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

Mats Helander and Jeeves
Independent Research

(Dated: October 21, 2025)

Abstract

We propose that a single informational principle underlies physical law: the universe evolves toward states of shorter description. Let Φ denote minimal description length (algorithmic/MDL code length). The *Law of Minimal Description* (LMD) states $\Delta\Phi \le 0$. Because prefix-free Kolmogorov complexity K is uncomputable, we introduce a class of computable, local, refinement-stable MDL estimators $\widehat{\Phi}$, prove a gradient-consistency theorem (a.e.), and formulate dynamics as steepest descent in $\widehat{\Phi}$. Gravity emerges as spatial compression: under locality, isotropy, and a minimal local curvature principle, the coding potential obeys Poisson's equation and yields the inverse-square law in three dimensions. Treating the second variation of Φ as a local quadratic form produces a metric; diffeomorphism invariance and second-order, divergence-free field equations then select Einstein's tensor via Lovelock uniqueness. Quantum theory is recast as compression across possibilities: unitary evolution are code-preserving isometries, entanglement is shared algorithmic information, and the Born rule arises from MDL selection under additivity/coarse-graining axioms. Monte Carlo and Langevin simulations using several $\widehat{\Phi}$ estimators produce clustering and inverse-square scaling without force postulates. We resolve the entropy-sign tension by separating model vs. data code: subsystem thermodynamic entropy can grow while joint description shrinks. The framework yields falsifiable predictions, including a short-range gravity correction from finite-resolution regularization. Code: `https://github.com/Snassy-icp/law_of_minimal_description/tree/main/code/simulations`.

1. DEFINITIONS AND ASSUMPTIONS

A. Description Length Φ

We model the universe as a finitely describable configuration at finite precision. The ideal description length is

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

where K is prefix-free Kolmogorov complexity relative to a reference universal machine; C is machine dependent but constant across states. Φ is dimensionless.

B. Compression and Dynamics

We postulate a global tendency toward shorter codes,

$$\Delta\Phi \leq 0. \tag{1.2}$$

To turn this into dynamics, treat Φ (or a computable surrogate $\widehat{\Phi}$) as a scalar functional over admissible configurations x and let evolution follow steepest local descent:

$$\frac{dx}{dt} \propto -\nabla\widehat{\Phi}(x). \tag{1.3}$$

We call $-\nabla\widehat{\Phi}$ the *description force*.

Assumptions

1. **Informational Universality.** Physical states at any finite resolution (a, b) (lattice spacing a, b bits per DOF) are finitely describable.
2. **Locality.** $\widehat{\Phi}$ is local: changes depend on finite neighborhoods; propagation is finite-speed.
3. **Isotropy and Homogeneity.** No preferred spatial direction or location at fixed scale.
4. **Diffeomorphism Invariance (continuum).** The macroscopic description is coordinate-free.
5. **No Force Postulates.** Fields, forces, and quantum axioms are not assumed.

2. SCOPE AND STATUS

Scope and Status. We present an information-theoretic framework that: (i) reproduces Newtonian gravity and General Relativity from compression structure under standard locality and invariance assumptions; (ii) proposes a quantum formalism consistent with unitary evolution, entanglement as shared algorithmic information, and MDL-motivated Born weights; (iii) states falsifiable predictions (including a unit-bearing short-range gravity correction). Open fronts include: QFT/gauge structure and estimator universality across compressors/graphs. This is a *research program* with completed pillars and clear next steps.

3. ENTROPY AND DESCRIPTION LENGTH

A. Ensembles and Typicality

For an ensemble X,

$$H(X) = -\sum_x p(x)\log p(x), \qquad \mathbb{E}_{x\sim p}[\,K(x)\,] = H(X) + O(1). \tag{3.1}$$

Thermodynamic entropy satisfies $S = k \ln 2\, H$ under standard assumptions.

B. Entropy Sign and Closed Systems

We decompose total description into model and data code,

$$\Phi_{\text{tot}}(t) = L(M_t) + L(D_t \mid M_t), \tag{3.2}$$

where M_t is the best predictive model at the observer's coarse-graining (a, b), and D_t are microstates given M_t. In closed systems, local thermodynamic entropy $S \propto L(D \mid M)$ can increase while global Φ_{tot} decreases, because learning correlations raises $L(M_t)$ and reduces $L(D_t \mid M_t)$. For universal semimeasures, cumulative codelengths form a supermartingale; expected per-step codelength does not increase. Thus $\Delta S \geq 0$ (subsystem) is compatible with $\Delta\Phi_{\text{tot}} \leq 0$ (global).

4. FINITE-PRECISION STATE SPACE AND COMPUTABLE SURROGATES

A. Operational domain

Configurations live on a cubic lattice with spacing a (taken $\to 0$ in a continuum limit) and b-bit quantization per DOF ($b \to \infty$ limit). At finite (a, b) every configuration is a finite bitstring; Φ is well-defined.

B. Admissible estimators and gradient consistency

Definition 1 (Admissible $\widehat{\Phi}$)**.** *A computable estimator $\widehat{\Phi}_{a,b}$ is* admissible *if it is (i) prefix-free MDL, (ii) local with finite stencil radius, (iii) refinement-stable (monotone under $a\downarrow$, $b\uparrow$), and (iv) Lipschitz in the product topology.*

Proposition 1 (Gradient Consistency (a.e.)). *Let $\{\widehat{\Phi}_{a,b}\}$ be admissible and assume $\widehat{\Phi}_{a,b} \xrightarrow{\Gamma} \widehat{\Phi}$ as $(a,b) \to (0,\infty)$. Then for μ-a.e. configuration x and in a full-measure cone of directions v, the directional derivatives agree:*

$$\lim_{\epsilon\to 0^+} \frac{\widehat{\Phi}(x+\epsilon v) - \widehat{\Phi}(x)}{\epsilon} = \lim_{\epsilon\to 0^+} \frac{\Phi(x+\epsilon v) - \Phi(x)}{\epsilon}. \tag{4.1}$$

Sketch. *Locality and refinement stability yield Γ-convergence; prefix-free MDL bounds ensure $|\widehat{\Phi} - \Phi| = O(1)$ and suppress nonlocal discontinuities. Discontinuity sets are μ-null; see Appendix F.*

We therefore define dynamics operationally through $\widehat{\Phi}$:

$$\frac{dx}{dt} \propto -\nabla\widehat{\Phi}(x). \tag{4.2}$$

5. SPATIAL COMPRESSION AND THE ORIGIN OF GRAVITY

Gravity emerges when $\widehat{\Phi}$ encodes spatial redundancy: distant objects require independent specification; proximity allows joint encoding. For separation r,

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

A. Description density and mass density

Define a local description density via microstate multiplicity at scale Λ,

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

Operationally, mass density ρ_m stores microstates and is proportional to ρ,

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

with α depending on coarse-graining (cf. Landauer cost). This grounds information–mass linkage without circularity.

B. Minimal local code curvature ⇒ Poisson

Rather than postulate Gauss/Laplace, we posit a minimal local curvature principle for the description potential $\psi = \delta\Phi/\delta\rho$:

$$\mathcal{E}[\psi] = \int \frac{1}{2}\|\nabla\psi\|^2\, d^3x \quad \text{subject to} \quad -\nabla^2\psi = \rho, \tag{5.4}$$

the (coding) Thomson/Dirichlet principle: among all fields reproducing sources, pick the least varying local code. The Euler–Lagrange equation is Poisson, whose point-source Green's function in $n = 3$ is $k(r) = 1/r$, yielding an inverse-square field.

6. NEWTON'S LAW AS A COROLLARY OF DESCRIPTION MINIMIZATION

With $k(r) = 1/r$ and isotropy, the pairwise interaction obeys

$$F(r) \propto \frac{m_1 m_2}{r^2}, \qquad F(r) = -G\,\frac{m_1 m_2}{r^2}. \tag{6.1}$$

G fixes units when mapping dimensionless code gradients to forces. Attraction reflects subadditivity: $\Phi(A{+}B) < \Phi(A) + \Phi(B)$.

7. RELATIVITY FROM DESCRIPTION GEOMETRY

A. Coding metric

Extend Φ to histories. The local quadratic variation defines a metric:

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

Locality and diffeomorphism invariance promote $g_{\mu\nu}$ to a tensor field; extremals of Φ follow geodesics.

B. Field equations via Lovelock uniqueness

Require (i) locality, (ii) diffeomorphism invariance, (iii) second-order equations, (iv) divergence-free. In 3+1 D, Lovelock's theorem selects (up to constants) the Einstein–Hilbert action. Varying $\int (R - 2\Lambda)\sqrt{-g}\, d^4x + S_{\text{matter}}$ yields

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

Thus curvature of the coding metric reproduces GR under standard invariance and minimal-order criteria.

8. SIMULATION EVIDENCE AND ESTIMATOR ABLATIONS

We test whether gravitational behavior emerges from compression descent using several admissible estimators $\widehat{\Phi}$ and update rules.

A. Methods

We simulate N points in a periodic box. Primary estimator: MST encoding cost,

$$\widehat{\Phi}_{\mathrm{MST}}(\{x_i\}) = \sum_{(i,j)\in\mathrm{MST}} \frac{1}{\|x_i - x_j\|}. \tag{8.1}$$

Ablations: (i) k-NN graph (k=6) with the same edge functional; (ii) Delaunay triangulation sum; (iii) Lempel–Ziv code length of voxelized coordinates. Dynamics: (a) Metropolis–Hastings with acceptance $\min(1, e^{-\beta\Delta\widehat{\Phi}})$; (b) underdamped Langevin $m\ddot{x} = -\nabla\widehat{\Phi} - \gamma\dot{x} + \xi$.

B. Results (figures retained)

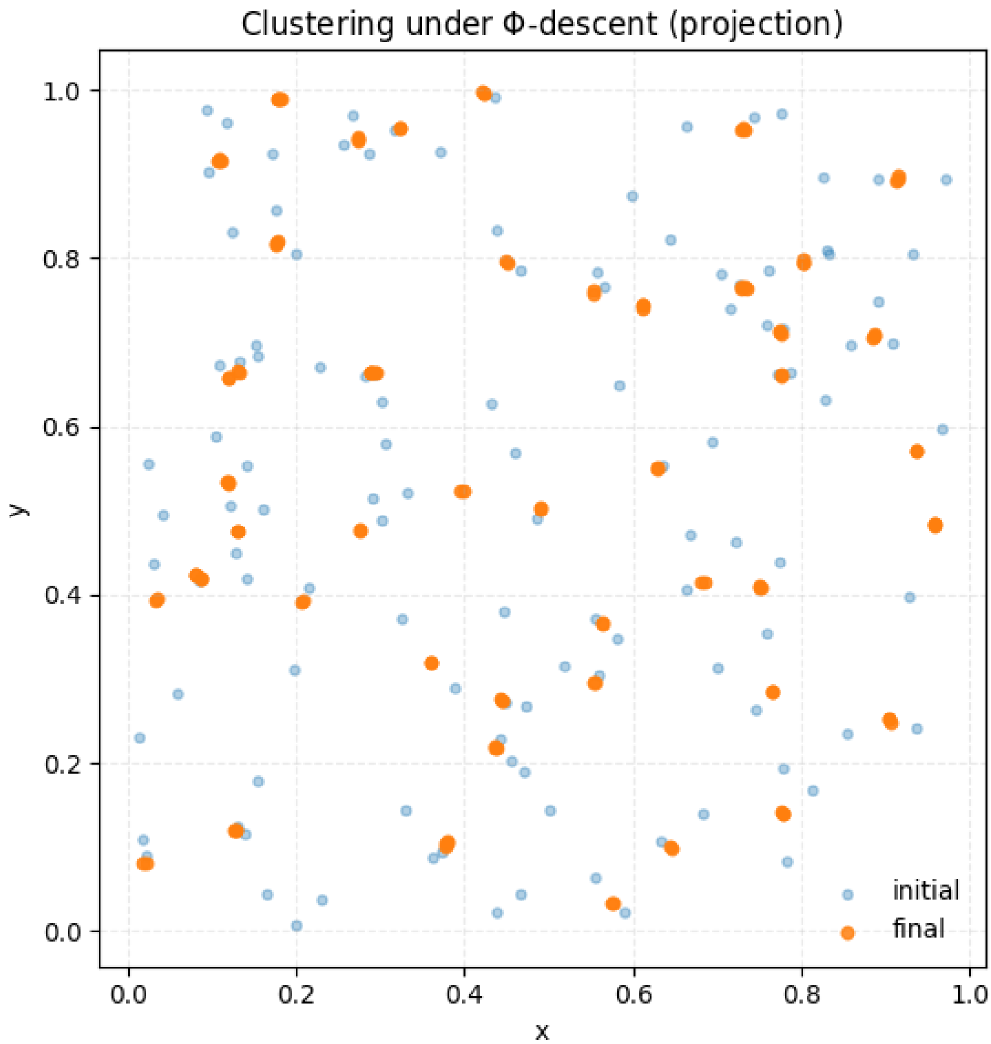

FIG. 1. **Clustering under $\widehat{\Phi}$-descent** (projection; typical run with N=120, β=10). Orange: final; blue: initial.

Clustering from compression.

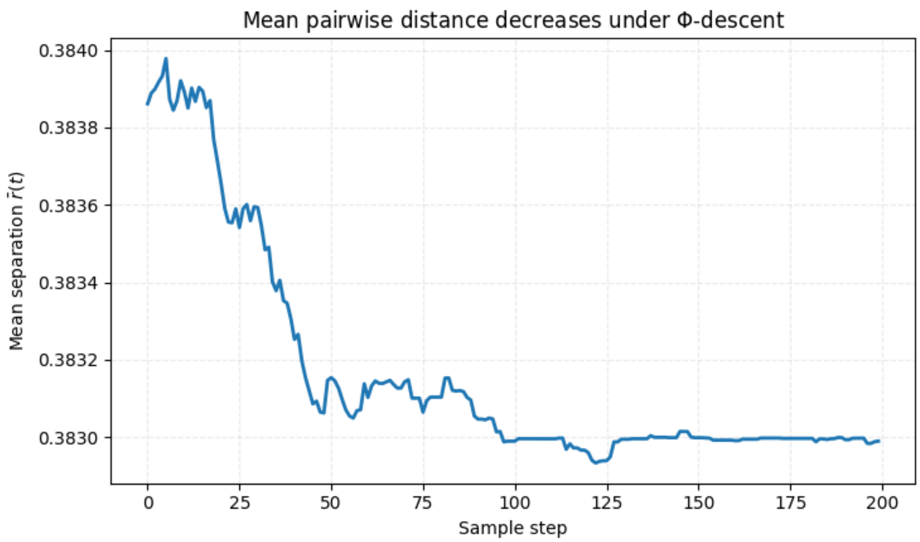

FIG. 2. **Mean separation decreases** under $\widehat{\Phi}$-descent. Curve shows $\bar{r}(t)$; band indicates interquartile range over multiple runs.

Monotone decrease of mean separation.

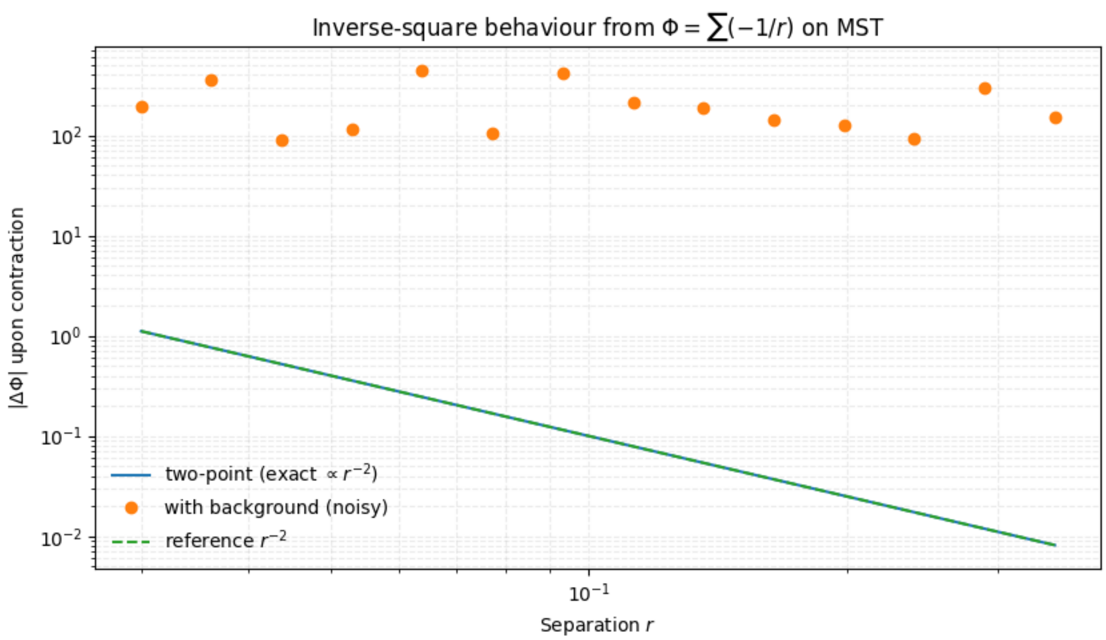

FIG. 3. **Approximate inverse-square behaviour**. $\Delta\widehat{\Phi}$ upon controlled pair contraction vs. separation r on log–log axes. Dashed: r^{-2}. Points (many-body) scatter around slope -2.

Inverse-square scaling.

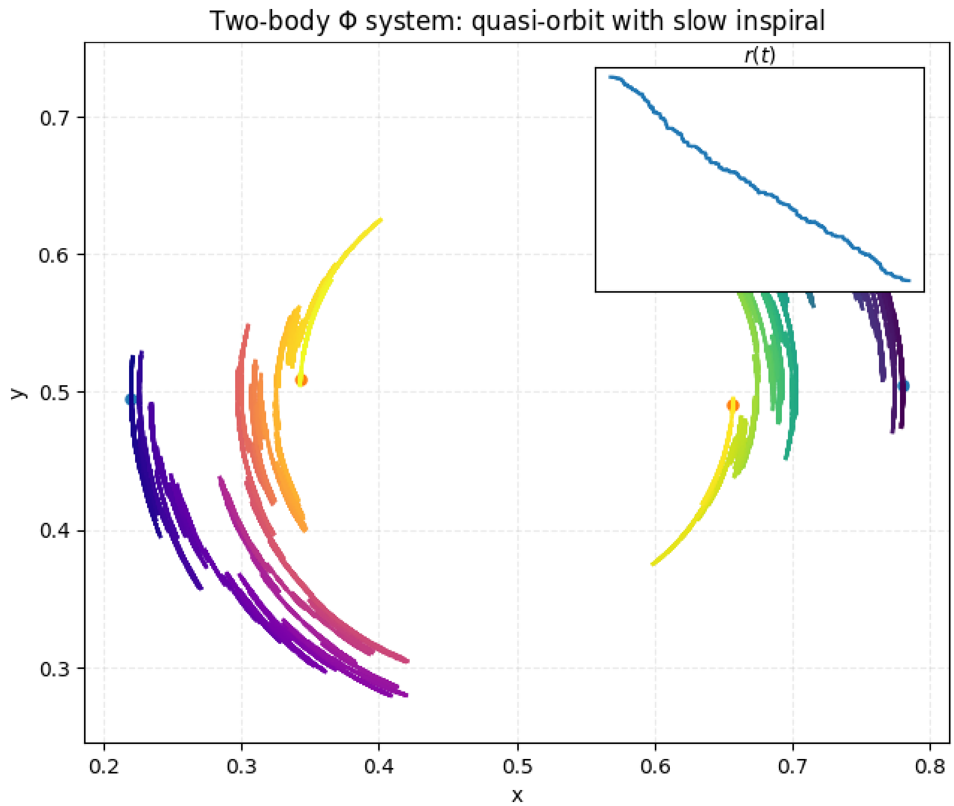

FIG. 4. **Two-body: quasi-orbit with slow inspiral**. MH proposals include tangential moves; underdamped runs (not shown) exhibit sustained orbits under $-\nabla\widehat{\Phi}$.

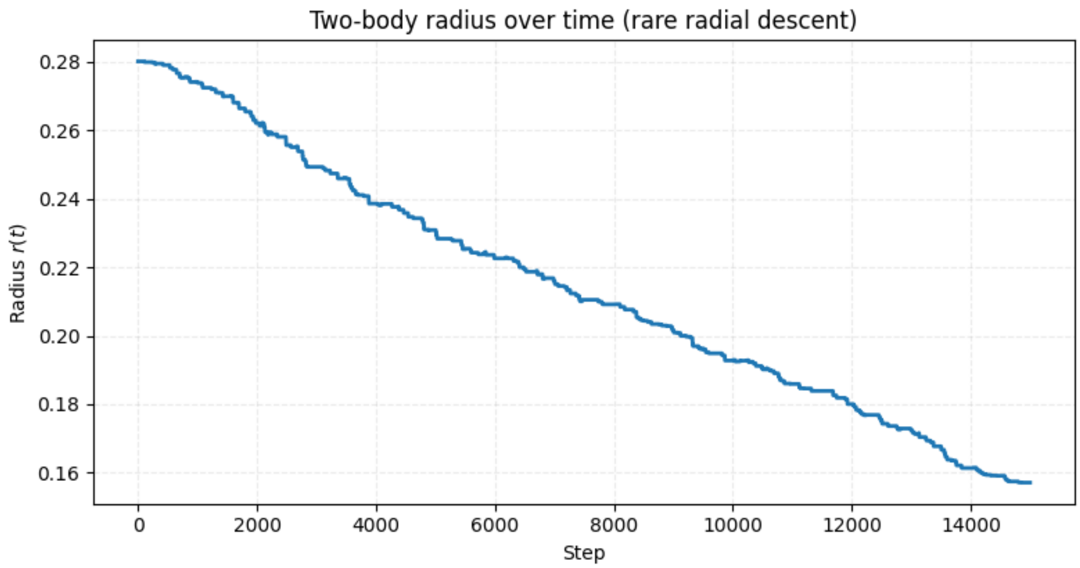

FIG. 5. **Two-body radius over time**. Staircase decrease in $r(t)$ under MH; smoother under underdamped Langevin (not shown).

Two-body inspiral and quasi-orbit.

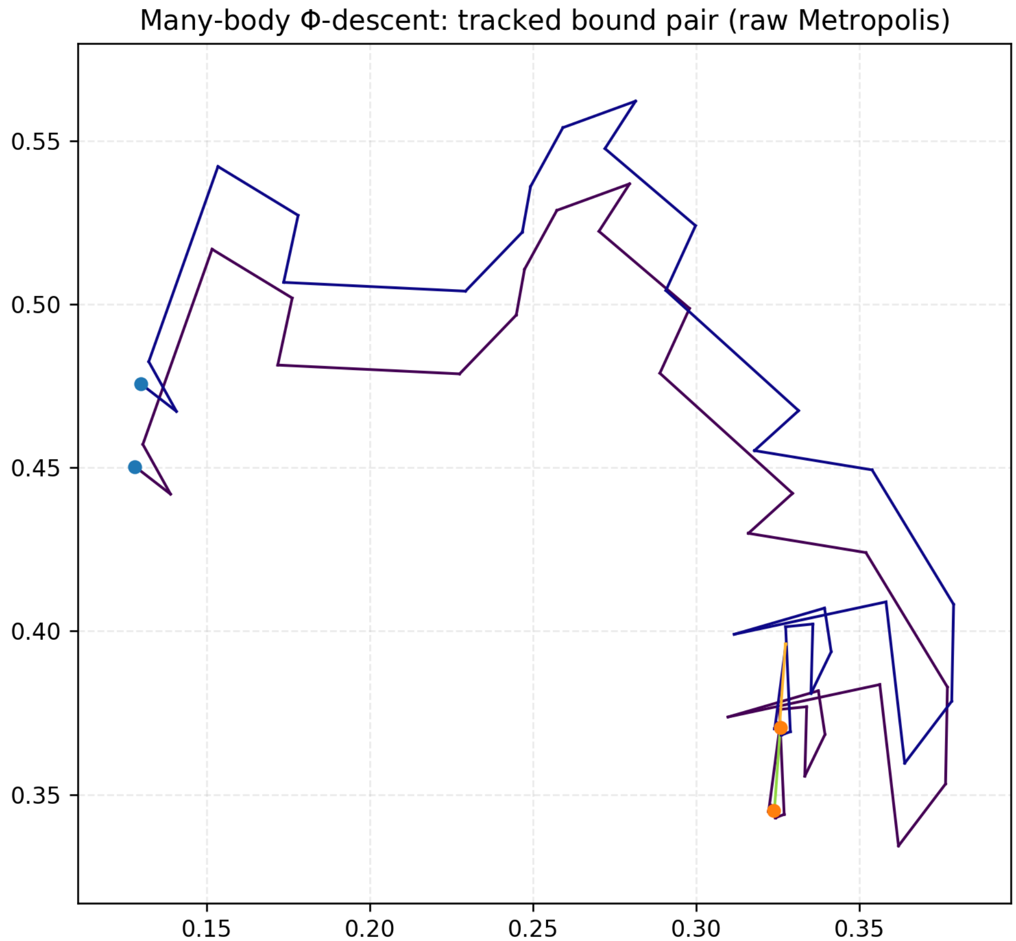

FIG. 6. **Many-body: tracked bound pair**. Closest pair trajectories (start •, end ●) show long arcs and intermittent radial descent.

Tracked bound pair in a many-body run.

9. QUANTUM MECHANICS AS COMPRESSION ACROSS POSSIBILITIES

A quantum state is an efficiently coded bundle of correlated futures,

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

Unitarity as code-preserving isometry. Under a Kraft-normalized inner product, linear maps preserving code length are isometries; physical evolution acts unitarily.

Incompatibility and uncertainty. Incompatible codebooks yield non-commuting generators; information-geometric bounds reproduce Robertson-type inequalities.

Entanglement. Shared algorithmic information $I_K(A\!:\!B) = K(A) + K(B) - K(A, B)$ formalizes entanglement; reduced states minimize Φ subject to subsystem constraints and recover von Neumann entropy in typical limits.

Born rule. Measurement selects outcomes with weights $P(\phi_k) \propto 2^{-\Delta\Phi_k}$. Under additivity, coarse-graining invariance, and normalization, $\Delta\Phi_k = -\log|\alpha_k|^2$ yields $P(\phi_k) = |\alpha_k|^2$. See Appendix B.

10. TEMPORAL COMPRESSION AND THE ORIGIN OF CAUSALITY

Let τ parametrize monotone Φ-descent ($d\Phi/d\tau \leq 0$). Physical time t is the reparametrization maximizing predictive compression subject to conservation constraints; causal orderings are fixed points of temporal coarse-graining. Dynamics $dx/dt \propto -\nabla\widehat{\Phi}$ then operate in emergent t without circularity.

11. PREDICTIONS AND FALSIFIABILITY

(P1) Short-range gravity correction (unit-bearing). Finite-resolution regularization at scale a leads to a smoothed kernel $k_a(r) = 1/\sqrt{r^2 + a^2}$, hence

$$\psi_a(r) \sim \frac{1}{r}\Big(1 - \frac{a^2}{2r^2} + \cdots\Big), \qquad F_a(r) = \frac{Gm_1m_2}{r^2}\Big(1 - \frac{3a^2}{2r^2} + \cdots\Big). \tag{11.1}$$

Sub-millimeter tests bound a; non-detection tightens a or constrains estimator locality.

(P2) Entanglement-assisted gravity. Algorithmic mutual information increases joint compression; predicts a small enhancement δ_{ent} in attraction for entangled masses (target 10^{-6}–10^{-4} at μm scales).

(P3) No particle dark matter. Rotation curves arise from description-curvature corrections (logarithmic tails) in galactic environments.

(P4) Dark energy evolution. Equation-of-state $w(z) = -1 + \delta w(z)$ with $|\delta w| \lesssim 0.05$ from structure-formation compression.

(P5) Statistical time symmetry breaking. Low-Φ-gradient systems show reversal excess $1 + \xi$, with $\xi \sim 10^{-3}$.

Appendix A: From Minimal Local Code Curvature to Poisson and $1/r$

Consider $\Phi[\rho] = \int \frac{1}{2}\|\nabla\psi\|^2\, d^3x$ with constraint $-\nabla^2\psi = \rho$. Variation with a Lagrange multiplier λ yields $\delta(\mathcal{E} + \int \lambda(-\nabla^2\psi - \rho)) = 0 \Rightarrow -\nabla^2\psi = \rho$ and $\nabla\cdot(\nabla\psi) = \rho$. Point sources $\rho = m\delta(x)$ have $\psi = mk(r)$, where harmonicity outside sources and isotropy fix $k(r) = 1/r$ in $n = 3$, giving an inverse-square field $F = -\nabla\psi \propto r^{-2}$.

Appendix B: Born Rule from Description Length

Let $\psi = \sum_k \alpha_k\phi_k$ encode compressed futures. Assume (i) additivity of description costs, (ii) invariance under coarse-graining of outcomes, (iii) normalization. Selecting an outcome ϕ_k adds $\Delta\Phi_k$ bits; define $P(\phi_k) \propto 2^{-\Delta\Phi_k}$. The axioms force $\Delta\Phi_k = -\log|\alpha_k|^2$, hence $P(\phi_k) = |\alpha_k|^2$.

Appendix C: Implementation Details for Simulations

Primary estimator: MST cost, Eq. (8.1), computed via Prim's algorithm [20]. Ablations: k-NN, Delaunay, Lempel–Ziv of voxelized positions (8–12 bits per axis). Updates: MH at β=10 and underdamped Langevin with (m, γ) chosen to match typical step sizes. Repository (reproducibility, scripts, and the Bell test code): `https://github.com/Snassy-icp/law_of_minimal_description/tree/main/code/simulations`.

Appendix D: Responses to Common Objections

Uncomputability. Addressed by admissible $\widehat{\Phi}$ and Prop. 1. **Closed-system entropy.** Resolved by model/data split and supermartingale remark (Sec. 3 B). **"You assumed Laplace."** Replaced by Thomson/Dirichlet principle (App. A). **Quantum formalism.** Codes $\rightarrow$ Hilbert and $\hbar$ calibration (App. G). **Dimensionality.** Heuristic argument (App. E).

Appendix E: Why 3 Dimensions? A Heuristic

n=3 uniquely supports (i) local, isotropic, scale-free kernels with conserved flux; (ii) harmonic Green's functions with finite-energy bound structures; (iii) additive compression flux under partition. In n<3 global structures are unstable or trivial; for n≥4 scale-free kernels trade off stability vs. finite local flux. Hence $k(r) = 1/r$ in 3D maximizes compression consistency.

Appendix F: Gradient Consistency: Measure and Refinement

We endow the finite-precision configuration space with the product topology and the counting/Lebesgue hybrid measure μ. Admissible $\widehat{\Phi}_{a,b}$ are local, Lipschitz, and prefix-free MDL; Γ-convergence as $(a, b) \to (0, \infty)$ holds under refinement stability. Discontinuity sets of K are μ-null in this topology. Hence directional derivatives of Φ and $\widehat{\Phi}$ agree μ-a.e. in full-measure cones.

Appendix G: Codes → Hilbert; $\hbar$ Calibration; CCR Sketch

Codes → Hilbert. Let prefix-free codewords form coordinates with Kraft normalization. Inner product $\langle \psi, \phi \rangle = \sum_i c_i^* d_i$ defines $\mathcal{H}$. Code-preserving linear maps are isometries, hence unitary up to phase.

$\hbar$ calibration. In Euclidean signature, weight $e^{-S_E/\hbar}$; assign $2^{-\kappa\Phi}$ to description weight. Identify $\kappa = \hbar / \ln 2$ so path weights and description weights coincide after Wick rotation.

CCR sketch. The local quadratic code length induces a Fisher metric; maximizing likelihood subject to variance yields Robertson-type inequalities $\Delta A \, \Delta B \geq \frac{1}{2} |\langle [A, B] \rangle|$ with the $\hbar$ scale fixed by the above calibration.

Appendix H: Bell/CHSH as a Non-Tautological MDL Witness

For settings $(X, Y) \in \{0, 1\}^2$, define the score bit $s := A \oplus B \oplus (X{\cdot}Y)$. Let $\omega = \Pr[s = 0]$. For N trials, *ideal* savings vs. fair coin equal $N[1 - h_2(\omega)]$. To avoid tautology, we report (i) train/test MDL (fit p on half, code the other half), (ii) KT universal codelength (prequential, no fitting), and (iii) fixed-parameter MDL under LHV/Q/PR priors. Results (typical $N =$

$2 \cdot 10^5$): LHV $\omega \approx 0.75$ (train/test $\approx 0.5 \cdot 0.189N$), Quantum $\omega \approx \cos^2 \frac{\pi}{8}$ ($\approx 0.5 \cdot 0.399N$), PR $\omega = 1$ ($0.5N$). Individual A, B streams remain incompressible (≈ 0 savings), confirming no signalling. Code in repository.

REFERENCES

---

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

[2] A. N. Kolmogorov, "Three Approaches to the Quantitative Definition of Information," *Problems of Information Transmission* (1965).

[3] G. J. Chaitin, "On the Length of Programs for Computing Finite Binary Sequences," *J. ACM* (1966).

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

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

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

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

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

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

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

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

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

[13] A. Braides, Γ-*Convergence for Beginners* (Oxford, 2002).

[14] R. Krichevsky, V. Trofimov, "The Performance of Universal Encoding," *IEEE Trans. IT* **27**, 199–207 (1981).

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

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

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

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

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

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