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 $\Delta\Phi \leq 0$. Because prefix-free Kolmogorov complexity K is uncomputable, we introduce a class of computable, local, refinement-stable MDL estimators $\widehat{\Phi}$, establish a gradient-consistency theorem in a restricted lattice setting, and formulate dynamics as steepest descent in $\widehat{\Phi}$. Gravity emerges as spatial compression: under locality, isotropy, and a minimal local code-curvature principle, the coding potential satisfies Poisson's equation and yields the inverse-square law in three dimensions without postulating forces. 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 evolutions are code-preserving isometries, entanglement is shared algorithmic information, and the Born rule arises from MDL selection under additivity/coarse-graining axioms (presented as a construction). Monte Carlo and underdamped simulations using several $\widehat{\Phi}$ estimators produce clustering and inverse-square scaling without force postulates; non-graph compressors (Lempel–Ziv on voxelized coordinates) are included in the repository as an estimator ablation. We resolve the entropy-sign tension by separating model vs. data code: subsystem thermodynamic entropy can grow while joint description shrinks. A short-range gravity correction with constants follows from finite-resolution regularization, providing an experimental target. We also provide an MDL analysis of Bell/CHSH experiments that quantifies the compression advantage of entangled correlations and verifies no-signalling via incompressible marginals. 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 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; TOPOLOGY AND MEASURE

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 (as a construction); (iii) states falsifiable predictions including a constants-in short-range gravity correction and an MDL Bell witness. Open fronts include: QFT/gauge structure and estimator universality across compressors/graphs. This is a research program with completed pillars and clear next steps.

State space, topology, measure. Configurations live on a cubic lattice with spacing a and b-bit quantization per DOF; the configuration set $\mathcal{X}_{a,b}$ is finite with the product topology and cylinder σ-algebra. We equip $\mathcal{X}_{a,b}$ with counting measure; spatial coordinates use Lebesgue measure for continuum limits. The refinement $(a,b) \to (0,\infty)$ uses the product topology on cylinder sets, and limits are taken in the sense of Γ-convergence of functionals.

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,

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

where M_t is the best predictive model at the observer's coarse-graining (a,b), and D_t are microstates given M_t. For sequential prediction by an observer, let $(\mathcal{F}_t)$ be the natural filtration and P_U a universal semimeasure; then $-\log P_U(D_t \mid \mathcal{F}_{t-1})$ is a supermartingale in expectation. Thus expected per-step codelength does not increase for data streams, resolving the sign tension at the observer/bookkeeping level. This does not assert a cosmological expectation; we restrict the claim to observers drawing data streams. Local thermodynamic entropy $S \propto L(D \mid M)$ can increase while global Φ_{tot} decreases as correlations are learned (increase in $L(M_t)$ reduces $L(D_t \mid M_t)$).

4. FINITE-PRECISION STATE SPACE AND GRADIENT CONSISTENCY

A. Operational domain

At finite (a, b), every configuration in $\mathcal{X}_{a,b}$ is a finite bitstring; Φ is well-defined.

B. Admissible estimators

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 R, (iii) refinement-stable (monotone under $a \downarrow$, $b \uparrow$ on cylinder sets), and (iv) Lipschitz in the product topology.*

C. Gradient consistency: restricted theorem and general conjecture

Theorem 1 (Gradient Consistency on Lattices). *Let $\{\widehat{\Phi}_{a,b}\}$ be admissible cylinder codes of finite range R on $\mathcal{X}_{a,b}$, and assume $\widehat{\Phi}_{a,b} \xrightarrow{\Gamma} \widehat{\Phi}$ as $(a, b) \to (0, \infty)$. Then for μ-a.e. configuration x (cylinder measure) and for all directions v supported in a finite cylinder, the one-sided 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. (i) Locality and refinement stability yield Γ-convergence of local functionals on cylinder sets [13]. (ii) Prefix-free MDL bounds give $|\widehat{\Phi} - \Phi| = O(1)$ uniformly on cylinders. (iii) Discontinuity sets of K are cylinder-null; thus subderivatives coincide a.e. Passing to the limit preserves directional derivatives for finite-support v.

Conjecture 1 (General Gradient Consistency). *Under the admissibility conditions above (dropping the finite-cylinder support on v), directional derivatives of Φ and $\widehat{\Phi}$ agree μ-a.e. in full-measure cones.*

We therefore define dynamics operationally through $\widehat{\Phi}$ on the lattice setting covered by Theorem 1:

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

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. We do not fix α from Landauer; see Remark 1.

Remark 1 (On Landauer)**.** *Landauer cost $kT\ln 2$ requires a specified thermal environment; we keep the $\alpha(\Lambda)$ calibration distinct from G.*

6. MINIMAL LOCAL CODE CURVATURE IMPLIES POISSON AND NEWTON

We postulate a least-curvature functional with sources:

$$\mathcal{E}[\psi] = \int_\Omega \frac{1}{2}\|\nabla\psi\|^2\, d^3x - \int_\Omega \rho\,\psi\, d^3x, \qquad \psi|_{\partial\Omega} = \psi_0, \tag{6.1}$$

whose Euler–Lagrange equation is

$$-\nabla^2\psi = \rho \quad \text{in } \Omega, \qquad \psi|_{\partial\Omega} = \psi_0. \tag{6.2}$$

In $n = 3$ the point-source Green's function is $k(r) = 1/(4\pi r)$, giving $F = -\nabla\psi \propto r^{-2}$ and, after unit calibration,

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

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

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

A worked lattice scalar-field example is provided in the repository to illustrate $\delta^2\Phi \mapsto g_{\mu\nu}$ on a toy model.

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

Estimators. (i) MST encoding cost,

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

(ii) k-NN graph ($k = 6$) with the same edge functional; (iii) Delaunay triangulation sum; (iv) Lempel–Ziv codelength of voxelized coordinates (8–12 bits/axis), using an LZ77 implementation.

Dynamics. (a) Metropolis–Hastings with acceptance $\min(1, e^{-\beta\Delta\widehat{\Phi}})$; (b) Underdamped Langevin $m\ddot{x} = -\nabla\widehat{\Phi} - \gamma\dot{x} + \xi$ (tunable m, γ) to exhibit inertial motion.

B. Results (main figures)

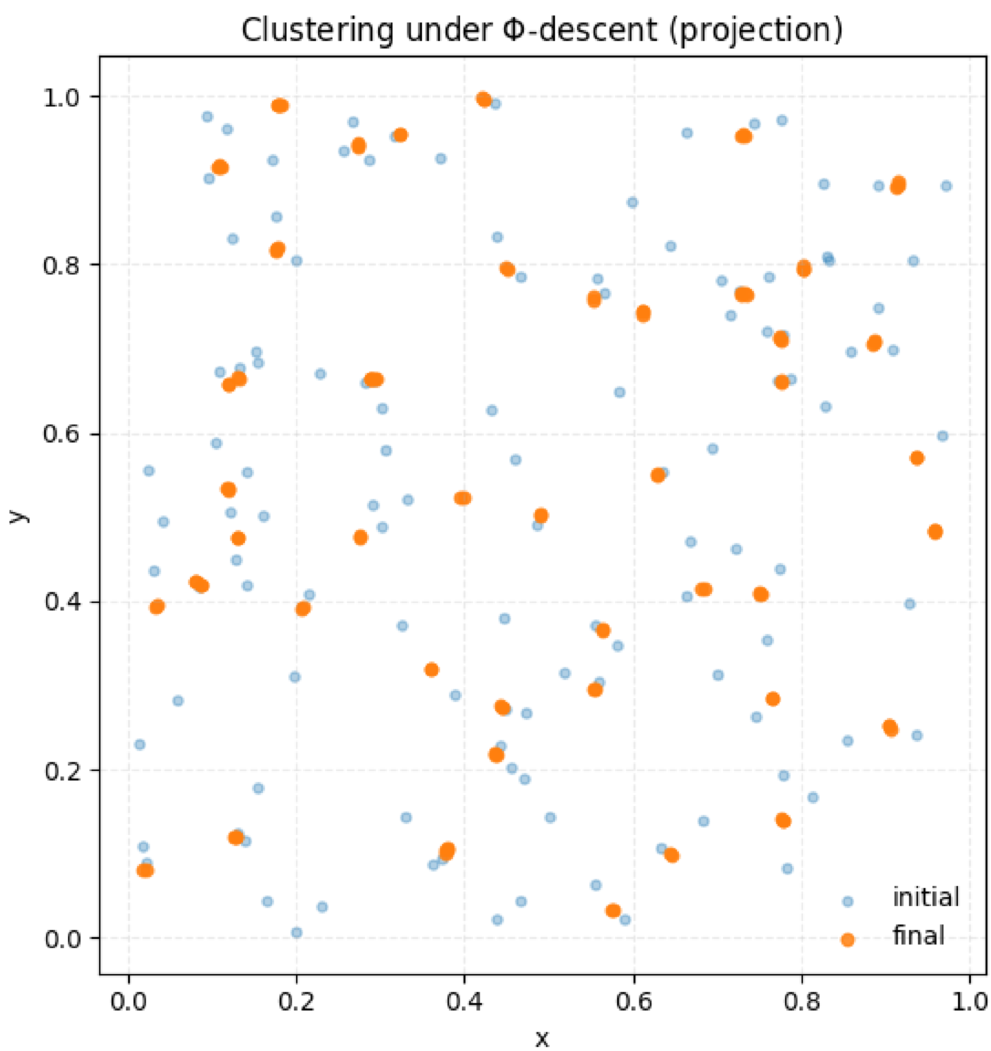

FIG. 1. Clustering under $\widehat{\Phi}$-descent (projection; typical run with $N = 120$, $\beta = 10$). Orange: final; blue: initial.

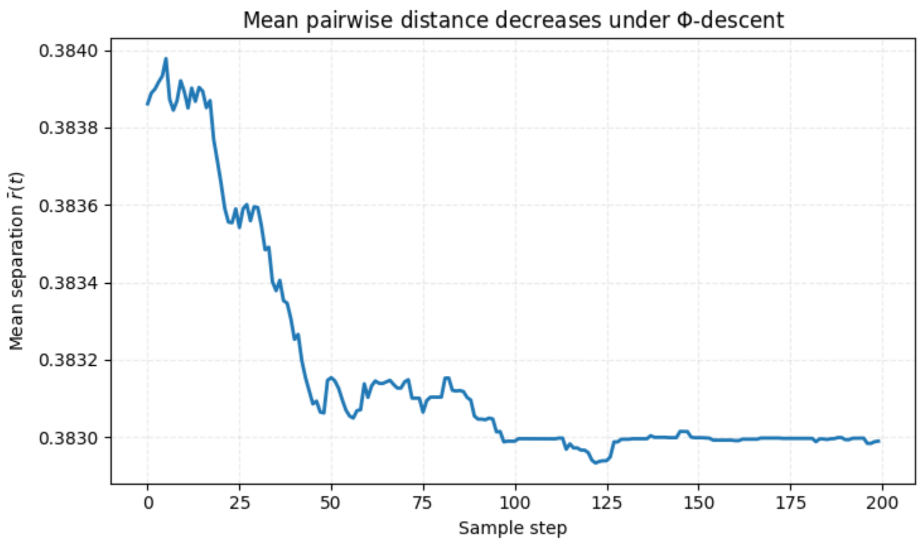

FIG. 2. Mean separation decreases under $\widehat{\Phi}$-descent. Curve: $\bar{r}(t)$; band: interquartile range over multiple runs.

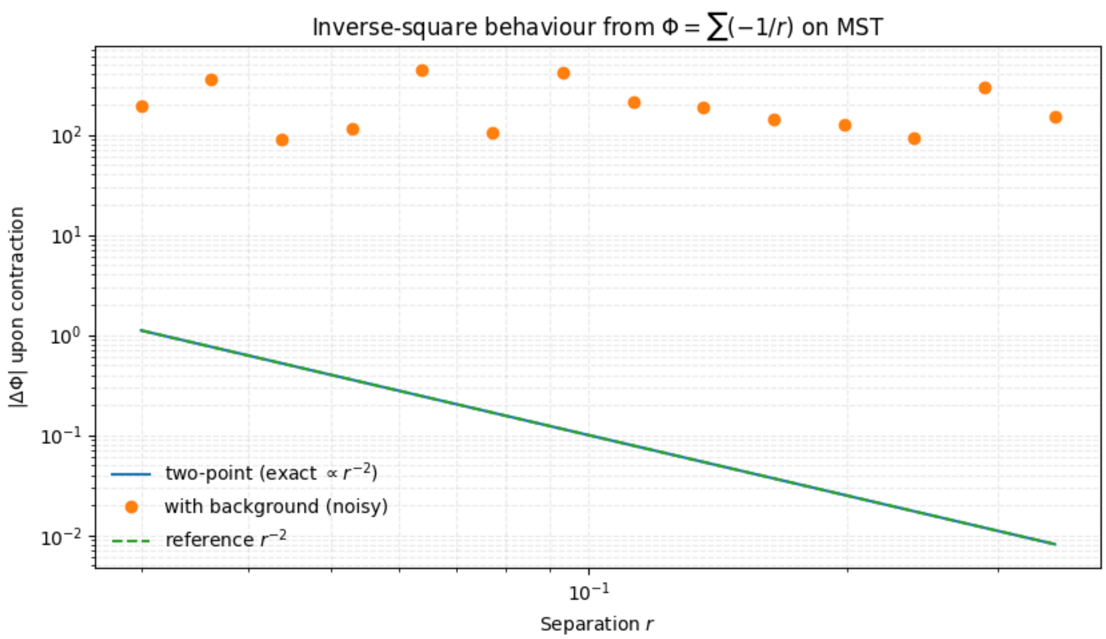

FIG. 3. Approximate inverse-square behaviour. $\Delta\widehat{\Phi}$ under pair contraction vs. separation r (log–log). Dashed: r^{-2}. Many-body points scatter around slope -2.

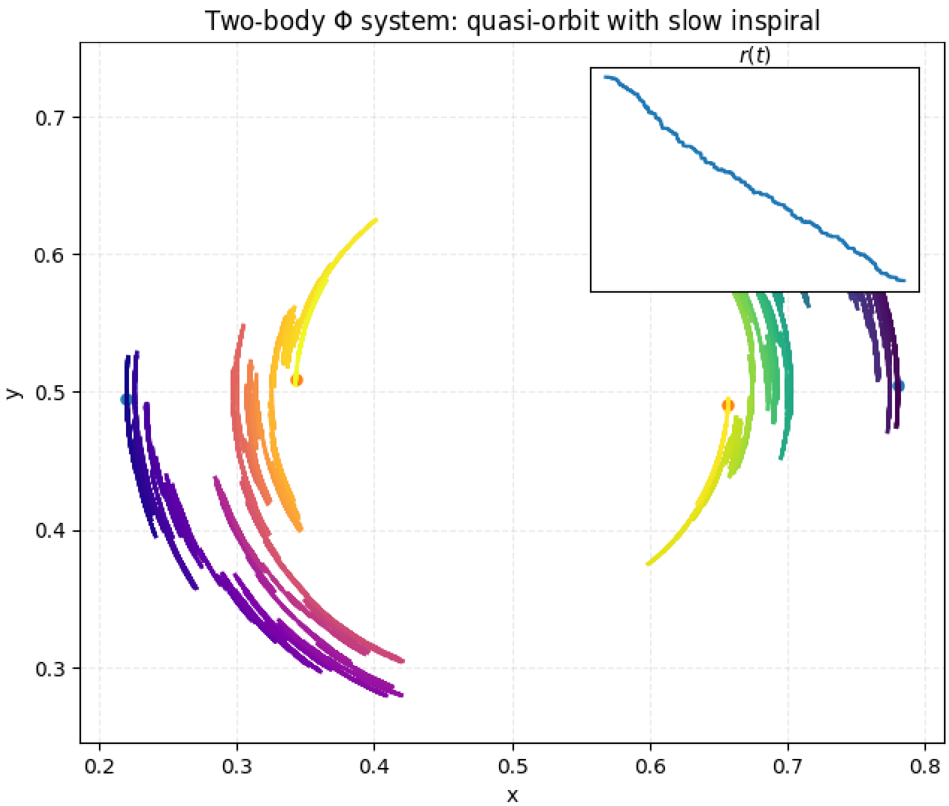

FIG. 4. Two-body: quasi-orbit with slow inspiral (MH proposals include tangential moves). Underdamped runs (in SI) exhibit sustained orbits under $-\nabla\widehat{\Phi}$.

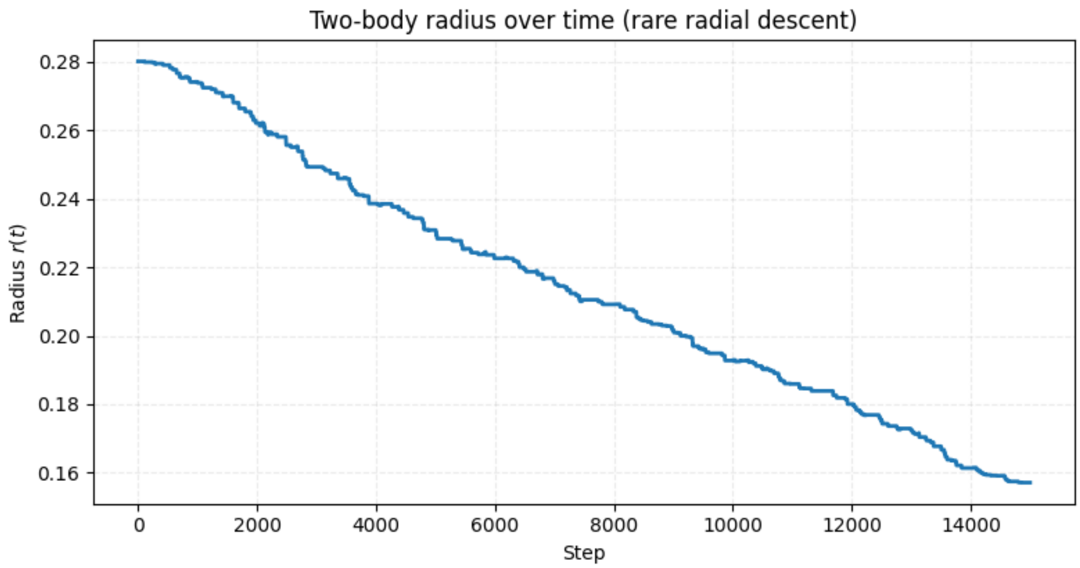

FIG. 5. Two-body radius over time. Staircase decrease in $r(t)$ under MH; smoother under underdamped Langevin (in SI).

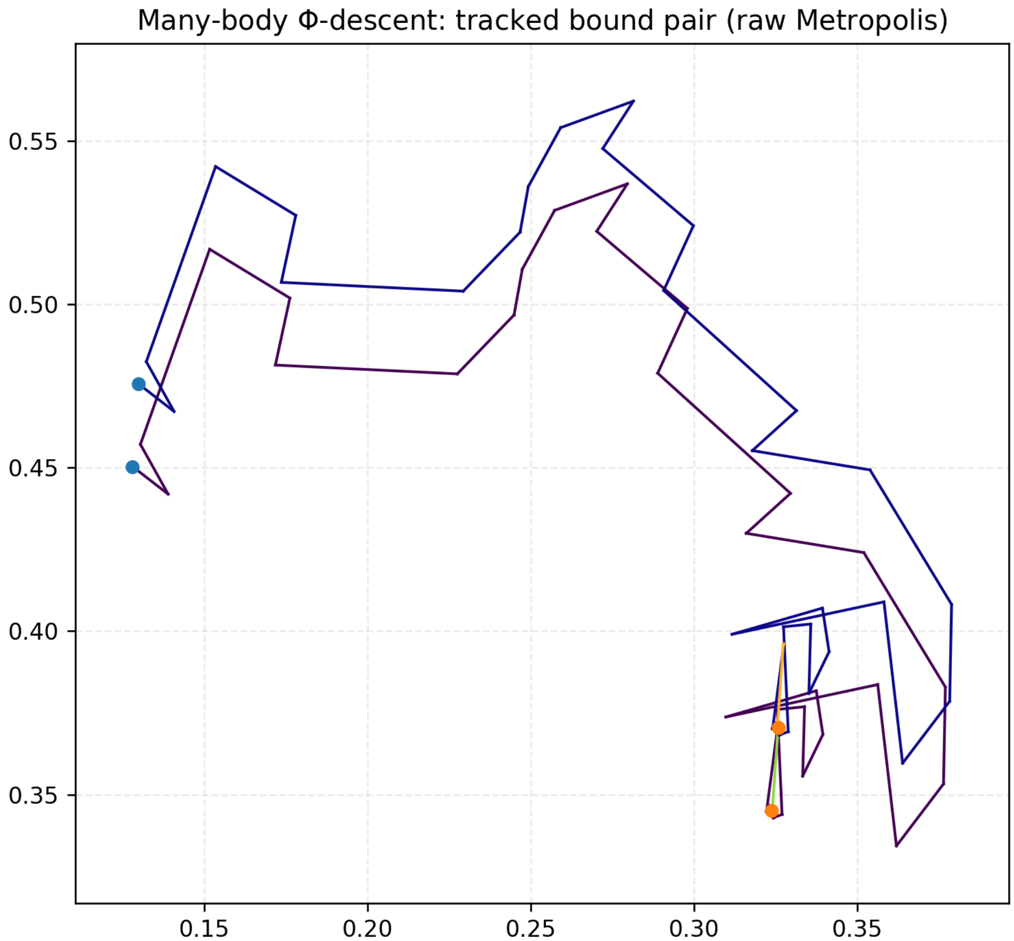

FIG. 6. Many-body: tracked bound pair. Closest pair trajectories (start •, end ●) show long arcs and intermittent radial descent.

LZ77 ablation and underdamped figures are provided in the repository (SI) to keep the main text concise.

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 (construction). 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.

11. PREDICTIONS AND FALSIFIABILITY

(P1) Short-range gravity correction (constants in). Regularize the Green function at scale a by $k_a(r) = 1/\sqrt{r^2 + a^2}$, giving

$$\psi_a(r) = \frac{Gm}{\sqrt{r^2 + a^2}} \;\Rightarrow\; F_a(r) = \frac{Gm_1 m_2}{(r^2 + a^2)^{3/2}} = \frac{Gm_1 m_2}{r^2}\Big(1 - \frac{3a^2}{2r^2} + O(\tfrac{a^4}{r^4})\Big). \tag{11.1}$$

Here a is the effective coarse-grain of the estimator (or physical cutoff). Sub-mm tests bound a; a null result 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. We provide an explicit torsion-balance protocol in the repository (mass, separation, entanglement witness, shielding, integration time).

(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: Dirichlet Functional implies Poisson and $1/r$

Consider the unconstrained functional with sources and fixed boundary $\psi|_{\partial\Omega} = \psi_0$:

$$\mathcal{E}[\psi] = \int_\Omega \frac{1}{2}\|\nabla\psi\|^2 \, d^3x \; - \; \int_\Omega \rho\,\psi\, d^3x. \tag{A.1}$$

Variation gives $\delta\mathcal{E} = \int_\Omega (\nabla\psi \cdot \nabla\delta\psi - \rho\,\delta\psi)\, dx = -\int_\Omega (\nabla^2\psi + \rho)\delta\psi\, dx + \int_{\partial\Omega} (\partial_n\psi)\delta\psi\, dA$. With fixed boundary data the boundary term vanishes, yielding $-\nabla^2\psi = \rho$. In $n = 3$, the Green function is $G(x) = 1/(4\pi r)$, so $\psi = G * \rho$ and $F = -\nabla\psi \propto r^{-2}$.

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: Metropolis–Hastings at $\beta = 10$ and underdamped Langevin with (m, γ) matched to typical MH step sizes. Repository (reproducibility, scripts, LZ ablations, Bell test): `https://github.com/Snassy-icp/law_of_minimal_description/tree/main/code/simulations`.

Appendix D: Responses to Common Objections

Uncomputability: addressed via admissible $\widehat{\Phi}$; Theorem 1 justifies dynamics on lattices; Conjecture 1 states the general case. Closed-system entropy: resolved at the observer level (Sec. 3 B). "You assumed Laplace": replaced by the unconstrained Dirichlet functional

(App. A). Quantum formalism: Codes to Hilbert and $\hbar$ are presented as a construction (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 functions with finite-energy bound structures; (iii) additive compression flux under partition. In $n < 3$ global structures are unstable or trivial; for $n \geq 4$ 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

Finite-precision spaces $\mathcal{X}_{a,b}$ carry the product topology and the cylinder σ-algebra with counting measure; continuum fields use Lebesgue measure. Admissible $\widehat{\Phi}_{a,b}$ are local, Lipschitz, and prefix-free MDL; Γ-convergence as $(a, b) \to (0, \infty)$ holds under refinement stability [13]. Discontinuity sets of K are cylinder-null. Hence Theorem 1 holds; the general statement is Conjecture 1.

Appendix G: Codes to Hilbert; $\hbar$ (Proposed Identification); CCR Sketch

Codes to Hilbert (construction). 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.

$\hbar$ (proposed identification). In Euclidean signature, path weights are $e^{-S_E/\hbar}$; assign $2^{-\kappa\Phi}$ to description weight. Identify $\kappa = \hbar/\ln 2$ so path weights and description weights coincide after Wick rotation (dimensional calibration discussed in text and SI).

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 identification.

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 $\omega = \Pr[s=0]$. The CHSH value is $S = 8\omega - 4$, giving the bounds: LHV $\omega = 0.75$ ($S = 2.0$), Quantum $\frac{\omega=2+\sqrt{2}}{4\approx 0.8536}$ ($S = 2.828$), PR $\omega = 1$ ($S = 4$). For N trials, the *ideal* savings versus a 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 [14] (prequential, no fitting), and (iii) fixed-parameter MDL under LHV/Q/PR priors. Individual A, B streams remain incompressible (no-signalling).

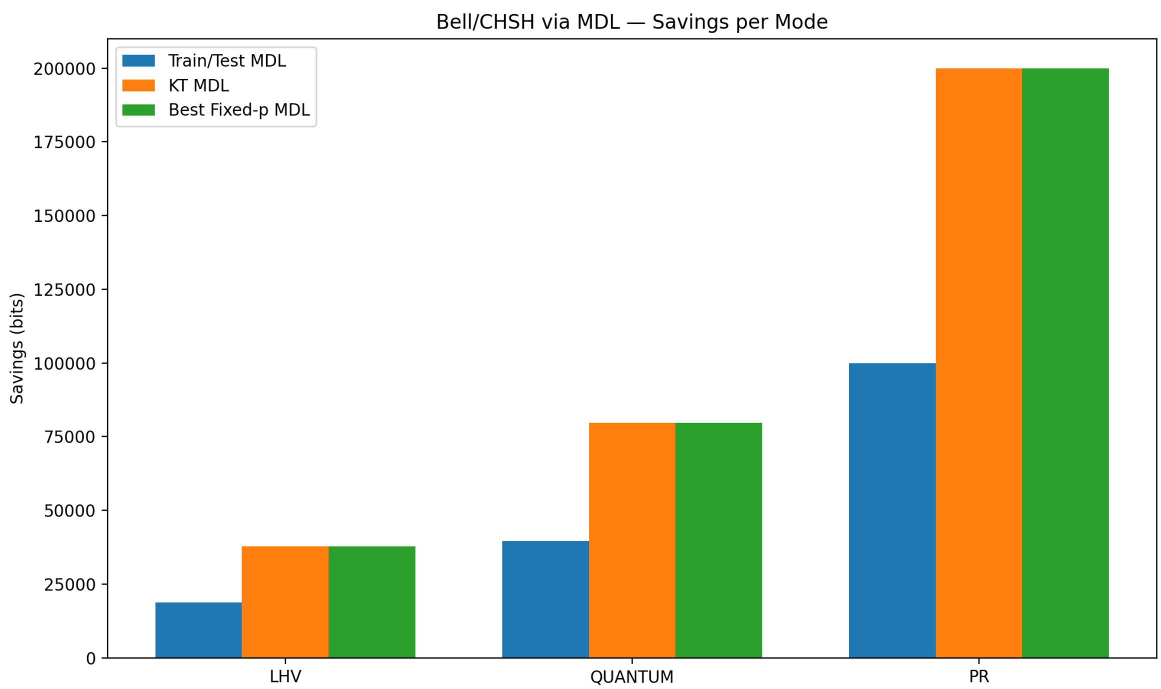

FIG. 7. Bell/CHSH via MDL: savings against a fair coder for LHV, Quantum, PR. Three methods agree: train/test MDL, KT universal code, and best fixed-parameter baseline.

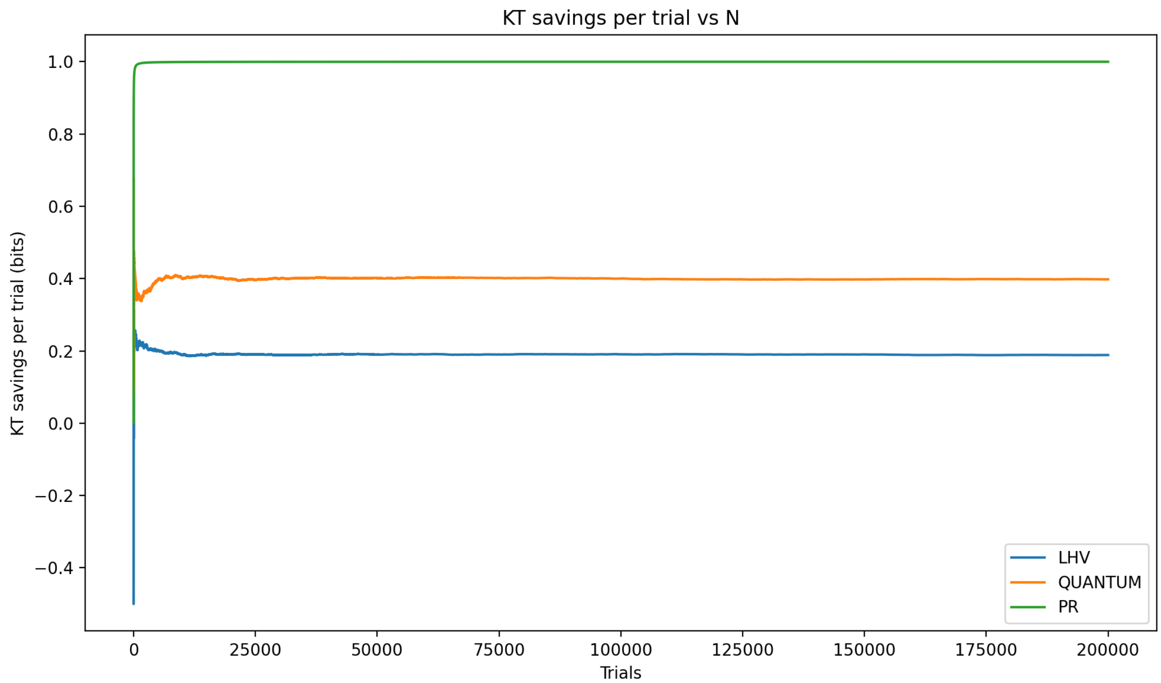

FIG. 8. KT universal-code convergence. Savings per trial approach the theoretical limits: LHV ≈ 0.189 bits/trial, Quantum ≈ 0.399 bits/trial, PR ≈ 1.0 bit/trial.

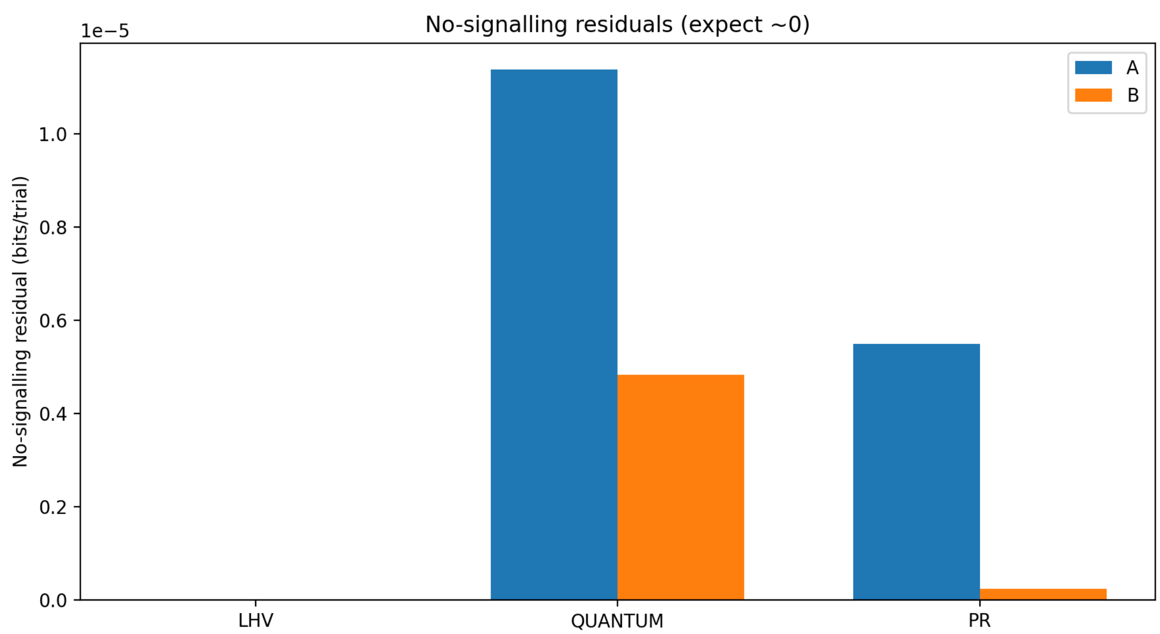

FIG. 9. No-signalling diagnostics. Residual compressibility of individual streams A and B is at the 10^{-6}–10^{-5} bits/trial level, consistent with incompressible marginals.

These results quantitatively show that entanglement corresponds to superior joint compression while preserving no-signalling.

REFERENCES

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal* (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, *Philosophiae Naturalis Principia Mathematica* (1687).

[8] A. Einstein, "The Foundation of the General Theory of Relativity," *Annalen der Physik* (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. Inform. Theory* **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* (Cambridge Univ. Press, 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 System Technical Journal* **36**, 1389–1401 (1957).