spontaneous Symmetry Breaking of Reference Frames as a Computational Cost Minimization Strategy*, Zenodo (2026), DOI: 10.5281/zenodo.18579703.

[145] S. Liu, *Fisher Information Geometry and the Thermodynamic Cost of Self-Referential Calibration*, Zenodo (2026), DOI: 10.5281/zenodo.18591771.