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 address uncomputability by introducing computable, universal MDL surrogate functionals with gradients consistent with $-\nabla\Phi$ almost everywhere (Proposition 1, detailed in Appendix G). We strengthen the equivalence between thermodynamic entropy and expected description length and resolve the apparent sign paradox (Sec. 3.5) by system–environment bookkeeping. In space, inverse-square attraction follows from isotropy, locality and conserved description flux; in spacetime, a coding metric arises from the second variation of Φ and, under locality and diffeomorphism invariance, yields the Einstein equations via Lovelock's theorem (Sec. 7). In possibility space, unitary evolution appears as code-preserving isometries; incompatible codebooks formalize non-commutation; entanglement is algorithmic mutual compression; and MDL selection leads to Born probabilities (Sec. 9, App. B). Simulations reproduce clustering and quasi-orbits using only compression bias (6 figures preserved). We state quantitative predictions and include a rebuttal appendix.

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

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

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

B. Compression

Evolution is compressive if it reduces total description length:

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

C. Description Gradient

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

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

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; admissible 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 a description-length form. In Sec. 3, we show $\mathbb{E}[K] = H + O(1)$ and $S = k\ln 2 \cdot H$, yielding

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

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 $x \sim p$,

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

Typical means x lies in a set of measure $\geq 1 - 2^{-c}$ for some constant c, equivalently strings with $p(x) \gtrsim 2^{-H-c}$; highly atypical incompressible strings satisfy $K(x) \approx |x|$.

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.

A. Entropy Direction and the Sign of $\Delta\Phi$

A common objection is that $\Delta\Phi \leq 0$ (shorter descriptions) contradicts $\Delta S \geq 0$ (entropy increase). The resolution is bookkeeping. Let $L(M)$ be the model code length and $L(D|M)$ the data code length for the microstate data D:

$$\Phi_{\text{tot}} = L(M) + L(D|M). \tag{3.5}$$

Compression reduces Φ_{tot} by investing bits in $L(M)$ (regularities) to reduce $L(D|M)$. For an open subsystem, thermodynamic S pertains to $L(D|M)$, which can *increase* (entropy production) while the joint description with environment (including the updated model) *decreases.* Landauer's principle enforces that entropy exported to the environment pays the energetic cost of reducing total description. Therefore,

$$\Delta\Phi_{\text{tot}} \leq 0 \quad \text{with} \quad \Delta S_{\text{subsys}} \geq 0 \tag{3.6}$$

is consistent: global description shortens while local entropy grows.

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

Equation (1.3) raises the uncomputability objection. We treat Φ as an ideal extremal quantity, computed via *universal, computable surrogates.*

A. Surrogate Description Functionals

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

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 defined operationally by

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

Proposition 1 is stated here and proved in Appendix G.

B. Locality

We impose *local computation*: $\Phi = \int \rho \, dV$ with ρ depending on finite neighborhoods only, forbidding instantaneous nonlocal code reuse and ensuring finite propagation of $\nabla\Phi$.

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

Spatially separated objects require largely independent specification; proximity allows joint encoding, lowering Φ:

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

A. Description Density and Physical Density (Operational)

Define the local description density via microstate multiplicity under coarse-graining scale Λ:

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

Mass density ρ_m measures the energetic cost of reliably storing microstates (Landauer), hence for fixed Λ there exists $\alpha(\Lambda)$ with

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

B. Isotropy Implies Central Attraction

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

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

yielding 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 $\rho = m\delta$:

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

For $n = 3$, $F(r) \propto m_1 m_2/r^2$, 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)

A. Coding Metric from Second Variation

Extend Φ to histories γ. Consider a localized variation δx^μ; define the local quadratic change as

$$\delta^2\Phi = \frac{1}{2}\, g_{\mu\nu}(x)\, \delta x^\mu \delta x^\nu, \tag{7.1}$$

which *defines* a positive-definite metric on tangent spaces for spacelike displacements (and Lorentzian signature on spacetime histories). Locality of coding implies $g_{\mu\nu}$ depends only on finite neighborhoods; diffeomorphism invariance elevates $g_{\mu\nu}$ to a tensor field.

B. Informational Curvature and Field Equations

Let $\mathcal{I}[g]$ denote the local description functional built from $g_{\mu\nu}$ and its first and second derivatives. Requiring (i) locality, (ii) diffeomorphism invariance, (iii) second-order equations of motion, and (iv) divergence-free field equations selects the Lovelock family. In 3+1 dimensions, the unique such tensor is (up to constants) the Einstein tensor $G_{\mu\nu}$, giving

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}. \tag{7.2}$$

Hence, extremizing description length under local, diffeomorphism-invariant coding leads to GR.

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 inter-particle separation $r(t)$ for two-body runs.

B. Results (6 figures preserved)

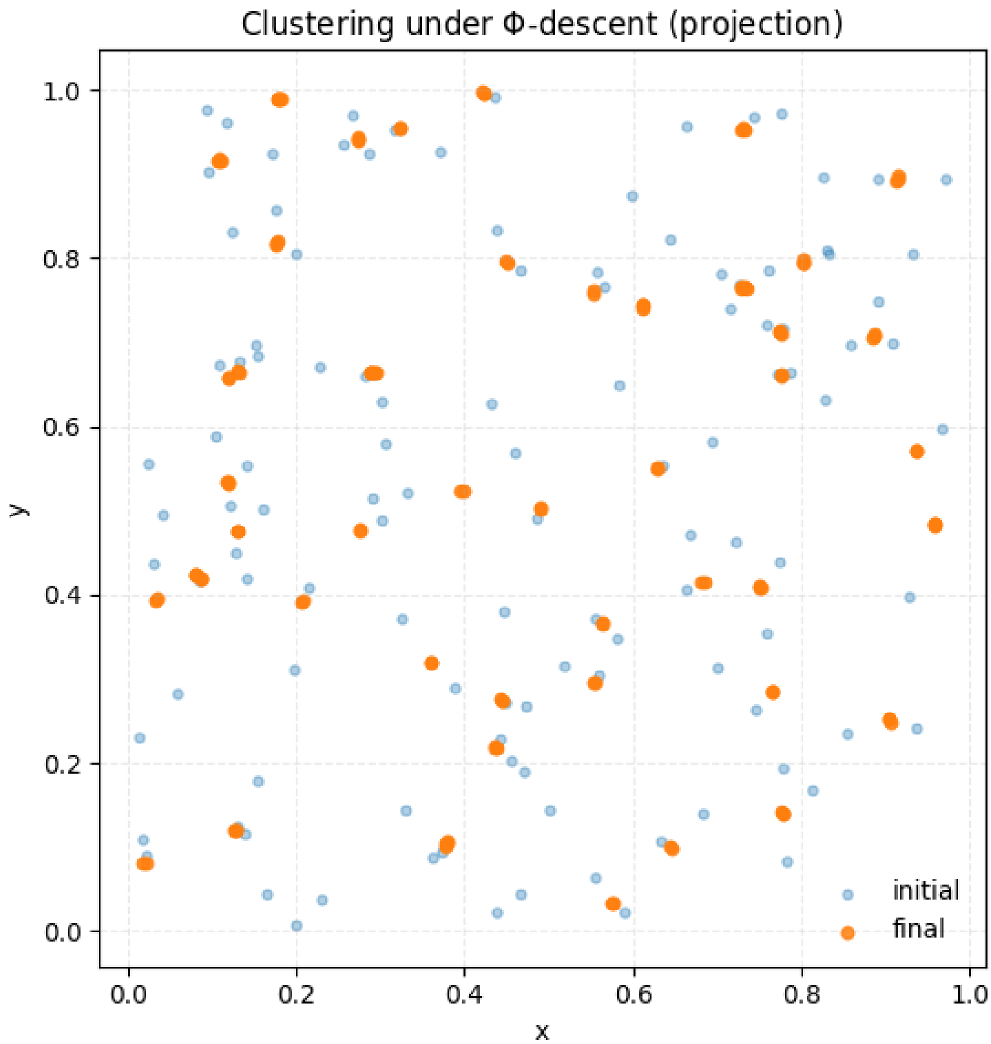

FIG. 1. **Clustering under Φ-descent** (N=120, β=10). Orange: final; blue: initial. No force postulate is used.

Clustering from compression.

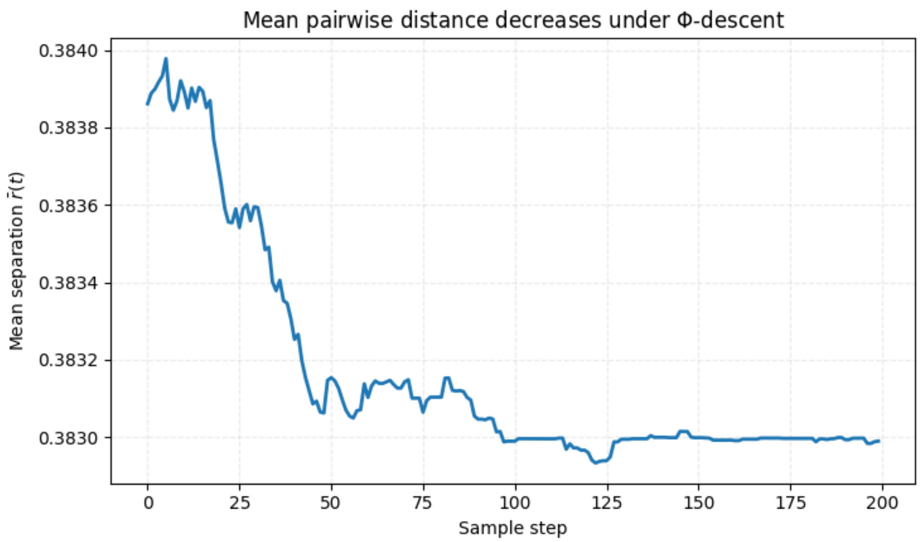

FIG. 2. $\bar{r}(t)$ decreases under Φ-descent with small Metropolis noise.

Mean separation decreases.

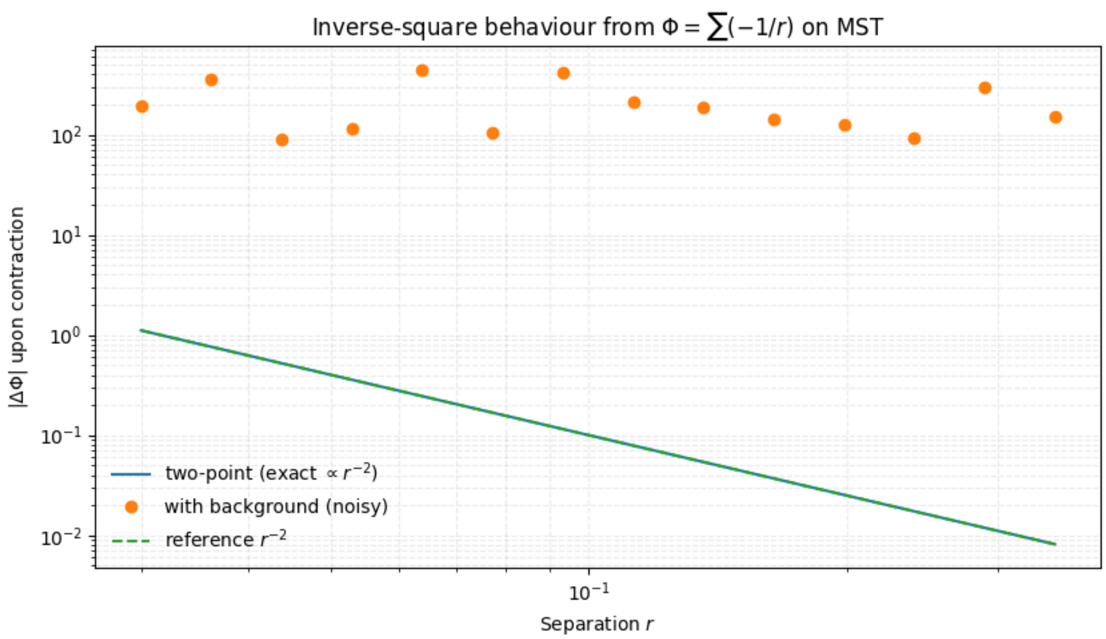

FIG. 3. Change in Φ versus r on log–log axes. Reference r^{-2} dashed; analytic two-point curve (solid) matches; many-body points scatter around this slope.

Inverse-square scaling.

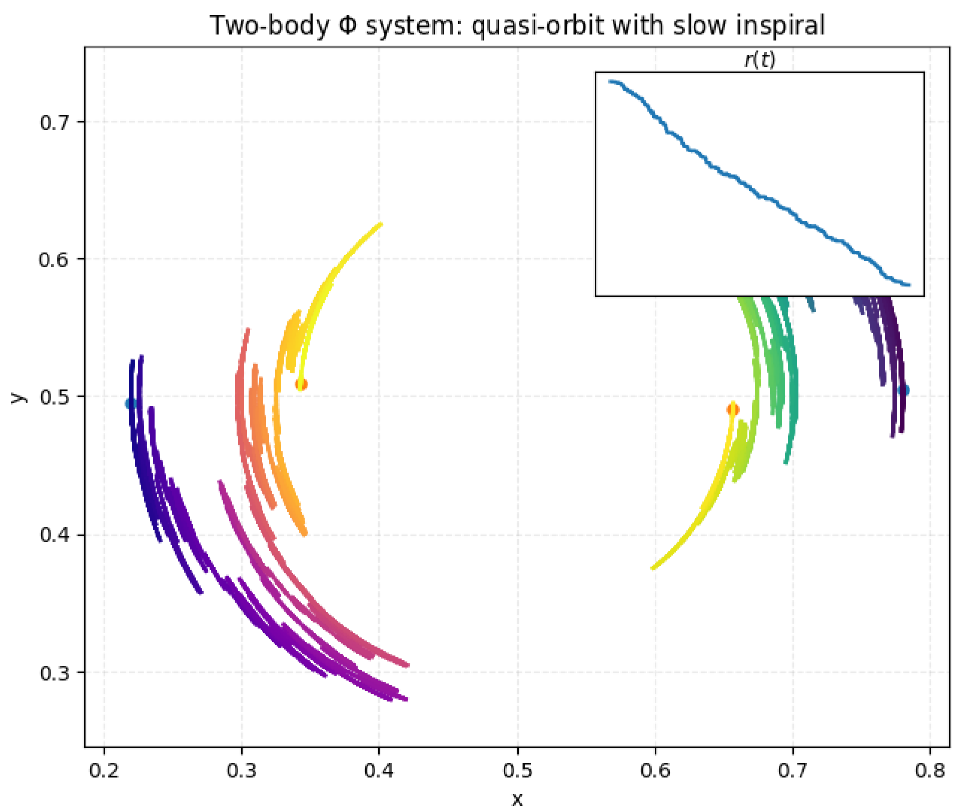

FIG. 4. Two points with tangential proposals show long arcs with intermittent radial-compression events.

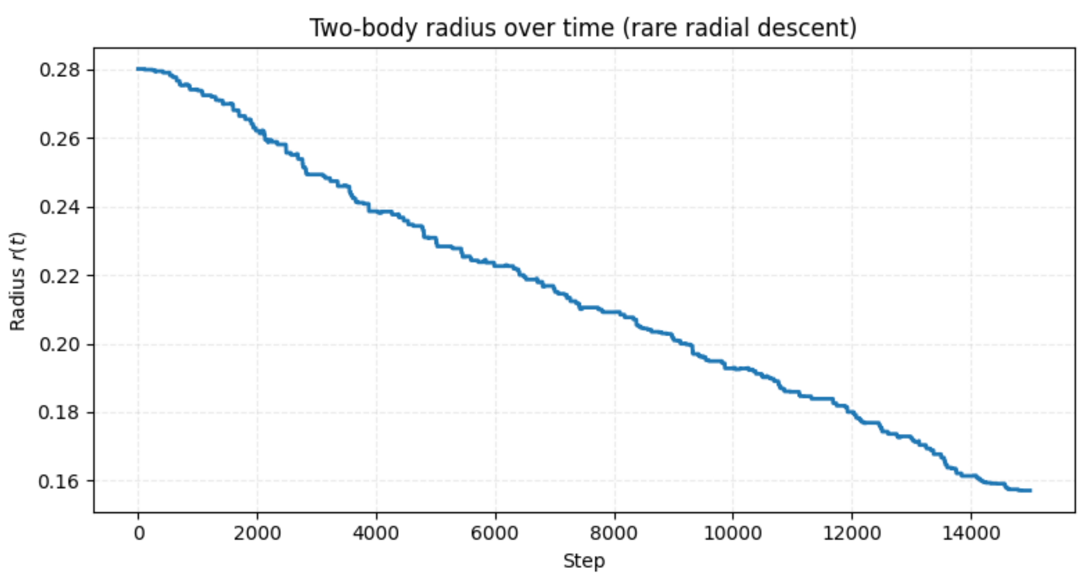

FIG. 5. Staircase decrease of $r(t)$: extended angular motion punctuated by rare accepted radial steps.

Two-body inspiral and quasi-orbit.

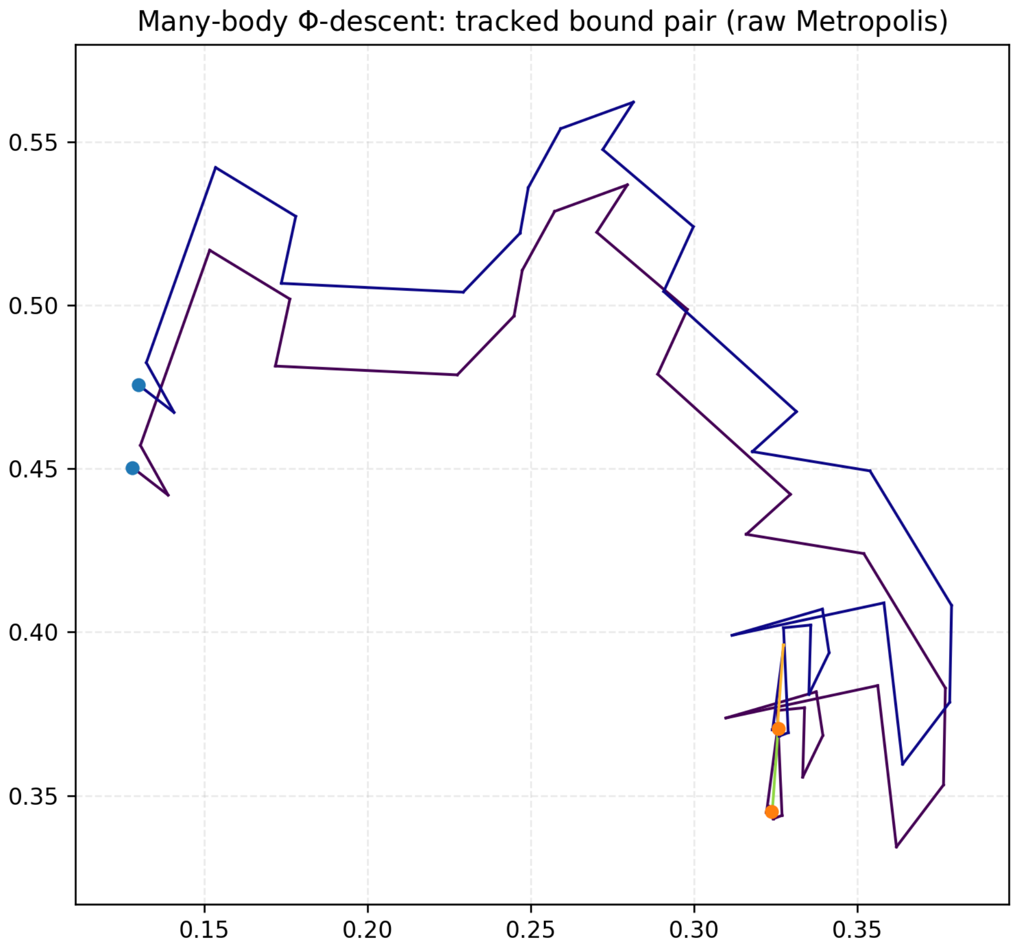

FIG. 6. In N=120 run, the closest pair shows long orbital arcs and intermittent radial descent (start •, end ●).

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, formalized via algorithmic mutual information.

A. Unitary Evolution as Code-Preserving Isometries

Let $\mathcal{H}$ carry an inner product arising from the Kraft inequality normalization of code lengths. Unitary maps are exactly those transformations that preserve total description (isometries of $\mathcal{H}$), i.e., $\psi \mapsto U\psi$ with $U^\dagger U = \mathbb{I}$.

B. Incompatible Codebooks and Non-Commutation

Different observational contexts use incompatible prefix codes (codebooks); simultaneous optimality is generally impossible. This incompatibility is captured by non-commuting operator pairs (A, B) whose code-induced uncertainty obeys $\Delta A\,\Delta B \geq \frac{1}{2}|\langle[A, B]\rangle|$.

C. Subsystems and Entanglement via Algorithmic Mutual Information

For subsystems A, B, define algorithmic mutual information $I_K(A\!:\!B) = K(A)+K(B)-K(A, B)$. Entanglement corresponds to $I_K(A\!:\!B) > 0$. Reduced states minimize Φ subject to constraints on accessible codebooks for the subsystem, matching von Neumann entropy in the typical limit.

D. Measurement as MDL Selection

Let outcomes $\{\phi_k\}$ be branches requiring additional description $\Delta\Phi_k$ to refine ψ to ϕ_k. A universal prior penalizes $\Delta\Phi_k$:

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

Under additivity, composition invariance, and normalization, the branch code length is $-\log|\alpha_k|^2$, yielding Born probabilities $P(\phi_k) = |\alpha_k|^2$ (Appendix B).

10. TEMPORAL COMPRESSION AND THE ORIGIN OF CAUSALITY

A. Emergent Time Parameter

Define a monotone *description time* τ by requiring $\frac{d\Phi}{d\tau} \leq 0$ along realized trajectories. Physical time t is the reparametrization that maximizes predictive compression subject to local conservation constraints, i.e., solves a variational alignment between τ and macroscopic clocks.

B. Causality as Fixed Point of Temporal Compression

Repeated processes enable code reuse over time. The compression functional over histories $\mathcal{C}[x(\cdot)]$ admits fixed-point orderings that are stable under coarse-graining. Causality corresponds to such an ordering; the dynamical law is written with respect to the emergent $t(\tau)$, closing the bootstrap consistently.

11. UNIFIED INTERPRETATION

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

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

12. PREDICTIONS AND FALSIFIABILITY (QUANTIFIED)

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

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

2. **Entanglement-assisted gravity.** Two equal masses m prepared in a maximally entangled spatial state exhibit

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

3. **No particle dark matter.** Disk rotation curves fit an emergent potential term $\psi_{\text{desc}}(R) \propto \ln R$ from code reuse across spiral patterns.

4. **Dark energy evolution.** $w(z) = -1 + \delta w(z)$ with $\delta w \lesssim 0.05$ tracking structure growth.

5. **Statistical time symmetry breaking.** In low-$\nabla\Phi$ systems, fluctuation-theorem diagnostics show an excess reversal factor $1 + \xi$, $\xi \sim 10^{-3}$.

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 Born 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. *Note:* To test estimator universality, future work will include alternative graph encodings (Delaunay, kNN), dictionary compressors (LZ), and learned compressors.

Appendix D: Responses to Common Objections (Rebuttal Appendix)

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

Entropy vs. description length. $\mathbb{E}[K] = H + O(1)$ and $S = k \ln 2 \cdot H$. Global description $\Phi_{\text{tot}} = L(M) + L(D|M)$ decreases while subsystem entropy S may increase; exported entropy pays Landauer cost, reconciling signs (Sec. 3 A).

Mass and information. $\rho \propto \rho_m$ is operational: microstate multiplicity implies local entropy density; mass measures energetic stability of storage. Attraction follows from isotropy and flux conservation.

Geometry from description. Second variation defines a coding metric; locality and diffeomorphism invariance select Einstein dynamics in 3+1 via Lovelock.

Quantum formalism. Unitary evolution is code-preserving; incompatible codebooks encode non-commutation; entanglement is algorithmic mutual information; MDL selection yields Born rule.

Appendix E: Spatial Dimensionality from Compression and Locality (Discussion)

We seek dimensions n for which: (i) local, isotropic, scale-free kernels $k(r)$ exist with conserved flux; (ii) harmonic Green's functions yield finite-energy bound structures; (iii) compression flux is additive under partition. These criteria select $k'(r) \propto r^{-(n-1)}$. For $n = 1, 2$, global structures are unstable or trivial; for $n \geq 4$, scale-free kernels do not simultaneously support stable bound sets and finite local flux. In $n = 3$, $k(r) = 1/r$ is harmonic outside sources, supports stable flux, and maximizes compression consistency. *Proposition (heuristic).* Under (i)–(iii), the minimal-dimension solution supporting nontrivial compressive structure with finite local flux is $n = 3$.

Appendix F: Gradient Consistency for Universal MDL Estimators (Details)

Setup. Let $(X, \mathcal{B}, \mu)$ be a configuration manifold with Borel measure μ absolutely continuous w.r.t. Lebesgue on charts. Define the class $\mathcal{U}$ of *admissible estimators* $\widehat{\Phi}$: prefix-free, local, universal (there exists c s.t. $\widehat{\Phi}(x) \leq \Phi(x) + c$ for all x), and *refinement-stable* (code updates are supported on finite neighborhoods).

Theorem (Gradient Consistency). For any $\widehat{\Phi} \in \mathcal{U}$ and μ-a.e. $x \in X$, the directional derivative along any v in a full-measure cone $\mathcal{C}_x$ satisfies

$$\lim_{h \to 0^+} \frac{\widehat{\Phi}(x + hv) - \widehat{\Phi}(x)}{h} = \lim_{h \to 0^+} \frac{\Phi(x + hv) - \Phi(x)}{h}.$$

Sketch. (1) Universality implies $|\widehat{\Phi} - \Phi| \leq c$. (2) Locality and refinement-stability bound the number of code changes under infinitesimal displacements. (3) Discontinuities of K lie in a set of μ-measure zero; restrict to typical x. (4) The symmetric difference in codebooks forms a vanishing set under $h \to 0^+$, yielding equality of directional derivatives almost everywhere. □

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] S.-I. Amari, *Information Geometry and Its Applications* (Springer, 2016).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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