Its From Bits - The Inertial Mass of Information

Robert C. Dennis[1,*]

[1] *Independent Researcher, Leander, Texas 78641, USA*

(Dated: November 22, 2025)

We demonstrate that inertial mass emerges as the statistical precision (the inverse covariance) of an agent's predictive model within a gauge-theoretic information geometry. By considering multi-variate Gaussians and a SO(N) gauge group we show that the second-order Taylor expansion of Kullback-Leibler divergence under belief updating yields a kinetic energy term $T = \frac{1}{2}\dot{\mu}^T \Sigma_p^{-1} \dot{\mu}$, in the variational free energy princple, where the Fisher information metric of the agent's predictive model serves as the mass matrix. This extends the variational free energy principle from a purely dissipative gradient dynamics to a fully Hamiltonian mechanics, with the Lagrangian structure $\mathcal{L} = T - V$. Crucially, mass and spacetime geometry are defined relative to each agent's prior with objective physics emerging only through inter-agent consensus driven by variational free energy minimization. We derive testable predictions including observer-dependent gravitational mass for quantum superpositions and information-theoretic interpretations of black hole entropy. The framework constitutes a concrete realization of Wheeler's participatory "it from bit" universe, wherein physical law emerges from the epistemic alignment of observing agents rather than existing a priori.

I. INTRODUCTION

The nature of inertial mass has long been a mystery to scientists and natural philosophers. It appears throughout physics as a primitive quantity: the m in Newton's $F = ma$, the source term in Einstein's field equations $G_{\mu\nu} = 8\pi T_{\mu\nu}$, and a parameter in the Standard Model Lagrangian. While the Higgs mechanism explains mass generation for elementary particles through spontaneous symmetry breaking, the deeper question of *why* mass exists

* cdenn016@gmail.com

and why matter resists change and curves spacetime remains unanswered within conventional frameworks.

Wheeler's vision of a participatory universe [1] and his aphorism "it from bit" suggest that physical quantities might emerge from information-theoretic primitives rather than being fundamental. This tantalizing program has inspired fruitful avenues of research [2, 3], yet concrete realizations connecting information to inertial mass have remained elusive.

Meanwhile, since Kant first described the mind's "forms of sensuous intuition," philosophers and neuroscientists have demonstrated empirically that the brain acts as a statistical model of reality, actively constructing representations of inaccessible underlying noumena [4, 5]. This presents a deep, yet often unspoken, tension between physics and neuroscience. On one hand, space, time, and matter are treated as primitives of our universe; on the other, they are actively constructed through perception. While many have pursued a "physics-first" perspective on cognition, few have taken the converse approach: a "cognition-first" perspective on physics itself.

In previous work [**?**], we showed that empirical results in machine learning such as attention mechanisms and transformer architectures emerge naturally from a cognition-first, gauge-theoretic framework built upon multi-agent variational inference. In companion work [**?**], we presented simulations of a Wheelerian participatory universe exhibiting multi-scale emergence, dynamical renormalization, and spontaneous symmetry breaking.

In the present work, we demonstrate that this cognitive-first framework yields a direct connection between informational quantities and observable matter. Concretely

1. We derive an informational Lagrangian $\mathcal{L} = \mathcal{T} - \mathcal{V}$ where the potential $\mathcal{V}$ is the generalized variational free energy and the kinetic term $\mathcal{T}$ emerges from the second-order expansion of the Kullback-Leibler divergence.

2. We identify inertial mass as statistical precision: $M = \Sigma_p^{-1}$, where Σ_p is the covariance of an agent's predictive model (prior). Mass emerges as the resistance to changing beliefs against a slowly-varying, quasi-static prior.

3. We show that standard dissipative dynamics (Friston's free energy principle) correspond to the *overdamped limit* of this richer Hamiltonian structure.

Finally, we present proof-of-principle simulations, discuss testable predictions, including observer-dependent gravitational mass for quantum superpositions, and explore general consequences of this framework for foundational physics.

A. Background: The Free Energy Principle

Friston's free energy principle (FEP) [5] provides a variational theory of self-organizing active systems, from cellular metabolism to neural computation. The variational free energy

$$\mathcal{F} = \mathrm{KL}(q\|p) - \mathbb{E}_q[\log p(o \mid c)] \tag{1}$$

is analogous to the Helmholtz free energy $F = U - TS$, where q represents an agent's beliefs about hidden states and p encodes its generative model (prior).

The associated dynamics take the form of gradient descent

$$\dot{\mu} = -\nabla_\mu \mathcal{F} \tag{2}$$

and is purely dissipative. All trajectories relax monotonically toward equilibrium with $\dot{\mathcal{F}} \leq 0$. There is no inertia, no oscillation, no wave propagation.

B. Extension to Multi-Agent Systems

In previous work [**?**], we extended this framework to multi-agent systems on principal bundles, deriving pairwise interaction terms from a normalized generative model. Full details of our gauge theoretic geometry may be found in [6].

Briefly, agents are modeled as smooth sections of an associated bundle $\mathcal{E}$ to a principal G bundle with statistical fibers $\mathcal{B}$ of K-dimensional multi-variate Gaussian (MVG) distributions where the structure group G acts on statistics $\mu_q(c)$ and $\Sigma_q(c)$ as

$$\rho(\Omega) \cdot (\mu, \Sigma) = (\Omega\mu, \Omega\Sigma\Omega^\top) \tag{3}$$

where $\Omega \in G$ and ρ is a K-dimensional representation of G.

In addition to the agents' statistics we define per-agent gauge frames $\phi_i(c) \in \mathfrak{g}$ at each point in the base manifold $\mathcal{C}$. In our studies we choose not to gauge fix any agents. This represents a geometric manifestation that any given agent may choose to fix their frames without consequence, however, the relational frames are what encode agent relationships. The gauge frames act as a local coordinate system with which agents embed their statistics.

In our associated bundle geometry we show that the variational free energy becomes

$$\boxed{\begin{aligned}\mathcal{V}[\{q_i\},\{p_i\},\{\phi_i\}] = &\sum_i \int_{\mathcal{C}} \chi_i(c)\, D_{\mathrm{KL}}(q_i(c)\|p_i(c))\, dc \\ &+ \sum_{ij} \int_{\mathcal{C}} \chi_{ij}(c)\, \beta_{ij}(c)\, D_{\mathrm{KL}}(q_i(c)\|\Omega_{ij}[q_j](c))\, dc \\ &+ \sum_{ij} \int_{\mathcal{C}} \chi_{ij}(c)\, \gamma_{ij}(c)\, D_{\mathrm{KL}}(p_i(c)\|\Omega_{ij}[p_j](c))\, dc \\ &- \sum_i \int_{\mathcal{C}} \chi_i(c)\, \mathbb{E}_{q_i}[\log p(o \mid c)]\, dc\end{aligned}} \tag{4}$$

Here $\chi_i(c)$ and $\chi_{ij}(c)$ are support functions encoding where agents exist and overlap, and $\Omega_{ij} \in G$ is the gauge transport operator rotating agent j's frame into agent i's perspective

$$\Omega_{ij} = e^{\phi_i} e^{-\phi_j} \tag{5}$$

where $\phi_i(c) \in \mathfrak{g}$ is agent i's local gauge frame in the Lie algebra $\mathfrak{g}$ of structure group G (we take $G = \mathrm{SO}(3)$ throughout, though the framework extends to $\mathrm{SO}(N)$ and $\mathrm{SU}(N)$ for MVG exponential families. Although we consider MVGs here for simplicity, other gauge groups may be applicable for more exotic mixture and exponential families).

As we have shown and validated elsewhere, the attention weights $\beta_{ij}(c)$ are softmax-normalized KL divergences

$$\beta_{ij}(c) = \frac{\exp\left[-\frac{1}{\kappa_\beta} D_{\mathrm{KL}}(q_i(c)\|\Omega_{ij}[q_j](c))\right]}{\sum_k \exp\left[-\frac{1}{\kappa_\beta} D_{\mathrm{KL}}(q_i(c)\|\Omega_{ik}[q_k](c))\right]} \tag{6}$$

where κ_β is the attention temperature. As shown in [**?**], this corresponds to transformer attention in the flat-bundle, isotropic, delta-function limit. The analogous weights $\gamma_{ij}(c)$ govern prior (model) alignment.

C. Observations as Environmental Agents

The observation term $\mathbb{E}_{q_i}[\log p_i(o \mid c)]$ represents agents reconciling their beliefs with sensory data according to their model p_i. We have shown previously that [**?**]

1. Observations manifestly break vacuum gauge symmetry, driving agents toward specialized configurations.

2. The observation term can be replaced by alignment with *environmental agents* $\{e_k\}$ as

$$\sum_i \int_{\mathcal{C}} \chi_i(c)\, \mathbb{E}_{q_i}[\log p_i(o_i \mid c)]\, dc \;\longleftrightarrow\; \sum_{i,k} \int_{\mathcal{C}} \chi_{ie_k}(c)\, \beta_{ie_k}(c)\, D_{\mathrm{KL}}(q_i(c) \| \Omega_{ie_k}[q_{e_k}](c))\, dc \tag{7}$$

In this view, observations *are* other agents with their own beliefs and priors. The participatory universe bootstraps itself into existence precisely as Wheeler envisioned [1].

D. Hierarchical Emergence

Under gradient descent of (4), natural hierarchical evolution unfolds as agents interact and reach consensus. When local clusters of agents align their beliefs (low pairwise KL divergence), higher-order *meta-agents* emerge as coarse-grained collective variables as a form of dynamical renormalization.

In our implementation, cross-scale couplings allow meta-agent beliefs to propagate downward as priors for constituent agents, completing an Ouroboros of self-consistent, self-referential dynamics, (i.e. a fully participatory universe of agents, beliefs, priors, and gauge frames) where the "top-observer" percolates beliefs down to lower levels.

II. MASS FROM STATISTICAL PRECISION

The framework presented in Section **??** defines a variational free energy $\mathcal{V}$ that serves as the potential for agent dynamics. We now show that promoting this to a full Lagrangian theory reveals a kinetic energy term hidden in the second-order structure of KL divergence and that this kinetic term identifies *inertial mass with statistical precision.*

A. The Second-Order KL Expansion

Consider an agent with beliefs $q = \mathcal{N}(\mu_q, \Sigma_q)$ and predictive model (prior) $p = \mathcal{N}(\mu_p, \Sigma_p)$. The KL divergence is

$$D_{\mathrm{KL}}(q\|p) = \frac{1}{2}\left[(\mu_q - \mu_p)^T\Sigma_p^{-1}(\mu_q - \mu_p) + \mathrm{tr}(\Sigma_p^{-1}\Sigma_q) - K + \ln\frac{|\Sigma_p|}{|\Sigma_q|}\right] \tag{8}$$

Under an infinitesimal belief update $\mu_q \to \mu_q + d\mu$, the change in KL divergence is

$$D_{\mathrm{KL}}(q + dq\|p) - D_{\mathrm{KL}}(q\|p) = \underbrace{(\mu_q - \mu_p)^T\Sigma_p^{-1}d\mu}_{\text{first order}} + \underbrace{\frac{1}{2}d\mu^T\Sigma_p^{-1}d\mu}_{\text{second order}} \tag{9}$$

The first-order term is the gradient, i.e. the ”force” driving belief updates. Standard treatments of the free energy principle retain only this term, yielding gradient descent dynamics.

The second-order term has traditionally been neglected. With $d\mu = \dot{\mu}\,dt$, it becomes

$$\frac{1}{2}d\mu^T\Sigma_p^{-1}d\mu = \frac{1}{2}\dot{\mu}^T\Sigma_p^{-1}\dot{\mu}\cdot dt^2 \tag{10}$$

This is analogous to kinetic energy.

B. The Fisher Metric as Mass Matrix

The natural metric on the statistical manifold of Gaussian distributions is the Fisher-Rao metric, manifestly positive. For the mean parameters, this metric is precisely

$$G_{\mu\mu} = \Sigma^{-1} \tag{11}$$

The second-order term in (9) can thus be written

$$\boxed{T = \frac{1}{2}\dot{\mu}^T G_{\mu\mu}\dot{\mu} = \frac{1}{2}\dot{\mu}^T\Sigma_p^{-1}\dot{\mu}} \tag{12}$$

This is kinetic energy with the Fisher information metric serving as the mass matrix

$$\boxed{M = \Sigma_p^{-1}} \tag{13}$$

If we take seriously Wheeler’s ”it from bit” participatory ontology then the identification $M = \Sigma_p^{-1}$ has profound physical consequences.

An agent with a precise prior (small Σ_p) has large inertial mass such that changing its beliefs requires effort proportional to Σ_p^{-1}. The more certain the agent is, the harder it is to move. Meanwhile, an agent with an uncertain prior (large Σ_p) has small inertial mass. Its beliefs are easily perturbed by external influences.

As a practical example we might say that "rocks are certain". A macroscopic object like a rock maintains an extremely precise self-model: $\Sigma_{\text{rock}} \approx \epsilon I$ with $\epsilon \sim \ell_P^2$ (Planck scale). This yields enormous inertial mass $M_{\text{rock}} \sim \ell_P^{-2}$.

Similarly, quantum superpositions are uncertain. A particle in superposition has larger effective position uncertainty than a localized particle. By (13), this means smaller inertial mass potentially allowing for a testable prediction developed in Section **??**.

C. The Intuition: Inertia as Epistemic Resistance

Why should certainty resist change? Consider the following thought experiment.

You believe a rock is at position x_0 with uncertainty $\sigma_{\text{rock}} \approx 0$. I claim the rock is actually at $x_1 \neq x_0$. To change your belief, you must traverse the statistical manifold from $\mu = x_0$ to $\mu = x_1$. The "distance" traveled, measured by the Fisher metric, is

$$d_F^2 = (x_1 - x_0)^T \Sigma_{\text{rock}}^{-1} (x_1 - x_0) \to \infty \quad \text{as } \Sigma_{\text{rock}} \to 0 \tag{14}$$

The more certain your belief, the greater the information-geometric distance required to change it. This is the epistemic origin of inertia.

D. Covariance and Gauge Kinetic Terms

For completeness, the full kinetic energy includes contributions from all dynamical fields. The covariance Σ_q evolves on the manifold of symmetric positive-definite (SPD) matrices, with Fisher metric

$$G_{\Sigma\Sigma}[\dot{\Sigma}, \dot{\Sigma}] = \frac{1}{2} \text{tr} \left(\Sigma^{-1} \dot{\Sigma} \Sigma^{-1} \dot{\Sigma} \right) \tag{15}$$

The gauge frame $\phi \in \mathfrak{g}$ evolves on the Lie algebra, with the Killing form providing a natural metric

$$\langle \dot{\phi}, \dot{\phi} \rangle_{\mathfrak{g}} = -\text{tr}(\dot{\phi}^2) \tag{16}$$

The full kinetic energy is thus:

$$\mathcal{T} = \sum_i \int_{\mathcal{C}} \chi_i(c) \left[\frac{1}{2} \dot{\mu}_{q,i}^T \Sigma_{p,i}^{-1} \dot{\mu}_{q,i} + \frac{1}{4} \mathrm{tr}(\Sigma_{q,i}^{-1} \dot{\Sigma}_{q,i})^2 + \frac{1}{2} |\dot{\phi}_i|^2 \right] dc \tag{17}$$

For the physical interpretation of inertial mass, we contend that the dominant contribution comes from the mean sector (in our toy model), since, in our simulations, covariance and gauge dynamics typically occur on slower timescales.

In our framework, there is no external time parameter t. Instead, time emerges from the information-processing dynamics themselves. We operationally define time such that each agent's clock advances by one "tick" when it updates beliefs by one bit of information: $\Delta\tau_i = \Delta I_i/(1 \text{ bit})$ where $\Delta I_i = \mathrm{KL}(q_i^{\text{new}} \| q_i^{\text{old}})$. This realizes Wheeler's "it from bit" at the level of temporal flow such that time is the counting parameter for information updates.

Importantly belief dynamics allows us a natural definition for the arrow of time as the direction of an agent's belief trajectory in the fiber. In our view we anticipate, but have not shown, that perpendicular directions to the belief trajectory may pullback on the base manifold as spatial directions. As different agents utilize different trajectories they experience different proper times. This suggests a route towards an information-theoretic time dilation analogous to relativistic time dilation but arising from information geometry rather than spacetime curvature.

III. THE FULL DYNAMICAL THEORY

Given the kinetic energy $\mathcal{T}$ from Section II and the potential $\mathcal{V}$, we now construct the complete Lagrangian and Hamiltonian formulations for epistemic dynamics.

A. The Lagrangian

The Lagrangian is

$$\boxed{\mathcal{L} = \mathcal{T} - \mathcal{V}} \tag{18}$$

Suppressing covariance and gauge kinetic terms for clarity, the mean-field sector Lagrangian is

$$
\begin{aligned}
\mathcal{L} = \sum_i \int_{\mathcal{C}} \chi_i(c)\, \frac{1}{2} \dot{\mu}_{q,i}^T \Sigma_{p,i}^{-1} \dot{\mu}_{q,i}\, dc \\
- \sum_i \int_{\mathcal{C}} \chi_i(c)\, D_{\mathrm{KL}}(q_i \| p_i)\, dc \\
- \sum_{ij} \int_{\mathcal{C}} \chi_{ij}(c)\, \beta_{ij}(c)\, D_{\mathrm{KL}}(q_i \| \Omega_{ij}[q_j])\, dc \\
- \sum_{ij} \int_{\mathcal{C}} \chi_{ij}(c)\, \gamma_{ij}(c)\, D_{\mathrm{KL}}(p_i \| \Omega_{ij}[p_j])\, dc \\
+ \sum_i \int_{\mathcal{C}} \chi_i(c)\, \mathbb{E}_{q_i}[\log p(o|c)]\, dc
\end{aligned}
\tag{19}
$$

The action is

$$
S = \int_{\tau_0}^{\tau_1} \mathcal{L}\, d\tau \tag{20}
$$

where τ represents a proper time.

B. Conjugate Momenta

The conjugate momentum to $\mu_{q,i}(c)$ is:

$$
\pi_{\mu,i}(c) = \frac{\delta \mathcal{L}}{\delta \dot{\mu}_{q,i}(c)} = \Sigma_{p,i}^{-1}(c)\, \dot{\mu}_{q,i}(c) \tag{21}
$$

Inverting:

$$
\dot{\mu}_{q,i}(c) = \Sigma_{p,i}(c)\, \pi_{\mu,i}(c) \tag{22}
$$

C. The Hamiltonian

The Hamiltonian is the Legendre transform of the Lagrangian

$$
\mathcal{H} = \sum_i \int_{\mathcal{C}} \pi_{\mu,i} \cdot \dot{\mu}_{q,i}\, dc - \mathcal{L} = \mathcal{T} + \mathcal{V} \tag{23}
$$

In terms of momenta we have

$$
\boxed{\mathcal{H} = \sum_i \int_{\mathcal{C}} \frac{1}{2} \pi_{\mu,i}^T(c)\, \Sigma_{p,i}(c)\, \pi_{\mu,i}(c)\, dc + \mathcal{V}[\{q_i\}, \{p_i\}, \{\phi_i\}]} \tag{24}
$$

where integration over agent support is understood.

The Hamiltonian $\mathcal{H}$ is conserved under the flow.

D. Hamilton's Equations

Given these results we have

$$\dot{\mu}_{q,i}(c) = \frac{\delta \mathcal{H}}{\delta \pi_{\mu,i}(c)} = \Sigma_{p,i}(c)\, \pi_{\mu,i}(c) \tag{25}$$

$$\dot{\pi}_{\mu,i}(c) = -\frac{\delta \mathcal{H}}{\delta \mu_{q,i}(c)} = -\frac{\delta \mathcal{V}}{\delta \mu_{q,i}(c)} \tag{26}$$

Combining (25) and (26) yields the second-order equation

$$\boxed{\Sigma_{p,i}^{-1}(c)\, \ddot{\mu}_{q,i}(c) = -\frac{\delta \mathcal{V}}{\delta \mu_{q,i}(c)}} \tag{27}$$

This is an informational Newton's second law on the statistical manifold

$$M_i(c)\, \ddot{\mu}_i(c) = F_i(c) \tag{28}$$

with

- Mass matrix: $M_i(c) = \Sigma_{p,i}^{-1}(c)$
- Force: $F_i(c) = -\frac{\delta \mathcal{V}}{\delta \mu_{q,i}(c)}$

We provide the complete gradients in the appendix.

Finally, we notice, in passing, that the full variational energy has the structure of the grand thermodynamic potential $\Upsilon = E - TS - \mu N$.

E. The Overdamped Limit

Adding dissipation (friction) γ_i to the dynamics:

$$\chi_i \Sigma_{p,i}^{-1} \ddot{\mu}_{q,i} + \gamma_i \dot{\mu}_{q,i} = -\frac{\delta \mathcal{V}}{\delta \mu_{q,i}} \tag{29}$$

In the **overdamped limit** $\gamma_i \to \infty$ (or equivalently $M_i \to 0$), the inertial term vanishes:

$$\gamma_i \dot{\mu}_{q,i} = -\frac{\delta \mathcal{V}}{\delta \mu_{q,i}} \tag{30}$$

This is standard gradient descent as popularized by Friston's free energy principle [5].

> The free energy principle is the overdamped (high-friction, zero-inertia) limit of a richer Hamiltonian theory. Neural systems may operate in this regime however, "it from bit" physics requires the full second-order structure.

F. Conservation Laws and Symmetries

By Noether's theorem, continuous symmetries of $\mathcal{L}$ yield conserved quantities:

a. Time translation invariance ($\mathcal{L}$ independent of t) implies conservation of energy:

$$\frac{d\mathcal{H}}{dt} = 0 \tag{31}$$

b. Gauge invariance ($\mathcal{L}$ invariant under $\phi_i \to \phi_i + \xi$) implies conservation of "gauge charge".

c. Translation invariance on $\mathcal{C}$ (if present) implies conservation of total informational momentum:

$$P_{\text{total}} = \sum_i \int_{\mathcal{C}} \pi_{\mu,i}(c)\, dc = \text{const} \tag{32}$$

G. Mass Relative to Prior

The inertial "self" mass of agent i is

$$M_i = \Sigma_{p_i}^{-1} \tag{33}$$

This is defined in terms of agent i's own prior p_i. However, the mass agent j assigns to agent i is

$$M_i^{(j)} = \Sigma_{p_j}^{-1}(\text{regarding } i\text{'s state}) \tag{34}$$

In general:

$$M_i^{(j)} \neq M_i^{(k)} \neq M_i^{(i)} \tag{35}$$

The perspectival nature extends quite beyond mass. The pullback metric also depends on each agent's priors

$$g_{\mu\nu}^{(i)}(x) \neq g_{\mu\nu}^{(j)}(x) \tag{36}$$

However, note that these assignments depend on agent priors which themselves depend on the generative model the agents are utilizing to describe their observations. We generally have for unaligned agents that

- Each agent experiences its own perceptual pullback geometry

- Causal structure is observer-dependent
- Most importantly "The laws of physics" differ between agents

However, this is not a failure of the theory but its central feature. There is no "view from nowhere". In the case of human agents we have evolved to largely share a consensus generative model of reality utilizing our own unique/similar gauge-frames. Deviations from these priors are well-documented (trauma, psychedelics, meditative states, etc) as are the perceptual "violations" of physical law (objects materializes, dynamic spatial and temporal geometries, etc). Traditionally, these "observations" have largely been dismissed as pathological and deviant. However, our framework naturally account for them.

H. Consensus as Objectivity

Consider that our free energy $\mathcal{V}$ includes a prior alignment term

$$\sum_{ij} \gamma_{ij}(c)\, D_{\mathrm{KL}}(p_i(c) \| \Omega_{ij}[p_j](c)) \tag{37}$$

Minimizing these terms drives priors toward agreement. At equilibrium

$$p_i \approx \Omega_{ij} p_j \quad \forall i, j \tag{38}$$

When (38) holds, the gauge-invariance of trace ensures that

$$\mathrm{tr}(\Sigma_{p_i}^{-1}) = \mathrm{tr}(\Omega_{ij}\Sigma_{p_j}^{-1}\Omega_{ij}^T) = \mathrm{tr}(\Sigma_{p_j}^{-1}) \tag{39}$$

Therefore we find that at consensus, all agents agree on masses and manifestly agree that their generative models of reality are gauge invariant.

Similarly, the emergent pullback metrics align

$$g_{\mu\nu}^{(i)}(x) = g_{\mu\nu}^{(j)}(x) = g_{\mu\nu}(x) \quad \text{(shared geometry)} \tag{40}$$

"Objective physics", our universal laws, shared spacetime, agreed-upon masses is manifestly, in "it from bit" the *intersubjective gauge-invariant consensus* that emerges when agents align their priors through free energy minimization.

I. Macroscopic Objects Impose Consensus

We have shown a mechanism by which agents can evolve to agree on generative models of reality. We may wonder, "why do everyday physics appear objective"? In the view of our framework we find that macroscopic objects enforce consensus through precision dominance.

A rock maintains beliefs about its own state with enormous precision

$$\Sigma_{\text{rock}}^{(\text{self})} \approx \epsilon I, \quad \epsilon \to 0 \tag{41}$$

When agent i interacts with the rock, their coupling term is

$$\beta_{i,\text{rock}}\, D_{\text{KL}}(q_i \| \Omega_{i,\text{rock}}[q_{\text{rock}}]) \tag{42}$$

Since $\Sigma_{\text{rock}}^{-1} \to \infty$, this term dominates i's free energy. Agent i's beliefs are forced to align with the rock's self-model.

The apparent objectivity of macroscopic physics arises because high-precision objects act as epistemic anchors that drag all observers into consensus.

J. Quantum Systems: Pre-Consensus Physics

We argue that quantum systems in superposition have not yet reached consensus. No agent has formed a sharp prior; different observers maintain different beliefs.

Consider a particle in superposition

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{43}$$

- Observer A with prior peaked on $|0\rangle$: assigns mass M_A
- Observer B with prior peaked on $|1\rangle$: assigns mass M_B
- Observer C with uniform prior: assigns mass M_C

The particle has no definite mass until measurement. Measurement is the process by which observers form sharp priors and reach consensus. Wavefunction collapse then, in this view, is consensus formation: the transition from perspectival to shared physics.

K. The Transition: From Perspectival to Objective

Regime	Prior State	Physics
Pre-measurement	$p_i \neq p_j$, broad	Perspectival, quantum
Partial consensus	$p_i \approx p_j$, narrowing	Decoherence, classicalization
Full consensus	$p_i = \Omega_{ij} p_j$, sharp	"Objective," classical

TABLE I. The emergence of objective physics through consensus.

IV. PERSPECTIVAL PHYSICS AND CONSENSUS

If we allow ourselves to take Wheeler's "it from bit" participatory view seriously [1] and couple it with the neuro-scientific view that our perceptual realities are constructed as an "evolutionarily controlled hallucination" [7], then we are resigned to abandon some of the most cherished bedrocks of our universe and its conceptualization.

A fundamental feature of our framework is that physical quantities are subjective: defined relative to each agent's prior. Objective physics emerges only through inter-agent consensus and belief/prior evolution.

This is however, not fatal to predictable and reliable physical processes. In fact, this novel view may allow novel interpretations and explanations of some of the most nagging issues in physics; such as the observer paradox, relativity, the measurement problem, and more.

The perspectival structure resonates with several philosophical traditions such as Kantian idealism where space and time are not "things in themselves" but rather "forms of sensuous intuitions" which the mind organizes and experiences. In Kant's view space-time is not fundamental but constructed from agent priors.

Similarly, Rovelli [8] argues that quantum states are relative to observers. We greatly extend this: not just states but masses, geometry, and physical law are all relative until consensus, the sharing of models of reality, is reached.

Fuchs et al. [9] interpret probabilities as subjective degrees of belief. Our framework shares this spirit while providing an informational mechanism (variational free energy minimization) for how intersubjective agreement emerges.

Meanwhile, Wheeler envisioned observers ”participating” in bringing reality into being. Our framework makes this precise: the pullback metric is constituted by the network of agent priors and their alignment dynamics.

We describe several possible routes towards falisifying or validating our framework in the next section.

A. Mass of Quantum Superpositions

The mass-precision correspondence yields a striking prediction for quantum systems: a particle in spatial superposition should have less inertial mass than the same particle in a localized state.

Consider a particle localized to position uncertainty σ_0. According to the framework, its inertial mass is

$$M_0 = \kappa \cdot \frac{1}{\sigma_0^2} \tag{44}$$

where $\kappa = \hbar \ell_P / c^2$ provides the dimensional coupling between information-geometric and gravitational quantities in natural units (if we tentatively assume inertial-gravitational mass equivalence). Now if we prepare the same particle in a spatial superposition with larger position uncertainty $\sigma_+ > \sigma_0$. The mass-precision identification implies

$$M_+ = \kappa \cdot \frac{1}{\sigma_+^2} < M_0 \tag{45}$$

The superposed particle is lighter than its localized counterpart. This is not a statement about rest mass in the conventional sense but about inertial response: the de-localized particle, having lower precision about its position, offers less resistance to acceleration. The mass deficit is

$$\Delta M = M_0 - M_+ = \frac{\hbar \ell_P}{c^2} \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_+^2} \right) \tag{46}$$

For a meso-scopic superposition with $\sigma_+ \sim 1\,\mu\text{m}$ and $\sigma_0 \sim 1\,\text{pm}$ (typical atomic localization), this yields $\Delta M \sim 10^{-34}\,\text{kg}$, corresponding to roughly 10^4 Planck masses. While extraordinarily small, this regime may be accessible to next-generation matter-wave interferometry experiments. Marshall, Simon, Penrose, and Bouwmeester have proposed placing a mirror of $\sim 10^{14}$ atoms into spatial superposition via single-photon interactions [10], and recent work has outlined a roadmap for matter-wave interference with masses up to 100 MDa [11]. Space-based experiments using optically trapped nanospheres, currently under development by Aspelmeyer and collaborators, aim to test precisely the regime where gravitational effects on superposition become significant [12].

This prediction is sharp: measuring the gravitational field (if equivalence holds) of a particle in superposition should reveal a mass deficit relative to the localized case, with the deficit scaling as the inverse square of the position uncertainties.

This connects to ongoing debates about gravitational decoherence and the Penrose-Diósi hypothesis. In those frameworks, gravity causes superposition collapse; here, superposition reduces effective gravitational mass. The two perspectives may be complementary faces of a deeper unity between quantum coherence and gravitational physics.

B. Black Hole Entropy as Precision Bound

The Bekenstein-Hawking entropy of a black hole,

$$S_{BH} = \frac{A}{4\ell_P^2} \tag{47}$$

counts the maximum information content of a region bounded by area A. In the mass-precision framework, this acquires a new interpretation: the entropy bound is a *precision bound.*

The total precision able to localize to a (percieved) spatial region cannot exceed what its boundary area permits. Translating the entropy into precision units,

$$\mathrm{tr}(\Sigma^{-1})_{\max} \sim \frac{A}{\ell_P^4} \tag{48}$$

where the ℓ_P^4 factor arises from dimensional analysis (precision has units of inverse length squared; integrating over a 2-surface of area A and converting to Planck units yields this scaling).

This reframes the holographic principle information-geometrically. The bound $S \leq A/4\ell_P^2$ becomes: no region can contain more certainty than its boundary area allows. A black hole saturates this bound. It is maximally precise given its surface area. Any attempt to localize more information within the region would require more area, causing the horizon to grow.

The second law of black hole thermodynamics ($dA \geq 0$) becomes a statement about precision conservation: total certainty cannot decrease, and concentrating certainty requires spatial resources.

The connection to the mass-precision correspondence is immediate. Since $M = \kappa \cdot \mathrm{tr}(\Sigma^{-1})$, a maximum precision bound implies a maximum mass bound for a given area. This is precisely the content of the hoop conjecture: a region of circumference C cannot contain mass exceeding $M \sim C/G$. The geometric bound on precision translates directly to the gravitational bound on mass.

C. Mass per Bit

The framework allows us to define a fundamental quantity: the mass per bit of information stored in a gravitating system. For a black hole of mass M containing $N = S_{BH}/\ln 2$ bits, the mass per bit is

$$m_{\mathrm{bit}} = \frac{M}{N} = \frac{M \ln 2}{S_{BH}} = \frac{\ln 2 \cdot M_P^2}{4\pi M} \tag{49}$$

where we have used $S_{BH} = 4\pi M^2/M_P^2$ for a Schwarzschild black hole.

This reveals a remarkable scaling: mass per bit is *inversely proportional* to total mass. A Planck-mass black hole has $m_{\mathrm{bit}} \approx 0.055\, M_P$, meaning each bit contributes significantly to the total mass. A solar-mass black hole has $m_{\mathrm{bit}} \approx 10^{-39}\, M_P$, with each bit contributing negligibly.

The interpretation is that larger black holes have lower precision per bit. Their information is spread out over more degrees of freedom. A small black hole packs its bits tightly, with each bit highly localized and therefore massive. A large black hole dilutes its information across an enormous entropy, with each individual bit contributing little to the total precision and hence little to the total mass.

This suggests a picture of black hole evaporation as precision concentration. As a black hole radiates and shrinks, m_{bit} increases. The final stages of evaporation involve highly massive bits (maximally localized information) which may explain the predicted burst of high-energy radiation at the endpoint. The information that emerges in Hawking radiation is initially dilute and low-precision; as evaporation proceeds, the remaining information becomes increasingly concentrated and massive.

D. Gravitational Waves as Precision Radiation

When compact objects merge, a fraction of the total mass-energy is radiated as gravitational waves. In the mass-precision framework, this radiation carries away precision.

For a merger releasing gravitational wave energy ΔE_{GW}, the corresponding precision loss is

$$\Delta[\mathrm{tr}(\Sigma^{-1})] = \frac{\Delta E_{GW}}{\kappa c^2} = \frac{\Delta E_{GW} \cdot c^2}{\hbar \ell_P} \tag{50}$$

A binary in-spiral can be understood as a process of precision redistribution. Initially, the two objects have separate, localized precision contributions. As they spiral inward, gravitational wave emission carries precision to infinity. The final merged object has less total precision than the sum of its progenitors, with the deficit encoded in the gravitational wave signal.

This yields a quantitative prediction: the information content of detected gravitational waves should satisfy

$$\Delta I = \frac{\Delta E_{GW}}{\hbar c^2/\ell_P} = \frac{\Delta E_{GW} \cdot \ell_P}{\hbar c^2} \tag{51}$$

in natural units. For the GW150914 event, which radiated approximately $3M_\odot c^2$ of energy, this corresponds to roughly 10^{77} bits of information—consistent with the entropy change expected from the area theorem applied to the merging horizons.

The gravitational wave strain $h(t)$ encodes this precision flow. Regions of high strain correspond to moments of rapid precision transfer; the ringdown phase represents the final object settling into its equilibrium precision distribution. Future gravitational wave observations, combined with precision measurements of the progenitor and remnant masses, could test whether the radiated energy precisely accounts for the precision deficit predicted by the framework.

E. Implications

If the mass-precision correspondence is empirically validated, several implications follow.

The framework provides a unified account of inertia across cognitive, informational, neural, and potentially physical domains. The same mathematical structure—Fisher information as mass—applies at all scales, suggesting deep connections between information processing and our universe. The correspondence also offers new experimental handles on cognitive and neural dynamics, since precision can be measured and manipulated independently of inertial response, providing a direct test of the identification.

Finally, the framework supports the broader program of deriving physics from information theory. If inertial mass is fundamentally precision, then mechanics is a special case of inference, and the laws of physics describe optimal information processing under the constraints of our shared evolutionary priors.

V. DISCUSSION

A. Why Was the Kinetic Term Missed?

The kinetic term $\frac{1}{2}d\mu^T\Sigma^{-1}d\mu$ is second-order in the belief update. Standard treatments of active inference and variational free energy minimization employ first-order gradient descent, implicitly taking the overdamped limit in which inertial terms are negligible. In the mechanical analogy, this corresponds to

$$m\ddot{q} + \gamma\dot{q} = -\nabla V \quad \xrightarrow{m/\gamma\to 0} \quad \gamma\dot{q} = -\nabla V \tag{52}$$

where the mass-to-damping ratio vanishes and acceleration becomes instantaneously yoked to the force.

Neural systems may well operate in this regime, where synaptic time constants dominate over any inertial effects. But fundamental physics requires the full second-order structure. The oversight is natural: when modeling brain dynamics, one reasonably neglects terms that are small on biological timescales. The present work recovers these terms by treating the variational principle as fundamental rather than approximate, revealing structure that was always present but previously discarded.

B. The Equivalence Principle

The framework predicts that inertial and gravitational mass share a common origin

$$M_{\text{inertial}} = M_{\text{gravitational}} = \kappa \cdot \text{tr}(\Sigma_p^{-1}) \tag{53}$$

The same precision that resists changes in belief/priors (inertia) also sources the metric that other agents experience as gravity. The equivalence principle, one of the deepest empirical facts in physics, thus emerges from the single origin of both effects in prior precision. This is not an assumption but a consequence: any framework in which mass derives from a single informational quantity will automatically satisfy equivalence.

C. Realization of Wheeler's Program

John Archibald Wheeler proposed that physics might ultimately rest on information-theoretic foundations, captured in slogans such as "it from bit" and "law without law" [1]. The present framework provides a concrete mathematical framework with falsifiable predictions for this vision.

In tandem Wheeler's "participatory universe" holds that observers constitute reality rather than merely measuring a pre-existing world. In our framework, spacetime structure is not ontologically prior to agents but emerges from their consensus. The coupling terms in the variational functional drive agents toward aligned beliefs; the geometry they collectively experience is this alignment made manifest.

Finally, "law without law" suggests that physical regularities emerge from deeper principles rather than being axiomatic. Here, the laws of mechanics emerge from variational free energy minimization on a principal bundle. No dynamical laws are assumed; they arise from the geometry of inference. Furthermore, gauge invariance of physical law is manifestly a consequence of agent-agent alignment and consensus. Any agents which agree on reality must be gauge equivalent. Our theory produces this as a consequence, rather than an assumption.

VI. RELATION TO OTHER APPROACHES

A. Entropic Gravity

Verlinde proposed that gravity is an entropic force arising from information changes at holographic screens [2]. While sharing the "gravity from information" spirit, our approach differs in its fundamental quantity. Verlinde takes entropy gradients at boundaries as primary, with temperature playing a central role; here we take precision, the inverse of entropy, as primary, with certainty rather than thermal fluctuation as the basic concept.

The two perspectives may be related through a kind of duality. Verlinde's picture emphasizes what is unknown (entropy, disorder), while ours emphasizes what is known (precision, certainty). A region of high certainty corresponds to low entropy and, in our framework, high mass. Whether these can be unified into a single coherent picture, perhaps with precision and entropy as complementary descriptions of the same underlying reality, remains an open question.

B. Thermodynamic Gravity

In 1995 Jacobson derived Einstein's equations from thermodynamic relations applied to local Rindler horizons [3], treating spacetime dynamics as an equation of state rather than a fundamental law. Our approach is complementary in spirit but different in execution. Jacobson assumes local equilibrium thermodynamics and derives gravitational dynamics from $\delta Q = T\,dS$. We derive dynamics from agent inference without assuming equilibrium, allowing treatment of far-from-equilibrium situations where Jacobson's thermodynamic approach may not apply. Indeed, as we discuss elsewhere, meta-agent hierarchies naturally emerge and bootstrap in an "Ourobouros" fashion via agent-agent attention patterns from excess variational free energy. As a first order gradient descent dynamic renormalization occurs with a proliferation of complexity.[**?**]

Both approaches connect the Einstein tensor $G_{\mu\nu}$ to entropic or informational quantities, suggesting that the geometric description of gravity may be a coarse-grained or emergent account of underlying information-theoretic dynamics. The precise relationship between Jacobson's horizon thermodynamics and our agent-based inference remains to be clarified.

C. Gravitational Decoherence

Penrose [13] and Diósi [14] proposed that gravity causes wavefunction collapse, with the decoherence rate scaling with the mass of the superposed object. In their picture, superpositions of massive objects are unstable because the gravitational self-energy difference between branches drives objective reduction.

Our framework suggests an inversion of this logic. Rather than mass causing decoherence, we propose that decoherence causes definite mass to emerge. A system in superposition has indefinite precision and therefore indefinite mass. When observation induces consensus, precision becomes definite, and a specific mass value emerges. This is not gravitational decoherence but a form of epistemic coalescence. It is the crystallization of mass from the resolution of uncertainty.

The experimental predictions differ subtly. Penrose-Diósi predicts that massive superpositions decay faster than less massive ones due to gravitational self-interaction. Our framework predicts that superpositions have reduced effective mass (lower precision), which should be detectable as a mass deficit in precision interferometry. Both predictions are testable in the experimental regime now becoming accessible [10–12].

D. Relational Quantum Mechanics and QBism

Our perspectival structure resonates with Rovelli's relational quantum mechanics [8] and the QBist interpretation developed by Fuchs, Mermin, and Schack [9]. All three approaches hold that physical properties are relative to observers rather than absolute features of a mind-independent world.

Relational quantum mechanics asserts that quantum states are relational and defined only with respect to a reference system. However, it takes the dynamical laws as given structure. QBism interprets quantum probabilities as subjective degrees of belief, constraints on rational agents, rather than objective features of reality. Our framework shares the perspectival commitments of both but goes further in two respects.

First, we derive the dynamics rather than assuming them. Second, we provide a mechanism for intersubjective agreement. The coupling terms in the free energy functional drive agents toward agreement; the objective world is the fixed (or rather, current) point of this dynamics rather than its presupposition. Where relational QM and QBism describe the perspectival character of physics, our framework explains how perspectives coalesce into a shared reality.

VII. CONCLUSION

We have shown that inertial mass emerges in the form of statistical precision via the inverse covariance of an agent's predictive model, within a gauge-theoretic informational bundle geometry. The second-order KL expansion provides kinetic energy with the Fisher metric as mass matrix. The Lagrangian structure $\mathcal{L} = T - V$ automatically yields Lorentzian signature when covariantized to field theory.

Most significantly, mass and geometry are perspectival, defined relative to each agent's prior, with objective physics emerging only through inter-agent consensus. Different groups of agents may assign completely different "physics" to their shared reality. While arbitrary foolishness might emerge at first glance we find deep self-consistency throughout the framework. This provides a novel and concrete realization of Wheeler's participatory universe wherein physical law emerges from the epistemic alignment of observing agents.

Our framework yields testable predictions about quantum superposition masses and information-theoretic black hole properties, opening new connections between foundations of physics and information geometry. While many questions remain, we have a princpled mathematical framework with which to explore that may find utility not only in physics but diverse fields where dynamical informational systems interact and evolve.

ACKNOWLEDGMENTS

---

[1] J. Wheeler, *Complexity, Entropy, and the Physics of Information* (Addison-Wesley, 1990) pp. 3–28.

[2] E. Verlinde, On the origin of gravity and the laws of Newton, Journal of High Energy Physics **2011**, 29 (2011).

[3] T. Jacobson, Thermodynamics of spacetime: The Einstein equation of state, Physical Review Letters **75**, 1260 (1995).

[4] H. von Helmholtz, *Handbuch der Physiologischen Optik* (Leopold Voss, Leipzig, 1867).

[5] K. Friston, The free-energy principle: a unified brain theory?, Nature Reviews Neuroscience **11**, 127 (2010).

[6] R. Dennis, Attention, transformers, and backpropagation are degenerate limits of the variational free energy principle, Journal of Machine Learning Research (2025), submitted.

[7] A. Seth, *Being You: A New Science of Consciousness* (Faber & Faber, London, 2021).

[8] C. Rovelli, Relational quantum mechanics, International Journal of Theoretical Physics **35**, 1637 (1996).

[9] C. A. Fuchs, N. D. Mermin, and R. Schack, An introduction to QBism with an application to the locality of quantum mechanics, American Journal of Physics **82**, 749 (2014).

[10] W. Marshall, C. Simon, R. Penrose, and D. Bouwmeester, Towards quantum superpositions of a mirror, Physical Review Letters **91**, 130401 (2003).

[11] F. Kiałka, Y. Y. Fein, S. Pedalino, S. Gerlich, and M. Arndt, A roadmap for universal high-mass matter-wave interferometry, AVS Quantum Science **4**, 020502 (2022).

[12] C. Pfister, M. Kaniber, N. Piro, D. Bouwmeester, M. Aspelmeyer, and Č. Brukner, A universal test for gravitational decoherence, Nature Communications **7**, 13022 (2016).

[13] R. Penrose, On the gravitization of quantum mechanics 1: Quantum state reduction, Foundations of Physics **44**, 557 (2014).

[14] L. Diósi, Models for universal reduction of macroscopic quantum fluctuations, Physical Review A **40**, 1165 (1989).

Appendix A: Detailed KL Expansion for Multivariate Gaussians

For $q = \mathcal{N}(\mu_q, \Sigma_q)$ and $p = \mathcal{N}(\mu_p, \Sigma_p)$, the KL divergence is:

$$D_{\mathrm{KL}}(q\|p) = \frac{1}{2}\left[\mathrm{tr}\big(\Sigma_p^{-1}\Sigma_q\big) + \Delta\mu^T\Sigma_p^{-1}\Delta\mu - K + \ln\frac{|\Sigma_p|}{|\Sigma_q|}\right] \tag{A1}$$

where $\Delta\mu = \mu_q - \mu_p$ and $K = \dim(\mu)$.

Under the shift $\mu_q \to \mu_q + d\mu$:

$$D_{\mathrm{KL}}(q+dq\|p) = \frac{1}{2}\left[\mathrm{tr}\big(\Sigma_p^{-1}\Sigma_q\big) + (\Delta\mu + d\mu)^T\Sigma_p^{-1}(\Delta\mu + d\mu) - K + \ln\frac{|\Sigma_p|}{|\Sigma_q|}\right] \tag{A2}$$

$$= D_{\mathrm{KL}}(q\|p) + \Delta\mu^T\Sigma_p^{-1}d\mu + \frac{1}{2}d\mu^T\Sigma_p^{-1}d\mu \tag{A3}$$

The first-order term $\Delta\mu^T\Sigma_p^{-1}d\mu$ is the directional derivative (force).

The second-order term $\frac{1}{2}d\mu^T\Sigma_p^{-1}d\mu$ is the kinetic energy contribution.

With $d\mu = \dot{\mu}\,dt$:

$$\Delta D_{\mathrm{KL}} = (\nabla_\mu D_{\mathrm{KL}})\cdot\dot{\mu}\,dt + \frac{1}{2}\dot{\mu}^T\Sigma_p^{-1}\dot{\mu}\,dt^2 \tag{A4}$$

Integrating over time and identifying the action:

$$S = \int dt \left[\frac{1}{2}\dot{\mu}^T \Sigma_p^{-1} \dot{\mu} - V(\mu)\right] \tag{A5}$$

Appendix B: Dimensional Analysis and Natural Units

1. The Planck Scale Anchor

In natural units $\hbar = c = G = 1$:

- Planck length: $\ell_P = \sqrt{\hbar G/c^3} \approx 1.6 \times 10^{-35}$ m
- Planck mass: $M_P = \sqrt{\hbar c/G} \approx 2.2 \times 10^{-8}$ kg
- Planck time: $t_P = \sqrt{\hbar G/c^5} \approx 5.4 \times 10^{-44}$ s

A maximally localized bit has uncertainty $\sigma \sim \ell_P$ and precision $\Sigma^{-1} \sim \ell_P^{-2}$.

2. Mass-Precision Relation

We propose:

$$M = \kappa \cdot \mathrm{tr}\left(\Sigma^{-1}\right) \tag{B1}$$

To find κ: a single maximally localized bit ($\sigma = \ell_P$, $\Sigma^{-1} = \ell_P^{-2}$) should have mass $\sim M_P$:

$$M_P = \kappa \cdot \ell_P^{-2} \implies \kappa = M_P \ell_P^2 = \frac{\hbar \ell_P}{c^2} \tag{B2}$$

Thus:

$$\boxed{M = \frac{\hbar \ell_P}{c^2 \sigma^2} = \frac{M_P}{(\sigma/\ell_P)^2}} \tag{B3}$$

For uncertainty $\sigma = n \cdot \ell_P$ (n Planck lengths):

$$M = \frac{M_P}{n^2} \tag{B4}$$

Appendix C: Experimental Estimates

1. Superposition Mass Deficit

Consider a particle with:

- Localized state uncertainty: $\sigma_0 = 1$ pm $= 10^{-12}$ m $\approx 10^{23}\ell_P$
- Superposition uncertainty: $\sigma_+ = 1$ μm $= 10^{-6}$ m $\approx 10^{29}\ell_P$

Mass difference:

$$\Delta M = \frac{\hbar \ell_P}{c^2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_+^2}\right) \tag{C1}$$

$$\approx \frac{10^{-34} \times 10^{-35}}{10^{-16}} \times \left(\frac{1}{10^{-24}} - \frac{1}{10^{-12}}\right) \tag{C2}$$

$$\approx 10^{-53} \times 10^{24} \tag{C3}$$

$$\approx 10^{-29}\,\text{kg} \tag{C4}$$

This is comparable to the electron mass ($\sim 10^{-30}$ kg) — potentially detectable with precision measurements!

2. Black Hole Consistency Check

For a solar-mass black hole ($M_\odot \approx 2 \times 10^{30}$ kg):

Schwarzschild radius: $R_s = 2GM_\odot/c^2 \approx 3$ km

Bekenstein-Hawking entropy:

$$S_{BH} = \frac{4\pi G M_\odot^2}{\hbar c} \approx 10^{77} \tag{C5}$$

Number of bits: $N \approx S_{BH}/\ln 2 \approx 10^{77}$

Average precision per bit:

$$\langle \Sigma^{-1} \rangle = \frac{M_\odot}{\kappa N} = \frac{M_\odot c^2}{\hbar \ell_P N} \tag{C6}$$

Effective uncertainty per bit:

$$\sigma_{\text{bit}} \approx \sqrt{\frac{\hbar \ell_P N}{M_\odot c^2}} \approx 10^{-3}\,\text{m} \tag{C7}$$

This is roughly the event horizon scale divided by $\sqrt{N}$ — consistent with bits being delocalized across the horizon.

Appendix D: Gauge Structure and the Emergent Metric

In the full gauge-theoretic framework [6], agents carry gauge frames $\phi_i(c) \in \mathfrak{so}(3)$ over the base manifold $\mathcal{C}$. The transport operator:

$$\Omega_{ij} = e^{\phi_i} e^{-\phi_j} \tag{D1}$$

The emergent metric has components:

$$g_{\mu\nu}(x) = \sum_i \alpha_i(x) \cdot G_{ab}^{(i)}(x) \frac{\partial \theta_i^a}{\partial x^\mu} \frac{\partial \theta_i^b}{\partial x^\nu} \tag{D2}$$

where $G_{ab}^{(i)}$ is the Fisher metric on agent i's statistical manifold.

The temporal component:

$$g_{00}(x) = -\sum_i \alpha_i(x) \cdot \mathrm{tr}\big(\Sigma_{p_i}^{-1}(x)\big) \tag{D3}$$

The participation weight α_i can be defined as:

$$\alpha_i(x) = p_i(o_i|x) + \lambda \sum_j \beta_{ji}(x) \tag{D4}$$

This encodes Wheeler's participatory principle: existence = self-observation + being observed by others.

EXPLICIT GRADIENT EXPRESSIONS FOR SO(3) GAUSSIAN AGENTS

We derive explicit natural gradient expressions for gauge-theoretic free energy minimization with Gaussian distributions and SO(3) transport. All operations are gauge covariant.

Setup and Notation

Agent i maintains belief $q_i = \mathcal{N}(\mu_i, \Sigma_i)$ and prior $p_i = \mathcal{N}(\mu_{p,i}, \Sigma_{p,i})$ over K-dimensional fiber. Transport operator $\Omega_{ij} = \exp(\phi_i)\exp(-\phi_j) \in \mathrm{SO}(3)$ acts via representation $\rho : \mathrm{SO}(3) \to \mathrm{GL}(K)$.

The free energy decomposes as:

$$S = \sum_i \mathrm{KL}(q_i \| p_i) + \sum_{ij} \beta_{ij} \mathrm{KL}(q_i \| \Omega_{ij}[q_j]) + \sum_{ij} \gamma_{ij} \mathrm{KL}(p_i \| \tilde{\Omega}_{ij}[p_j]) \tag{5}$$

Attention weights are defined by:

$$\beta_{ij} = \frac{\exp\left[-\kappa_\beta^{-1}\mathrm{KL}(q_i \| \Omega_{ij}[q_j])\right]}{\sum_k \exp\left[-\kappa_\beta^{-1}\mathrm{KL}(q_i \| \Omega_{ik}[q_k])\right]} \tag{6}$$

Free Energy Gradient: Belief Means (Complete)

The belief alignment term is:

$$S_{\text{belief}} = \sum_j \beta_{ij}(\mu_i, \Sigma_i)\,\mathrm{KL}(q_i \| \Omega_{ij}[q_j]) \tag{7}$$

Applying the product rule:

$$\nabla_{\mu_i} S_{\text{belief}} = \sum_j \left[\frac{\partial \beta_{ij}}{\partial \mu_i}\mathrm{KL}_{ij} + \beta_{ij}\frac{\partial \mathrm{KL}_{ij}}{\partial \mu_i}\right] \tag{8}$$

The softmax derivative for attention weights gives:

$$\frac{\partial \beta_{ij}}{\partial \mu_i} = -\frac{1}{\kappa_\beta}\beta_{ij}\left[\frac{\partial \mathrm{KL}_{ij}}{\partial \mu_i} - \sum_k \beta_{ik}\frac{\partial \mathrm{KL}_{ik}}{\partial \mu_i}\right] \tag{9}$$

For the direct KL gradient:

$$\frac{\partial \mathrm{KL}_{ij}}{\partial \mu_i} = (\Omega_{ij}\Sigma_j\Omega_{ij}^T)^{-1}(\mu_i - \Omega_{ij}\mu_j) \tag{10}$$

Substituting and simplifying:

$$\begin{aligned}\nabla_{\mu_i} S_{\text{belief}} = &\sum_j \beta_{ij}(\Omega_{ij}\Sigma_j\Omega_{ij}^T)^{-1}(\mu_i - \Omega_{ij}\mu_j)\left[1 - \frac{\mathrm{KL}_{ij}}{\kappa_\beta}\right] \\ &+ \frac{1}{\kappa_\beta}\left(\sum_j \beta_{ij}\mathrm{KL}_{ij}\right)\sum_k \beta_{ik}(\Omega_{ik}\Sigma_k\Omega_{ik}^T)^{-1}(\mu_i - \Omega_{ik}\mu_k)\end{aligned} \tag{11}$$

Including the self-term $\mathrm{KL}(q_i \| p_i)$:

$$\boxed{\nabla_{\mu_i} S = \Sigma_i^{-1}(\mu_i - \mu_{p,i}) + \sum_j \beta_{ij}(\Omega_{ij}\Sigma_j\Omega_{ij}^T)^{-1}(\mu_i - \Omega_{ij}\mu_j)\left[1 - \frac{\mathrm{KL}_{ij}}{\kappa_\beta}\right] + R_\mu} \tag{12}$$

where the reweighting term is:

$$R_\mu = \frac{1}{\kappa_\beta}\left(\sum_j \beta_{ij}\mathrm{KL}_{ij}\right)\sum_k \beta_{ik}(\Omega_{ik}\Sigma_k\Omega_{ik}^T)^{-1}(\mu_i - \Omega_{ik}\mu_k) \tag{13}$$

The natural gradient is:

$$\tilde{\nabla}_{\mu_i} S = \Sigma_i \nabla_{\mu_i} S \tag{14}$$

Free Energy Gradient: Belief Covariances (Complete)

Similarly for covariances, the product rule gives:

$$\nabla_{\Sigma_i} S_{\text{belief}} = \sum_j \left[\frac{\partial \beta_{ij}}{\partial \Sigma_i} \text{KL}_{ij} + \beta_{ij} \frac{\partial \text{KL}_{ij}}{\partial \Sigma_i} \right] \tag{15}$$

The softmax derivative is:

$$\frac{\partial \beta_{ij}}{\partial \Sigma_i} = -\frac{1}{\kappa_\beta} \beta_{ij} \left[\frac{\partial \text{KL}_{ij}}{\partial \Sigma_i} - \sum_k \beta_{ik} \frac{\partial \text{KL}_{ik}}{\partial \Sigma_i} \right] \tag{16}$$

For Gaussian KL divergence:

$$\frac{\partial \text{KL}_{ij}}{\partial \Sigma_i} = \frac{1}{2} \left[-\Sigma_i^{-1} + (\Omega_{ij} \Sigma_j \Omega_{ij}^T)^{-1} \right] \tag{17}$$

The complete gradient including self-term is:

$$\begin{aligned} \nabla_{\Sigma_i} S = & \frac{1}{2} \left[-\Sigma_i^{-1} + \Sigma_{p,i}^{-1} \right] \\ & + \frac{1}{2} \sum_j \beta_{ij} \left[-\Sigma_i^{-1} + (\Omega_{ij} \Sigma_j \Omega_{ij}^T)^{-1} \right] \left[1 - \frac{\text{KL}_{ij}}{\kappa_\beta} \right] \\ & + \frac{1}{2\kappa_\beta} \left(\sum_j \beta_{ij} \text{KL}_{ij} \right) \sum_k \beta_{ik} \left[-\Sigma_i^{-1} + (\Omega_{ik} \Sigma_k \Omega_{ik}^T)^{-1} \right] \end{aligned} \tag{18}$$

The natural gradient on $\mathcal{S}_{++}^K$ is:

$$\tilde{\nabla}_{\Sigma_i} S = \Sigma_i (\nabla_{\Sigma_i} S) \Sigma_i \tag{19}$$

Gauge Covariance: Under gauge transformation $g \in \text{SO}(3)$:

$$\Sigma_i \mapsto g \Sigma_i g^T \tag{20}$$

$$\nabla_{\Sigma_i} S \mapsto g (\nabla_{\Sigma_i} S) g^T \tag{21}$$

$$\tilde{\nabla}_{\Sigma_i} S \mapsto g (\tilde{\nabla}_{\Sigma_i} S) g^T \tag{22}$$

This bilinear form $\Sigma(\cdot)\Sigma$ is the unique gauge-covariant gradient structure on $\mathcal{S}_{++}^K$.

Retraction: Staying on the SPD Manifold

After computing $\tilde{\nabla}_{\Sigma_i} S$, we update via retraction to ensure Σ_i remains positive-definite:

$$\Sigma_i^{\text{new}} = \mathcal{R}_{\Sigma_i} (-\eta_\Sigma \tilde{\nabla}_{\Sigma_i} S) \tag{23}$$

We use the **affine-invariant exponential map**:

$$\mathcal{R}_\Sigma(\Delta) = \Sigma^{1/2} \exp\left(\Sigma^{-1/2}\Delta\Sigma^{-1/2}\right)\Sigma^{1/2} \tag{24}$$

where $\exp(\cdot)$ is the matrix exponential. This retraction:

- Preserves positive-definiteness (exponential of symmetric matrix is SPD)
- Is affine-invariant under $\Sigma \mapsto A\Sigma A^T$ for invertible A
- Is gauge covariant: $\mathcal{R}_{g\Sigma g^T}(g\Delta g^T) = g\mathcal{R}_\Sigma(\Delta)g^T$

Computational Implementation: Using eigendecomposition $\Sigma = V\Lambda V^T$:

$$\mathcal{R}_\Sigma(\Delta) = V\sqrt{\Lambda}\exp\left(\sqrt{\Lambda}^{-1}V^T\Delta V\sqrt{\Lambda}^{-1}\right)\sqrt{\Lambda}V^T \tag{25}$$

This is the Riemannian exponential map—the geodesic retraction following the unique geodesic from Σ in direction Δ for unit parameter.

Complete Gauge Frame Gradients (Product Rule)

The free energy contains gauge frames ϕ_i through both transport operators and attention weights. For the belief alignment term:

$$S_{\text{belief}} = \sum_j \beta_{ij}(\phi_i)\,\mathrm{KL}(q_i\|\Omega_{ij}(\phi_i)[q_j]) \tag{26}$$

Applying the product rule:

$$\nabla_{\phi_i} S_{\text{belief}} = \sum_j \left[\frac{\partial\beta_{ij}}{\partial\phi_i}\mathrm{KL}(q_i\|\Omega_{ij}[q_j]) + \beta_{ij}\frac{\partial}{\partial\phi_i}\mathrm{KL}(q_i\|\Omega_{ij}[q_j])\right] \tag{27}$$

Term 1: Direct KL Gradient

For the transported KL divergence, using $\frac{\partial\Omega_{ij}}{\partial\phi_i^a} = G_a\Omega_{ij}$ where $\{G_a\}$ are SO(3) generators:

$$\frac{\partial}{\partial\phi_i^a}\mathrm{KL}(q_i\|\Omega_{ij}[q_j]) = \mathrm{tr}\left[G_a\Omega_{ij}\frac{\partial}{\partial\Omega_{ij}}\mathrm{KL}(q_i\|\Omega_{ij}[q_j])\right] \tag{28}$$

For Gaussians, the functional derivative is:

$$\frac{\partial}{\partial\Omega_{ij}}\mathrm{KL}(q_i\|\Omega_{ij}[q_j]) = (\Omega_{ij}\Sigma_j\Omega_{ij}^T)^{-1}(\mu_i - \Omega_{ij}\mu_j)\mu_j^T + \mathrm{tr}\left[(\Omega_{ij}\Sigma_j\Omega_{ij}^T)^{-1}\Sigma_j\right]\Omega_{ij} \tag{29}$$

Term 2: Attention Weight Gradient (Softmax Derivative)

The softmax structure of β_{ij} yields:

$$\frac{\partial \beta_{ij}}{\partial \phi_i^a} = -\frac{1}{\kappa_\beta}\beta_{ij}\left[\frac{\partial}{\partial \phi_i^a}\mathrm{KL}(q_i \| \Omega_{ij}[q_j]) - \sum_k \beta_{ik}\frac{\partial}{\partial \phi_i^a}\mathrm{KL}(q_i \| \Omega_{ik}[q_k])\right] \tag{30}$$

This follows from $\frac{\partial}{\partial x_i}\mathrm{softmax}(x)_j = \mathrm{softmax}(x)_j(\delta_{ij} - \mathrm{softmax}(x)_i)$ applied with $x_j = -\kappa_\beta^{-1}\mathrm{KL}_j$.

Combined Gradient (Two-Term Structure)

Substituting the softmax derivative and simplifying:

$$\begin{aligned}\nabla_{\phi_i^a} S_{\text{belief}} = \sum_j \beta_{ij}\frac{\partial \mathrm{KL}_{ij}}{\partial \phi_i^a}\left[1 - \frac{\mathrm{KL}_{ij}}{\kappa_\beta}\right] \\ + \frac{1}{\kappa_\beta}\left(\sum_j \beta_{ij}\mathrm{KL}_{ij}\right)\left(\sum_k \beta_{ik}\frac{\partial \mathrm{KL}_{ik}}{\partial \phi_i^a}\right)\end{aligned} \tag{31}$$

where $\mathrm{KL}_{ij} \equiv \mathrm{KL}(q_i \| \Omega_{ij}[q_j])$.

Physical Interpretation

The complete gradient has two interpretable terms:

$$\boxed{\nabla_{\phi_i} S = \underbrace{\sum_j \beta_{ij}\nabla_{\phi_i}\mathrm{KL}_{ij}}_{\text{Direct alignment}} + \underbrace{\frac{1}{\kappa_\beta}\left(\sum_j \beta_{ij}\mathrm{KL}_{ij}\right)\left(\sum_k \beta_{ik}\nabla_{\phi_i}\mathrm{KL}_{ik}\right)}_{\text{Attention reweighting}}} \tag{32}$$

- **Direct alignment:** Gradient of weighted KL divergences (as if β were constant)
- **Attention reweighting:** How changing ϕ_i redistributes attention across neighbors

When $\kappa_\beta \to \infty$ (soft/uniform attention), the reweighting term vanishes. When $\kappa_\beta \to 0$ (hard attention), reweighting dominates.

Prior Alignment (Similar Structure)

The prior alignment gradient has identical form:

$$\nabla_{\phi_i} S_{\text{prior}} = \sum_j \gamma_{ij} \nabla_{\phi_i} \mathrm{KL}(p_i \| \tilde{\Omega}_{ij}[p_j]) + \frac{1}{\kappa_\gamma} \left(\sum_j \gamma_{ij} \mathrm{KL}(p_i \| \tilde{\Omega}_{ij}[p_j]) \right) \left(\sum_k \gamma_{ik} \nabla_{\phi_i} \mathrm{KL}^p_{ik} \right) \tag{33}$$

Complete Update Equations

The gauge-covariant update scheme is:

$$\mu_i^{t+1} = \mu_i^t - \eta_\mu \Sigma_i^t \nabla_{\mu_i} S \tag{34}$$

$$\Sigma_i^{t+1} = \left(\Sigma_i^t\right)^{1/2} \exp\left[\left(\Sigma_i^t\right)^{-1/2} \left(-\eta_\Sigma \tilde{\nabla}_{\Sigma_i} S \right) \left(\Sigma_i^t\right)^{-1/2} \right] \left(\Sigma_i^t\right)^{1/2} \tag{35}$$

$$\phi_i^{t+1} = \phi_i^t - \eta_\phi \nabla_{\phi_i} S \tag{36}$$

where learning rates satisfy timescale separation $\eta_\mu : \eta_p : \eta_\phi \sim 1 : \epsilon : \epsilon^2$ with $\epsilon \ll 1$.

Computational Complexity

Per gradient step with N agents, dimension K:

- Computing all Ω_{ij}: $O(N^2 K^3)$ matrix exponentials
- Computing all transported KL divergences: $O(N^2 K^3)$ inverse covariances
- Natural gradient projections: $O(NK^3)$ per agent
- Exponential map retraction: $O(NK^3)$ eigendecomposition per agent

Total complexity: $O(N^2 K^3)$ per step, dominated by pairwise coupling.

Numerical Stability and Verification

Positive-Definiteness: The exponential map retraction $\mathcal{R}_\Sigma(\Delta)$ maps to $\mathcal{S}^K_{++}$ for any symmetric Δ, guaranteeing $\Sigma_i \succ 0$ throughout optimization.

Gauge Orbit Preservation: Under simultaneous transformation $\mu_i \mapsto g\mu_i$, $\Sigma_i \mapsto g\Sigma_i g^T$, $\phi_i \mapsto \phi_i + \xi$ for all i, the gradients transform covariantly and updates preserve gauge invariance of the free energy functional.

Gradient Verification: All analytical gradients verified against:

- Numerical finite differences (tolerance 10^{-6})
- PyTorch automatic differentiation (relative error $< 10^{-8}$)