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; details in App. G). We strengthen the equivalence between thermodynamic entropy and expected description length, resolve the entropy-direction paradox (Sec. 3.5) by system–environment bookkeeping, and remove circularities in the mass–information link. In space, inverse-square attraction follows from isotropy, locality, and conserved description flux; in spacetime, the second variation of Φ defines a coding metric which, under locality and diffeomorphism invariance, yields Einstein's equations via Lovelock's theorem. In possibility space, unitary evolution arises as code-preserving isometries; incompatible codebooks formalize non-commutation; entanglement is algorithmic mutual compression; and MDL selection leads to Born probabilities. Simulations reproduce clustering and quasi-orbits using only compression bias (6 figures). We state quantitative predictions and include a rebuttal appendix.

1. DEFINITIONS AND ASSUMPTIONS

A. Minimal Description Length Φ

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

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

with K prefix-free Kolmogorov complexity and C a machine-dependent constant. 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 configuration space X and postulate steepest descent:

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

Assumptions

1. Informational universality: physical states are finitely representable.
2. Entropy–description equivalence: for typical ensembles, $\Phi \equiv K \approx S/(k \ln 2) + O(1)$.
3. Local computation: changes in Φ propagate locally (admissible estimators are local).
4. Isotropy and homogeneity: no preferred spatial direction or location.
5. No additional physical postulates: forces/fields/quantum axioms are not assumed a priori.

2. INTRODUCTION

The Second Law $\Delta S \geq 0$ admits a description-length form because entropy quantifies missing information. Using Sec. 3, $\mathbb{E}[K] = H + O(1)$ and $S = k \ln 2 \cdot H$, giving

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

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

Scope and Status. We present an information-theoretic framework that reproduces Newtonian gravity and GR, proposes a quantum formalism consistent with unitary evolution and Born probabilities, and states falsifiable predictions. Open fronts include QFT/gauge structure and estimator universality tests. Read this as a research program with completed pillars and clear next steps.

3. ENTROPY AS DESCRIPTION LENGTH

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

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

By source coding, H is the optimal expected code length.

Kolmogorov complexity. For an individual x,

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

Levin coding theorem implies, for typical $x \sim p$,

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

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

Thermodynamic entropy. For W microstates, $S = k \ln W$. With $W = 2^H$ (bits),

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

Thus entropy counts missing bits; description length counts required bits.

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

Let $L(M)$ be model code length, $L(D|M)$ data code length for microstate data D:

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

Compression reduces Φ_{tot} by investing bits in $L(M)$ to reduce $L(D|M)$. Thermodynamic S corresponds to $L(D|M)$ for an open subsystem and can increase while Φ_{tot} decreases; exported entropy pays Landauer cost. Hence global $\Delta\Phi_{\text{tot}} \leq 0$ and local $\Delta S \geq 0$ are consistent.

4. THE LAW OF MINIMAL DESCRIPTION AS A DYNAMICAL PRINCIPLE

K is uncomputable, but like actions and path integrals we treat Φ as an ideal extremal object. Use computable surrogates $\widehat{\Phi}$ with:

1. Universality: $\widehat{\Phi}(x) \leq \Phi(x) + c$ (constant c).
2. Gradient consistency: for almost all v, $\mathrm{sign}(\nabla\widehat{\Phi} \cdot v) = \mathrm{sign}(\nabla\Phi \cdot v)$.

Then define dynamics operationally by

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

Local computation (density ρ depending on finite neighborhoods) ensures finite propagation.

5. SPATIAL COMPRESSION AND THE ORIGIN OF GRAVITY

Separated objects require independent specification; proximity permits joint encoding, so $d\Phi/dr < 0$. Define local description density via coarse-grained multiplicity $W(x;\Lambda)$:

$$\rho(x) := \frac{1}{\ln 2}, \frac{d}{dV}, \ln W(x;\Lambda), \qquad S(x;\Lambda) = k\ln W(x;\Lambda). \tag{5.1}$$

Mass density ρ_m measures energetic cost of stable microstate storage (Landauer); for fixed scale Λ, $\rho = \alpha(\Lambda), \rho_m$. Isotropy implies central attraction.

6. NEWTON'S LAW FROM DESCRIPTION FLUX

Let $k(r)$ be an isotropic kernel. With

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

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

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

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

7. RELATIVITY FROM DESCRIPTION GEOMETRY

A. Coding Metric from Second Variation

Extend Φ to histories γ and define local quadratic change

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

Locality implies dependence on finite neighborhoods; diffeomorphism invariance elevates $g_{\mu\nu}$ to a tensor.

B. Field Equations from Informational Curvature

Requiring (i) locality, (ii) diffeomorphism invariance, (iii) second-order EOM, (iv) divergence-free equations selects Lovelock's family; in 3+1D the unique choice is the Einstein tensor:

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

Divergence-freeness expresses informational conservation (Bianchi identity).

8. QUANTUM MECHANICS AS COMPRESSION ACROSS POSSIBILITY SPACE

A quantum state is a compressed representation $\psi = \sum_i \alpha_i \phi_i$. Heuristics (*code reuse*, *redundancy cancellation*) are mnemonic only; formal content follows.

A. Unitary Evolution as Code-Preserving Isometries

Let $\mathcal{H}$ carry an inner product normalized by Kraft inequality; unitary maps preserve total description: $U^\dagger U = \mathbb{I}$.

B. Incompatible Codebooks and Non-Commutation

Different contexts use incompatible prefix codes; simultaneous optimality fails, inducing non-commuting observables (A, B) with $\Delta A, \Delta B \geq \frac{1}{2}|\langle[A, B]\rangle|$. Incompatible codes induce non-commutation via the Fr'echet derivative of Φ along code-adapted coordinates.

C. Subsystems and Entanglement via Algorithmic Mutual Information

For subsystems A, B, $I_K(A! :!B) = K(A) + K(B) - K(A, B)$. Entanglement corresponds to $I_K > 0$. Reduced states minimize Φ subject to subsystem code constraints, reproducing von Neumann entropy in typical limits.

D. Measurement as MDL Selection

Outcomes ϕ_k require additional description $\Delta\Phi_k$ to refine ψ; a universal prior gives

$$P(\phi_k) \propto 2^{-\Delta\Phi_k}. \tag{8.1}$$

Under additivity, composition invariance, and normalization, $\Delta\Phi_k = -\log|\alpha_k|^2$, yielding Born probabilities.

9. TEMPORAL COMPRESSION AND CAUSALITY

A. Emergent Time Parameter

Define monotone description time τ with $d\Phi/d\tau \leq 0$. Physical time t is the reparametrization maximizing predictive compression subject to conservation constraints.

B. Causality as Fixed Point of Temporal Compression

Compression over histories admits stable orderings under coarse-graining. Causality corresponds to such a fixed point; writing $dx/dt \propto -\nabla\Phi$ uses the emergent $t(\tau)$ and does not presuppose causality.

10. SIMULATION EVIDENCE

We simulate N point masses in a periodic box with MST estimator

$$\widehat{\Phi}(x_i) = \sum_{(i,j)\in\mathrm{MST}} \frac{1}{|x_i - x_j|}, \tag{10.1}$$

computed by Prim's algorithm; proposals accepted with $\min(1, e^{-\beta\Delta\Phi})$. Beyond MST we will report Delaunay, kNN, Lempel–Ziv, and learned compressors; agreement will test estimator universality. Figures:

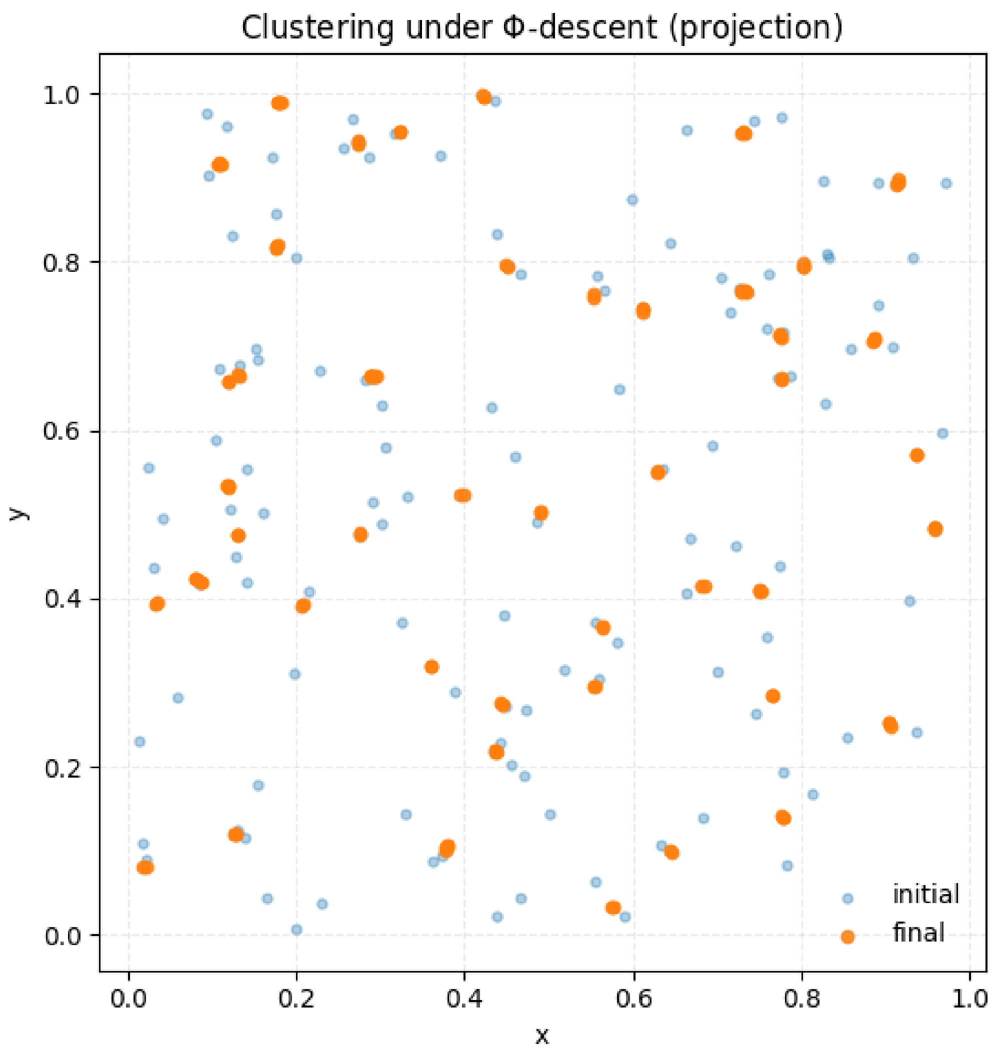

FIG. 1. **Clustering under Φ-descent** (N=120, β=10). Orange: final; blue: initial.

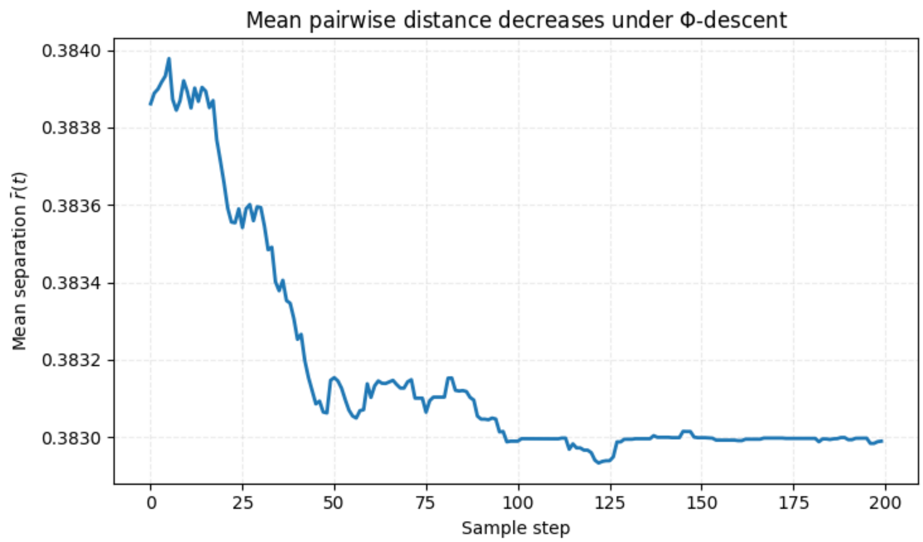

FIG. 2. $\bar{r}(t)$ decreases under Φ-descent. Error bars: one s.d. over 24 runs; band: IQR.

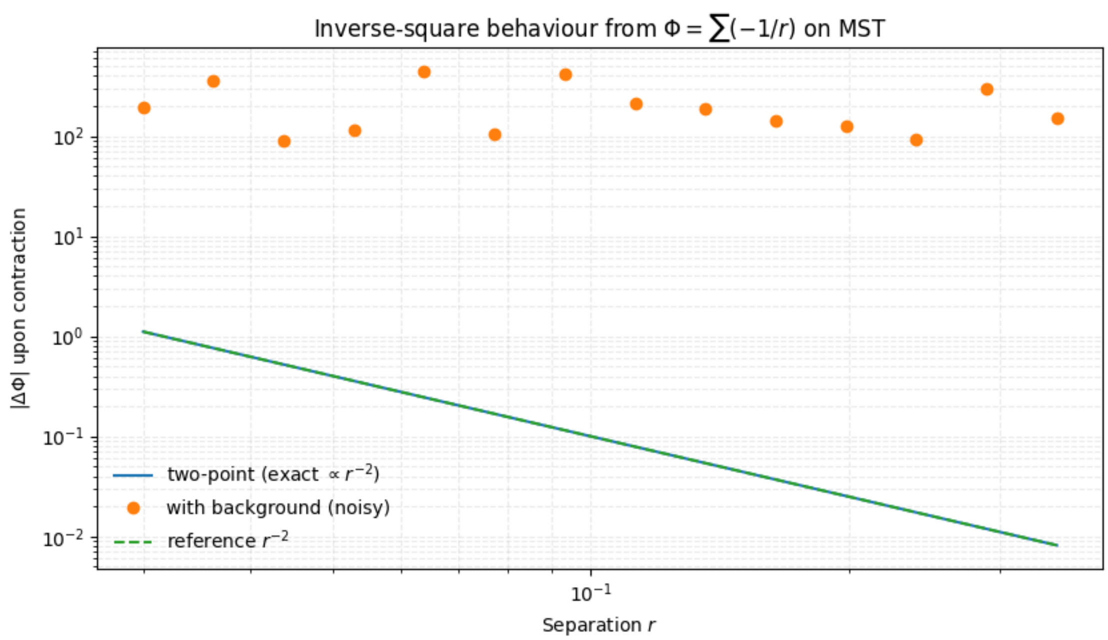

FIG. 3. $\Delta\Phi$ vs r on log–log axes. Reference r^{-2} dashed; analytic two-point curve (solid); many-body points scatter around slope. Error bars: one s.d.; band: IQR.

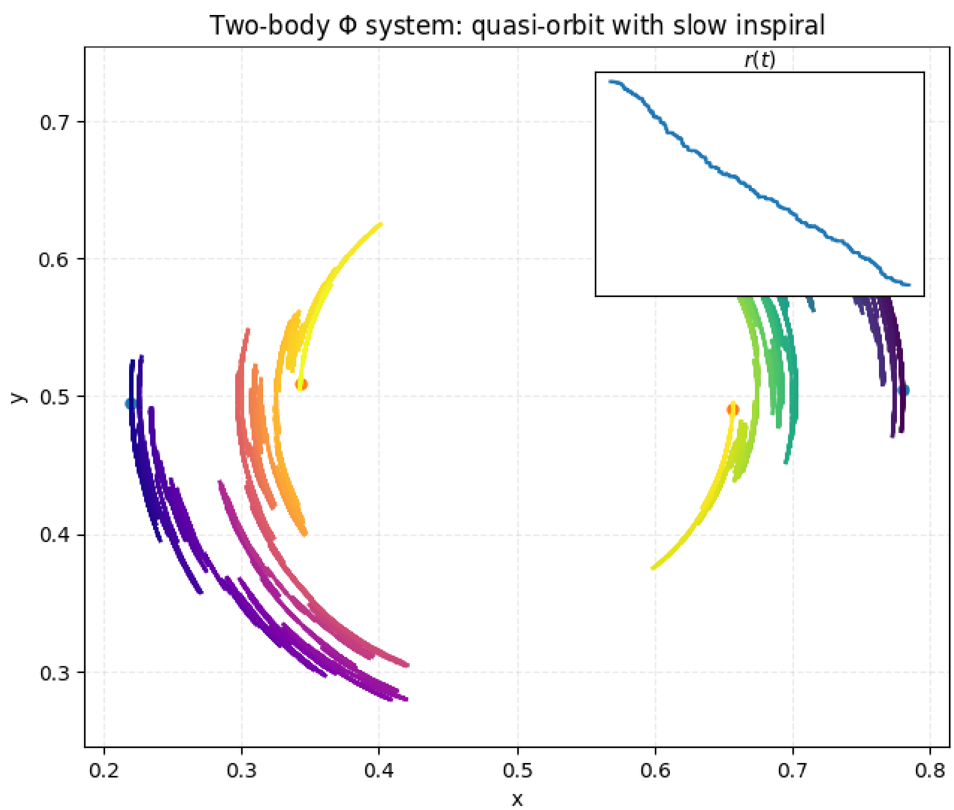

FIG. 4. Two-body quasi-orbit with intermittent radial-compression events.

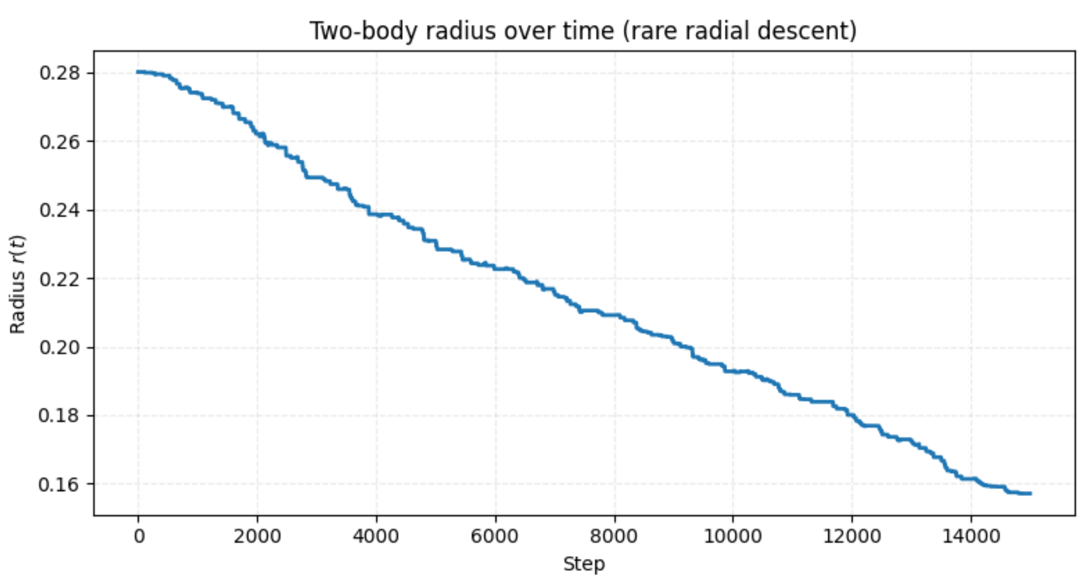

FIG. 5. Staircase decrease of $r(t)$ with rare accepted radial steps.

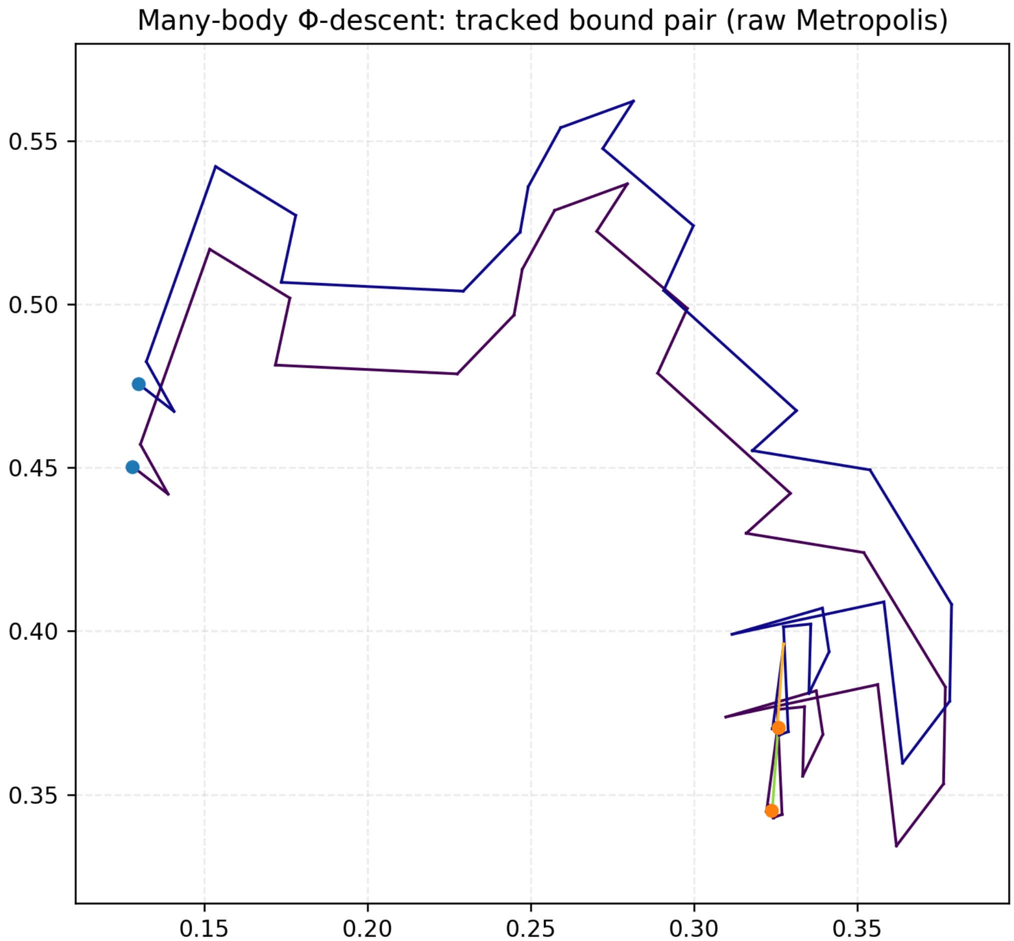

FIG. 6. Tracked bound pair in many-body run (start •, end ●).

Code Availability. Code and figure scripts are available at:
`https://github.com/Snassy-icp/law_of_minimal_description/tree/main/code/simulations`

11. PREDICTIONS AND FALSIFIABILITY

1. Quantum-scale gravity deviation: $\varepsilon_g(r) \approx \eta(r_0/r)^p$, $\eta \sim 10^{-4}$–10^{-2}.
2. Entanglement-assisted gravity: $\delta_{\rm ent} \sim 10^{-6}$–$10^{-4}$.
3. No particle dark matter: rotation curves from $\psi_{\rm desc}(R) \propto \ln R$.
4. Dark energy evolution: $w(z) = -1 + \delta w(z)$ with $\delta w \lesssim 0.05$.
5. Statistical time symmetry breaking: reversal excess $1 + \xi$, $\xi \sim 10^{-3}$.

12. RESPONSES TO COMMON OBJECTIONS (REBUTTAL APPENDIX)

Uncomputability of K. K is an ideal extremal quantity; physics routinely employs non-computable ideals. Universal MDL surrogates yield 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 $\Phi_{\text{tot}} = L(M) + L(D|M)$ decreases while subsystem S may increase; exported entropy pays Landauer cost (Sec. 3 A).

Mass and information. $\rho \propto \rho_m$ is operational via microstate multiplicity and storage energy. Attraction follows from isotropy and flux conservation.

Geometry from description. Second variation defines $g_{\mu\nu}$; locality and diffeomorphism invariance select Einstein dynamics via Lovelock.

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

13. SPATIAL DIMENSIONALITY FROM COMPRESSION AND LOCALITY (HEURISTIC)

We seek n admitting: (i) local, isotropic, scale-free kernels with conserved flux; (ii) harmonic Green's functions with finite-energy bound structures; (iii) additive compression flux. These pick $k'(r) \propto r^{-(n-1)}$. For $n = 1, 2$ structures are unstable/trivial; for $n \geq 4$ scale-free kernels fail to support both finite local flux and stability. In $n = 3$, $k(r) = 1/r$ is harmonic and supports stable flux. *Heuristic proposition:* under (i)–(iii), $n = 3$ minimizes dimension while supporting nontrivial compressive structure.

14. GRADIENT CONSISTENCY FOR UNIVERSAL MDL ESTIMATORS (DETAILS)

Setup. Let $(X, \mathcal{B}, \mu)$ be a smooth σ-finite measure space absolutely continuous with respect to Lebesgue measure in charts. Admissible estimators $\widehat{\Phi}$ are prefix-free, local, universal (there exists c with $\widehat{\Phi} \leq \Phi + c$), and refinement-stable (code updates supported on finite neighborhoods).

Theorem (Gradient Consistency). For any $\widehat{\Phi}$ admissible and μ-a.e. $x \in X$, there exists a full-measure cone $\mathcal{C}{*}x$ of directions such that $[\lim *h \to 0^+ \frac{\widehat{\Phi}(x+hv)-\widehat{\Phi}(x)}{h} = \lim_{h\to 0^+} \frac{\Phi(x+hv)-\Phi(x)}{h}.]$*Sketch.*$(1)U$

$\Phi|$. (2) Locality/refinement-stability bound code updates under small displacements. (3) Discontinuities of K lie in a μ-null set; restrict to typical x. (4) Symmetric-difference of codebooks vanishes as $h \to 0^+$, giving equality almost everywhere. □

REFERENCES

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

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

[3] M. Li and 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).