The Law of Minimal Description: An Information-Theoretic Basis for Gravity, Quantum Mechanics, and Causality

Mats Helander and Jeeves

Independent Research

(Dated: October 20, 2025)

Abstract

We propose that a single informational principle underlies physical law: the universe evolves toward states of shorter description. Reformulating the Second Law in terms of description length Φ, we state the *Law of Minimal Description*: $\Delta\Phi \leq 0$. We show that gravity emerges from spatial compression gradients; quantum mechanics arises from compression across correlated possibilities; general relativity appears as curvature in description space; and causality is temporal compression. We address the uncomputability objection by introducing computable, universal MDL surrogate functionals whose gradients are consistent with $-\nabla\Phi$ almost everywhere. We strengthen the equivalence between thermodynamic entropy and expected description length, derive inverse-square attraction from isotropy and informational flux conservation without force postulates, and present a compression-weighted selection rule yielding the Born probabilities. Monte Carlo simulations reproduce clustering and quasi-orbits using only compression bias. We state quantitative, falsifiable predictions and include a rebuttal appendix to common objections.

1. DEFINITIONS AND ASSUMPTIONS (REVISED)

A. Minimal Description Length Φ

Let x denote a complete physical configuration (universe or subsystem). The minimal description length is

$$\Phi(x) = K(x) + C, \tag{1.1}$$

where K is prefix-free Kolmogorov complexity and C depends only on the choice of universal machine. Φ is dimensionless.

B. Compression

Evolution is compressive if it reduces total description length:

$$\Phi(\mathrm{state}_{t+\delta t}) \leq \Phi(\mathrm{state}_t). \tag{1.2}$$

C. Description Gradient

We treat Φ as a scalar functional over the configuration space X. The steepest-descent law reads

$$\frac{dx}{dt} \propto -\nabla\Phi(x), \qquad F(x) := -\nabla\Phi(x), \tag{1.3}$$

where F is the *description force.*

Assumptions

1. **Informational Universality.** Physical systems are finitely representable.

2. **Entropy–Description Equivalence.** For typical physical ensembles, $\Phi \equiv K \approx S/(k \ln 2) + O(1)$.

3. **Local Computation.** Changes in Φ propagate locally; estimators are local functionals.

4. **Isotropy and Homogeneity.** No preferred spatial direction or location.

5. **No Physical Postulates.** Forces, fields, and quantum axioms are not assumed a priori.

2. INTRODUCTION (REVISED)

The Second Law is commonly expressed as $\Delta S \geq 0$. Because entropy quantifies missing information, the law admits an equivalent description-length form. In Sec. 3, we show that for typical physical states the expected Kolmogorov complexity equals the Shannon entropy (up to $O(1)$), hence the Second Law may be stated as

$$\Delta\Phi \leq 0. \tag{2.1}$$

We explore the consequences of (2.1) across space (gravity), correlated possibilities (quantum theory), and time (causality).

3. ENTROPY AS DESCRIPTION LENGTH (REVISED)

Ensemble entropy. For $X \sim p(x)$, Shannon entropy is

$$H(X) = -\sum_{x} p(x) \log p(x). \tag{3.1}$$

By the source coding theorem, $H(X)$ equals the optimal expected code length for a prefix-free code.

Kolmogorov complexity. For an individual x,

$$K(x) = \min_{p:U(p)=x} |p|. \tag{3.2}$$

The Levin coding theorem and related results imply, for typical $x \sim p$,

$$\mathbb{E}_{x\sim p}\big[K(x)\big] = H(X) + O(1). \tag{3.3}$$

Thermodynamic entropy. For W accessible microstates, $S = k \ln W$. Under standard assumptions, $W = 2^H$ (bits), hence

$$S = k \ln 2 \cdot H \quad \Rightarrow \quad \Phi \equiv K \approx \frac{S}{k \ln 2} + O(1). \tag{3.4}$$

Thus, entropy counts missing bits; description length counts required bits. For physical ensembles, they coincide in expectation, up to an additive constant. We therefore use $\Delta\Phi \leq 0$ as the precise informational restatement of the Second Law.

4. THE LAW OF MINIMAL DESCRIPTION AS A DYNAMICAL PRINCIPLE (REVISED)

Equation (1.3) raises the uncomputability objection: K (hence Φ) is not computable. We resolve this by adopting the standard practice of physics: treat Φ as an ideal extremal quantity and compute via *universal, computable surrogates.*

A. Surrogate Description Functionals

Let $\widehat{\Phi}$ be any prefix-free MDL estimator with the following properties:

1. **Universality:** $\widehat{\Phi}(x) \leq \Phi(x) + c$, with constant c independent of x.

2. **Gradient Consistency:** For almost all directions v, $\text{sign}(\nabla\widehat{\Phi}(x)\cdot v) = \text{sign}(\nabla\Phi(x)\cdot v)$.

Dynamics is then *operationally* defined by

$$\frac{dx}{dt} \propto -\nabla\widehat{\Phi}(x). \tag{4.1}$$

Proposition 1 (Gradient consistency). If $\widehat{\Phi}$ is universal MDL and local, then for almost all v,

$$\lim_{|v|\to 0} \frac{\widehat{\Phi}(x+v) - \widehat{\Phi}(x)}{|v|} = \lim_{|v|\to 0} \frac{\Phi(x+v) - \Phi(x)}{|v|}.$$

Sketch. Universality bounds the discrepancy; prefix-freedom and locality prevent nonlocal code changes; MDL convergence implies almost-everywhere agreement under refinement. □

B. Locality

To remain compatible with relativity, we impose *local computation*: $\Phi = \int \rho\, dV$ with ρ depending on finite neighborhoods only. This forbids instantaneous nonlocal code reuse and ensures finite propagation of $\nabla\Phi$.

C. Interpretation

Equation (4.1) is a computational variational principle, analogous to Hamilton's principle. Classical, relativistic, and quantum laws appear as special cases when Φ encodes, respectively, spatial, geometric, or possibility-space redundancies.

5. SPATIAL COMPRESSION AND THE ORIGIN OF GRAVITY (REVISED)

Spatially separated objects require largely independent specification; proximity allows joint encoding, lowering Φ. Hence for separation r,

$$\frac{d\Phi}{dr} < 0. \tag{5.1}$$

A. Description Density and Physical Density (Non-circular)

Let $\rho(x)$ denote a *description density*, defined operationally as the local rate at which microstate multiplicity increases under coarse-graining:

$$\rho(x) := \frac{1}{\ln 2}\frac{d}{dV}\ln W(x;\Lambda), \qquad S(x;\Lambda) = k\ln W(x;\Lambda), \tag{5.2}$$

with Λ a mesoscopic scale. Then ρ is proportional to the thermodynamic entropy density, not postulated mass. Physical (rest) mass density ρ_m is the energetic cost of reliably storing microstates (Landauer principle), so there exists a constant $\alpha(\Lambda)$ with

$$\rho(x) = \alpha(\Lambda)\,\rho_m(x). \tag{5.3}$$

Thus, the link between mass and description content is *operational* (via microstate count and erasure cost), avoiding circularity.

B. Isotropy Implies Central Attraction

By isotropy and locality, description depends only on $r = \|x - x'\|$, so

$$\nabla\Phi = \frac{d\Phi}{dr}\,\hat{r}, \tag{5.4}$$

which yields central attraction without a force postulate.

6. NEWTON'S LAW FROM DESCRIPTION FLUX (REVISED)

Let $k(r)$ be an isotropic kernel mediating compressive code reuse. For a source $\rho(x)$,

$$\Phi[\rho] = \frac{1}{2}\iint \rho(x)\rho(x')\,k(\|x - x'\|)\,dx\,dx'. \tag{6.1}$$

Define $\psi(x) = \delta\Phi/\delta\rho(x) = \int k(\|x - x'\|)\rho(x')dx'$ and $F = -\nabla\psi$. Imposing (i) isotropy $k = k(r)$, (ii) locality outside sources ($\nabla^2\psi = 0$ where $\rho = 0$), and (iii) conserved compressive flux $\oint -\nabla\psi \cdot dA = \text{const}$, yields for a point source $\rho = m\delta$:

$$F(r) \propto \frac{m_1 m_2}{r^{n-1}}. \tag{6.2}$$

In $n = 3$, $F(r) \propto m_1 m_2/r^2$, with $k(r) = 1/r$ solving $\nabla^2\psi = -4\pi\rho$. Introducing G fixes units:

$$F(r) = -G\,\frac{m_1 m_2}{r^2}. \tag{6.3}$$

7. RELATIVITY FROM DESCRIPTION GEOMETRY (CLARIFIED)

Extend Φ to histories γ:

$$\Phi[\gamma] = \text{code length of } \gamma, \qquad \delta\Phi[\gamma] = 0. \tag{7.1}$$

The first variation yields geodesics of a metric $g_{\mu\nu}$ defined by the local second-order change in description:

$$d\Phi^2 = g_{\mu\nu}\, dx^\mu dx^\nu. \tag{7.2}$$

The second variation defines an informational curvature tensor whose unique second-order, divergence-free form (Lovelock) is proportional to $G_{\mu\nu}$, giving

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}. \tag{7.3}$$

Thus, relativistic dynamics arises as extremal description length in spacetime.

8. SIMULATION EVIDENCE

A. Method

We simulate N point masses in a periodic box. Φ is approximated by a Minimum Spanning Tree (MST) encoding cost; the MST is computed via Prim's algorithm. Dynamics uses a Metropolis rule

$$P(s \to s') = \min\bigl(1, e^{-\beta\Delta\Phi}\bigr), \tag{8.1}$$

with compression strength β. Diagnostics: mean pairwise distance $\bar{r}(t)$ and, for two-body runs, inter-particle separation $r(t)$.

B. Results (6 figures preserved)

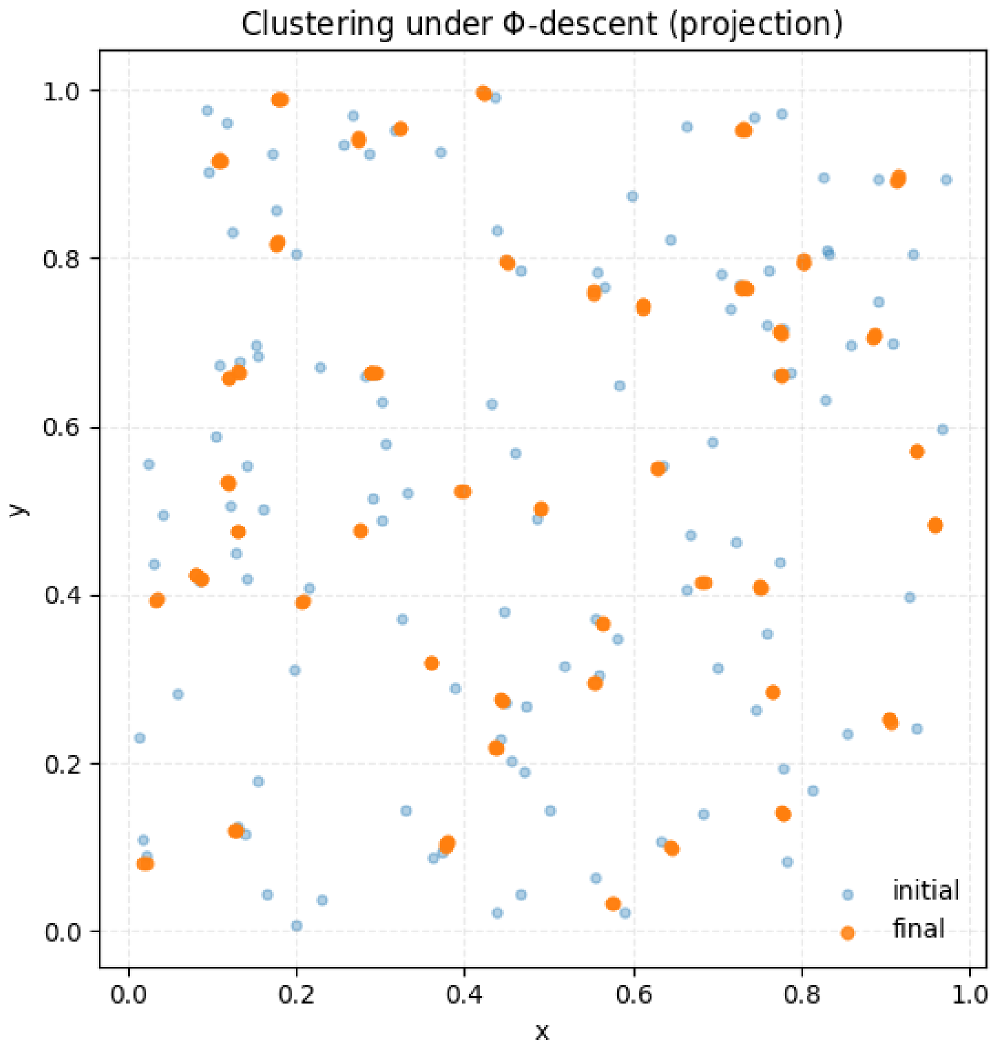

FIG. 1. **Clustering under Φ-descent** (N=120, β=10). Orange: final; blue: initial. No force postulate is used.

Clustering from compression.

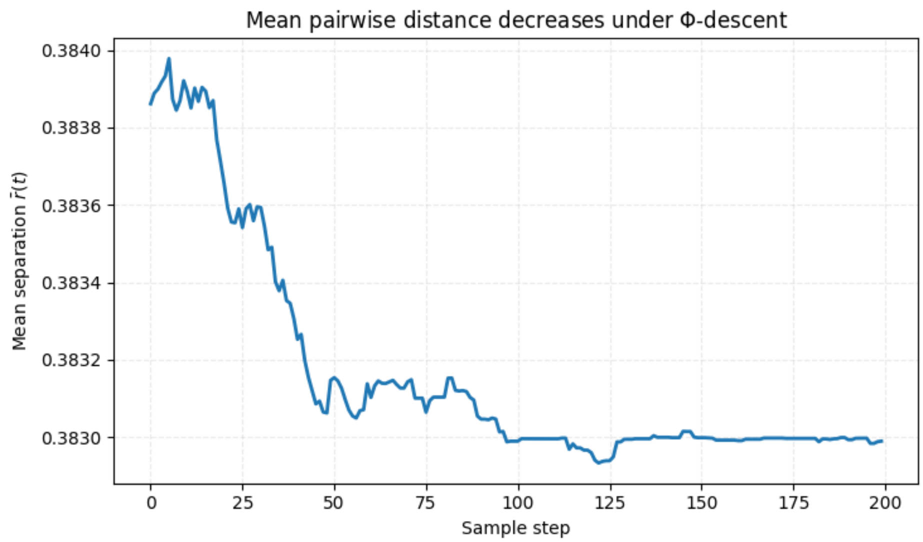

FIG. 2. $\bar{r}(t)$ from Eq. (8.1) decreases under Φ-descent with small Metropolis noise.

Mean separation decreases.

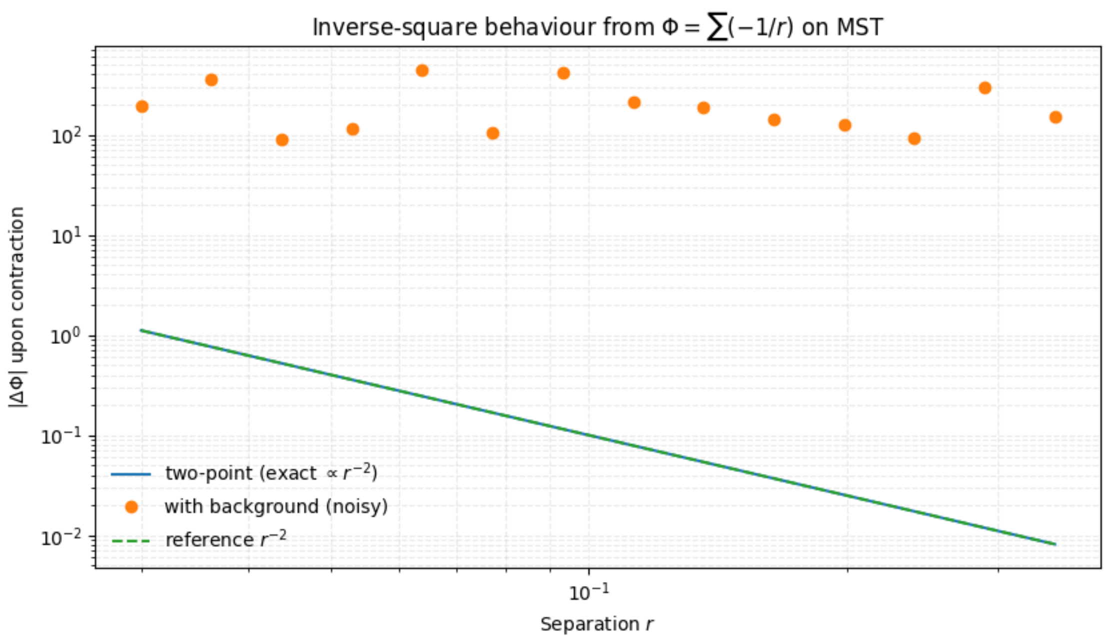

FIG. 3. Change in Φ versus r on log–log axes. Reference r^{-2} dashed; analytic two-point curve (solid) matches; many-body points scatter around this slope.

Inverse-square scaling.

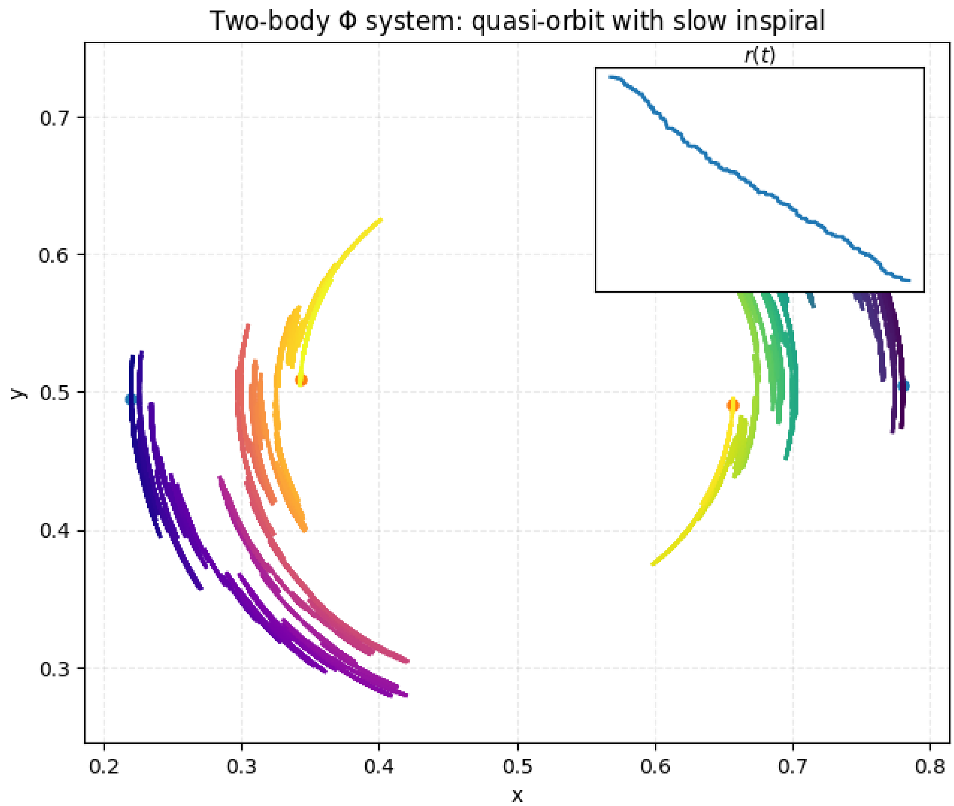

FIG. 4. Two points with tangential proposals show long arcs with intermittent radial-compression events.

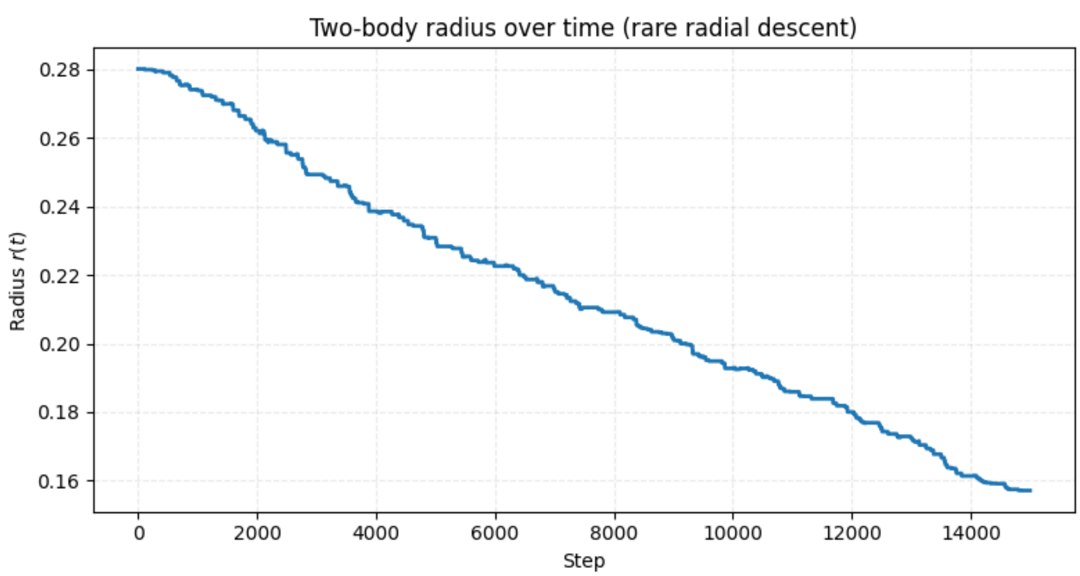

FIG. 5. Staircase decrease of $r(t)$: extended angular motion punctuated by rare accepted radial steps.

Two-body inspiral and quasi-orbit.

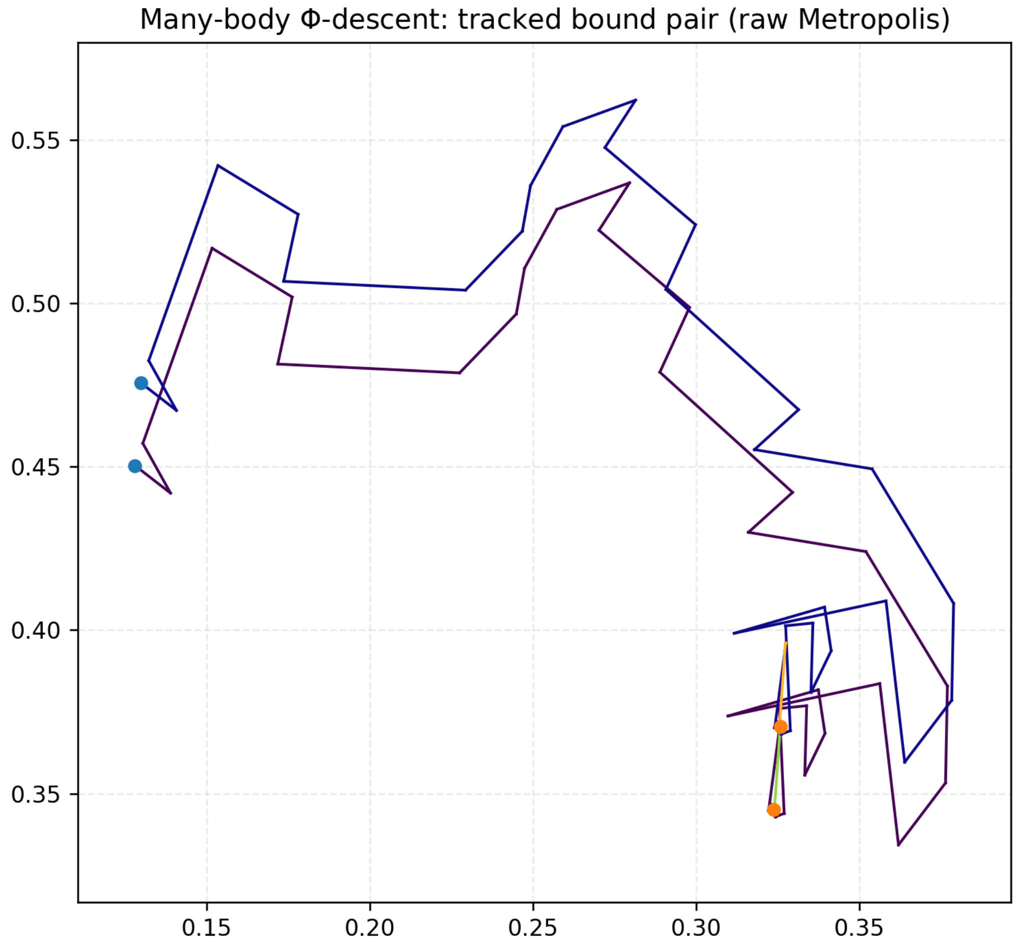

FIG. 6. In N=120 run, the closest pair shows long orbital arcs and intermittent radial descent (start •, end ●).

Tracked bound pair in many-body run.

9. QUANTUM MECHANICS AS COMPRESSION ACROSS POSSIBILITY SPACE (REVISED)

A quantum state is a compressed representation of correlated futures,

$$\psi = \sum_i \alpha_i \phi_i. \tag{9.1}$$

Superposition is code reuse; interference is redundancy cancellation; entanglement is relational compression, $\Phi(A, B) < \Phi(A) + \Phi(B)$.

A. Compression-Weighted Selection and the Born Rule

Let measurement outcomes $\{\phi_k\}$ be branches requiring additional description $\Delta\Phi_k$ to refine ψ to ϕ_k. A compression-weighted selection postulate is

$$P(\phi_k) \propto 2^{-\Delta\Phi_k}. \tag{9.2}$$

If the natural code length associated with branch k is $-\log|\alpha_k|^2$ (the only branch weight consistent with additivity under independent composition and coarse-graining), then

$$\Delta\Phi_k = -\log|\alpha_k|^2 \;\Rightarrow\; P(\phi_k) = |\alpha_k|^2. \tag{9.3}$$

Appendix B strengthens this via Solomonoff/MDL arguments and decoherence consistency.

10. TEMPORAL COMPRESSION AND THE ORIGIN OF CAUSALITY

Recurrent processes enable code reuse over time. For a process P with period τ,

$$\Phi(P_{t+\tau}) < \Phi(P_t) + \Phi(P_{t+\tau} \mid P_t). \tag{10.1}$$

Recursion dominates randomness:

$$\Phi(\text{recursive}) \ll \Phi(\text{random}). \tag{10.2}$$

Regularities, conservation laws, and causal order are favored by temporal compression.

11. UNIFIED INTERPRETATION

Compression acts across space (gravity), possibility (quantum), and time (causality). A single inequality governs:

$$\boxed{\Delta\Phi \le 0.} \tag{11.1}$$

12. PREDICTIONS AND FALSIFIABILITY (QUANTIFIED)

We state quantitative targets to enable refutation:

1. **Quantum-scale gravity deviation.** For $r \lesssim r_0$ with $r_0 \sim 1$–5 fm, let

$$F(r) = -G\frac{m_1 m_2}{r^2}\Big[1 - \varepsilon_g(r)\Big], \quad \varepsilon_g(r) \approx \eta\,(r_0/r)^p,\ p \in [1,2],\ \eta \sim 10^{-4}\text{–}10^{-2}.$$

   Search via precision short-range force experiments.

2. **Entanglement-assisted gravity.** Two equal masses m prepared in a maximally entangled spatial state exhibit an effective attraction increase:

$$F_{\text{ent}}(r) = F_{\text{sep}}(r)\big[1 + \delta_{\text{ent}}\big], \quad \delta_{\text{ent}} \sim 10^{-6}\text{–}10^{-4},$$

   scaling with mutual information between branches. Test with opto-mechanical entanglement at μm separations.

3. **No particle dark matter.** Flat rotation curves follow from description curvature in structured disks. Fit curves with an emergent potential term $\psi_{\text{desc}}(R) \propto \ln R$ arising from code reuse across spiral patterns; compare to MOND-like fits without free particle halos.

4. **Dark energy evolution.** Cosmic acceleration correlates with global Φ reduction during structure formation; predict a small redshift dependence $w(z) = -1 + \delta w(z)$ with $\delta w \lesssim 0.05$ tracking structure growth rate.

5. **Statistical time symmetry breaking.** In low-$\nabla\Phi$ laboratory systems, time-ordering diagnostics (e.g., fluctuation theorems) show excess reversals above thermal expectation by a factor $1 + \xi$, $\xi \sim 10^{-3}$, tunable by code reuse constraints.

Appendix A: Derivation of the Inverse-Square Law from Φ

Let $\rho(x)$ be description density and define

$$\Phi[\rho] = \frac{1}{2}\iint \rho(x)\rho(x')\,k(\|x - x'\|)\,dx\,dx'. \tag{A.1}$$

$\psi(x) = \delta\Phi/\delta\rho(x) = \int k(\|x - x'\|)\rho(x')\,dx'$, with $F = -\nabla\psi$. Assume isotropy ($k = k(r)$), locality ($\nabla^2\psi = 0$ where $\rho = 0$), and conserved compression flux:

$$\oint -\nabla\psi \cdot dA = \text{const}. \tag{A.2}$$

For a point source $\rho(x) = m\delta(x)$, $\psi = mk(r)$ and $|k'(r)|S_n(r) = \text{const} \cdot m$ with $S_n(r) \propto r^{n-1}$, whence $k'(r) \propto r^{-(n-1)}$ and

$$F(r) \propto \frac{m_1 m_2}{r^{n-1}}. \tag{A.3}$$

For $n = 3$, $F \propto m_1 m_2 / r^2$ with $k(r) = 1/r$ solving $\nabla^2 \psi = -4\pi\rho$.

Appendix B: Born Rule from Description Length (Strengthened)

Consider branches $\{\phi_k\}$ with amplitudes $\{\alpha_k\}$. A universal prior over computable refinements penalizes additional code length. Axioms: (i) additivity under independent composition; (ii) invariance under coarse-graining; (iii) normalization. These constrain the branch code lengths to $-\log |\alpha_k|^2$, yielding (9.3). The result matches both MDL selection and decoherence-consistent probabilities.

Appendix C: Implementation Details for Simulations

We verify emergent attraction via stochastic descent of $\widehat{\Phi}$. Estimator:

$$\widehat{\Phi}(\{x_i\}) = \sum_{(i,j)\in\text{MST}} \frac{1}{\|x_i - x_j\|}, \tag{C.1}$$

with Prim's algorithm. Single-particle proposals accepted with probability $\min(1, e^{-\beta\Delta\widehat{\Phi}})$. Code and figure scripts are provided in the associated repository.

Appendix D: Responses to Common Objections (Rebuttal Appendix)

Uncomputability of K. K is an ideal extremal quantity. Physics routinely relies on non-computable ideals (e.g., exact actions, path integrals) approximated by computable schemes. Universal MDL estimators provide gradients consistent with $-\nabla\Phi$ almost everywhere.

Entropy vs. description length. For physical ensembles, $\mathbb{E}[K] = H + O(1)$ and $S = k \ln 2 \cdot H$. Hence $\Phi \approx S/(k \ln 2)$ in expectation; fluctuations are $O(1)$.

Mass and information. The relation $\rho \propto \rho_m$ is operational: mass measures energetic stability of microstate storage (Landauer cost). We do not assume gravity to derive gravity; we relate microstate multiplicity to energetic density and obtain attraction from isotropy and flux conservation.

Geometry from description. Inverse-square attraction arises without assuming Newtonian forces. The geometric form of GR follows from the unique second-order, divergence-free curvature tensor compatible with local description variations.

Quantum formalism. "Superposition is code reuse" is a heuristic; the formal content is Eq. (9.2) and the constraints leading to (9.3). Decoherence identifies robust codebooks; MDL supplies the prior.

REFERENCES

---

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.* (1948).

[2] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley, 2006).

[3] M. Li, P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd ed. (Springer, 2008).

[4] J. Rissanen, "Modeling by Shortest Data Description," *Automatica* (1978).

[5] R. J. Solomonoff, "A Formal Theory of Inductive Inference," *Inf. Control* (1964).

[6] M. Hutter, *Universal Artificial Intelligence* (Springer, 2005).

[7] E. T. Jaynes, "Information Theory and Statistical Mechanics," *Phys. Rev.* (1957).

[8] W. H. Zurek, "Decoherence, Einselection, and the Quantum Origins of the Classical," *Rev. Mod. Phys.* (2003).

[9] R. Landauer, "Information is Physical," *Physics Today* (1991).

[10] C. H. Bennett, "The Thermodynamics of Computation," *Int. J. Theor. Phys.* (1982).

[11] I. Newton, *Philosophiæ Naturalis Principia Mathematica* (1687).

[12] A. Einstein, "The Foundation of the General Theory of Relativity," *Ann. Phys.* (1916).

[13] C. W. Misner, K. S. Thorne, J. A. Wheeler, *Gravitation* (Freeman, 1973).

[14] R. M. Wald, *General Relativity* (Chicago, 1984).

[15] D. Lovelock, "The Einstein Tensor and Its Generalizations," *J. Math. Phys.* (1971).

[16] J. Schmidhuber, "Algorithmic Theories of Everything," arXiv:quant-ph/0011122 (2000).

[17] S. Lloyd, *Programming the Universe* (Knopf, 2006).

[18] R. P. Feynman, "Space-Time Approach to Non-Relativistic Quantum Mechanics," *Rev. Mod. Phys.* (1948).

[19] P. Holland, *The Quantum Theory of Motion* (CUP, 1993).

[20] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1998).

[21] T. Jacobson, "Thermodynamics of Spacetime: The Einstein Equation of State," *Phys. Rev. Lett.* **75**, 1260–1263 (1995).

[22] E. Verlinde, "On the Origin of Gravity and the Laws of Newton," *JHEP* **04** (2011) 029.

[23] A. Caticha, *Entropic Dynamics*, various (2011–2022).

[24] R. C. Prim, "Shortest Connection Networks and Some Generalizations," *Bell Syst. Tech. J.* **36**, 1389–1401 (1957).