Waddington Landscape Cell Fate via the Latent Framework

Dr. Tamás Nagy

tnagyphd@gmail.com

draft • 2026-04-08

Executive summary

Cell-fate decisions are still explained with Waddington-style landscape language, but the metaphor rarely connects cleanly to quantitative barriers, transition rates, and compressibility of variation in high-dimensional measurements. This paper's aim is modest and practical: package that story inside the Latent framework so that canalization becomes two numbers—compressibility and an effective dimension—together with barrier-based rate heuristics.

The formal contribution is a machine-checked chain of thirty-six real-arithmetic lemmas, grouped to mirror attractors, rates, reprogramming, pluripotency, differentiation, and cross-domain analogies. The lemmas record ordering relationships among scalar summaries (barriers, gaps, depths, rates), not a full PDE or stochastic-process theory in the body of the proof file.

The numerical section uses deliberately simple synthetic potentials in one dimension. The reported rho values there come from a spectral dominance ratio between the two leading discrete modes of the sampled potential (a proxy for how peaked the energy profile is), while N-star is the rank needed to capture ninety percent of spectral power; these operational definitions coincide with the §2.2 Latent narrative only at the level of interpretation, not as a claim that the full covariance-based compressibility ratio from §2.2 was estimated from gene-expression tensors.

Abstract

Waddington's epigenetic landscape metaphor remains the dominant intuitive picture for cell fate: stable types are valleys, differentiation is downhill flow, and reprogramming lifts cells across ridges. Despite its influence, the metaphor has lacked a quantitative backbone that connects barrier heights, transition rates, and high-dimensional gene-regulatory dynamics in a single coordinate system usable across datasets.

This paper formalizes Waddington-style landscapes in the Latent framework. Phenotypic states live in a reduced coordinate system; differentiation and reprogramming correspond to flows and controlled barrier modifications of a Latent potential. The Latent Number ρ measures how compressible fate-associated variation is relative to a maximal-entropy reference, while the effective dimension N^* counts the modes needed to reconstruct fate coordinates at fixed accuracy. Together they quantify canalization: deep, low-N^* basins correspond to strong commitment.

We machine-check thirty-six theorems in six groups covering attractor summaries, rate-scale inequalities, reprogramming-related order constraints, pluripotency-style shallow-basin regimes, differentiation-style deep-basin regimes, and cross-domain bridge inequalities. Numerical case studies on one-dimensional two-state, three-state, and reprogrammed potentials report a spectral compressibility proxy $\rho \in [1.4, 2.4]$ and an N^*/N ratio below 1% under the §4 measurement

convention; the bundled script evaluates analytic Kramers factors from extracted barriers (no path sampling), and the reprogramming construction lowers the dominant barrier gap as checked against an unperturbed two-well control. Fifteen of fifteen automated tests pass in numerical_validation.py.

The emphasis on synthetic potentials is deliberate: they isolate the logical commitments of the Latent calculus before empirical identification noise enters. Subsequent papers will invert the pipeline, learning Φ and V jointly from paired time-series and snapshot atlases while reusing the same theorem groups as consistency filters.

1. Introduction

1.1 Why Waddington still matters

Cell fate decisions are central to development, regeneration, and cancer. Modern single-cell atlases provide millions of cells in high-dimensional expression space, yet biologists still explain outcomes with landscape language. The gap is mathematical: vector fields, potentials, and barriers must be inferred under noise, sparsity, and missing time resolution.

1.2 Latent landscapes and canalization

The Latent framework supplies three commitments: (i) a low-dimensional embedding Φ of cell states; (ii) a scalar potential V on Latent coordinates whose minima are attractors; (iii) quantitative complexity measures ρ and N^* that summarize how "tight" fate structure is. Canalization corresponds jointly to large ρ (strong compression of fate variation) and small N^*/N (few latent directions matter).

1.3 Contributions

We build a machine-checked chain of thirty-six theorems in six thematic groups, validate ordering predictions on synthetic one-dimensional potentials with controlled barrier heights, and record cross-domain bridge inequalities that parallel protein-folding funnels and evolutionary fitness landscapes at the level of barrier and compression summaries.

1.4 Reader map

Experimental biologists can read §1–§4 for the operational meaning of ρ, N^*, and barrier heights; mathematicians can prioritize §2–§3 and the Kramers consistency checks in §4. The cross-domain section is intentionally brief: full functor diagrams live in the formal proof artifact.

2. Mathematical Framework

2.1 State space and noise

Notation: $\mathbb{R}^N$ carries the standard inner product; $\|\cdot\|$ is Euclidean norm; $B_r(x)$ is the closed ball of radius r.

Let $x \in \mathbb{R}^N$ denote a high-dimensional molecular profile (e.g., log-expression). A stochastic differential model

$$\mathrm{d}x_t = -\nabla U(x_t)\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}W_t$$

encodes drift toward attractors of an energy-like function U plus diffusion strength D. The Latent map $\Phi : \mathbb{R}^N \to \mathbb{R}^d$ with $d \ll N$ induces a reduced potential $V(z) = U(\Phi^\dagger z)$ on coordinates $z = \Phi(x)$, where $\Phi^\dagger$ is a right inverse chosen by the ridge-regularized pseudoinverse in the numerical suite.

2.2 Latent Number and effective dimension

Define $\rho = \sigma_0^2/\sigma_*^2$ as in the general Latent protocol: σ_0^2 is baseline variation of fate coordinates under a reference ensemble, and σ_*^2 is residual variance after optimal rank-N^* truncation in the Latent basis. Effective dimension N^* is the smallest rank achieving reconstruction error below ε.

2.3 Barriers and Kramers scaling

For a bistable Latent potential with barrier ΔV between basins, overdamped Kramers theory predicts escape rates $\kappa \propto \exp(-\Delta V/D)$ up to prefactor corrections. The bundled numerical harness (§4) evaluates the **analytic** Kramers factor $\exp(-\Delta V/D)$ from extracted discrete barriers on fixed one-dimensional grids; it does not integrate Langevin trajectories or estimate transition frequencies from particle paths.

2.4 Reprogramming as barrier surgery

Reprogramming interventions are modeled as a parameterized family V_θ with $\theta = 0$ the wild-type landscape and $\theta > 0$ a protocol that flattens specific saddles. Theorems in Group C formalize monotonicity of $\Delta V(\theta)$ along convex combinations of potentials under stated curvature constraints.

2.5 Identifiability note

Latent potentials are not unique: any reparameterization of coordinate z that preserves basins induces an equivalent landscape description. Empirical estimates of ρ and N^* generally depend on the chosen embedding Φ and on how V is normalized, even when the underlying minima and saddles are unchanged. Practitioners should therefore report both Φ and the normalization convention for V, and treat cross-paper comparisons of ρ as meaningful only when those choices are aligned.

2.6 Discrete cell-state graphs

Many single-cell pipelines first cluster cells into a graph. The Latent framework admits a graph discretization: nodes are metastable clusters, edge weights encode transition flux, and V can be related to graph-Laplacian constructions in the usual way. The lemmas in Group B are chosen so their real-arithmetic conclusions echo Kramers-type scaling and gap–mixing heuristics on such discretizations; a quantitative mesh-refinement theorem identifying discrete spectra with a continuous generator is outside the scope of the present proof file.

3. Formal Proof Chain

Formal scope. The machine-checked lemmas are universal implications in real arithmetic: they encode the biological headings below as inequalities among nonnegative summary parameters (barriers, attempt rates, gaps, depths, doses). They should be read as a disciplined "order-calculus" for

landscape cartoons, not as a self-contained existence theory for global Lyapunov functions, PDE generators, or categorical diagrams.

Each group lists six theorems; proofs are machine-checked in elysium/ fields /bio_waddington/platonic.py (thirty-six p.prove targets; run python3 on that file to reproduce).

Group A — Attractor properties (6 theorems). Nonnegativity of basin-depth summaries; positivity of barrier heights under slack hypotheses; ordering of attractor versus saddle energies; basin minima lying below reference levels; gradient-scale versus depth inequalities; positivity of stability margins.

Group B — Transition rates (6 theorems). Exponential barrier dependence in Arrhenius-style parameters; rate bounds tied to gap parameters; mixing-time inequalities from spectral-gap summaries; positivity and ordering of attempt rates; faster transitions under smaller barriers; curvature-strength effects on attempt rates.

Group C — Reprogramming (6 theorems). Monotone barrier–rate relationships under lowering; Yamanaka-style shallowing bounds; monotone dose response for reprogramming factors; overlap constraints for partial protocols; stochastic boost inequalities; ordering of metastable energies.

Group D — Pluripotency (6 theorems). Shallow multi-attractor regimes; high plasticity versus committed gaps; flat-region path counts; entropy summaries bounded by diversity parameters; reversible barrier symmetry; naive stem-basin width inequalities.

Group E — Differentiation (6 theorems). Deep-basin slow escape; low transition rates in committed regimes; barrier growth along maturation summaries; lineage separation energies; canalization-depth inequalities; terminal-attractor stability margins.

Group F — Cross-domain (6 theorems). Parallel barrier inequalities between cell-fate and folding-funnel summaries; funnel-like basin width constraints; spectral coupling inequalities between fitness and epigenetic barrier proxies; subadditivity-style triple inequalities for composed barriers; universal escape scaling; landscape compression ratios as cross-domain analogy.

Dependency sketch. Groups A and B provide the common ordering spine. Groups D–E specialize the same inequality vocabulary to shallow- versus deep-basin regimes. Group C interpolates barrier and dose parameters between them. Group F reuses barrier, width, and compression summaries for folding and fitness metaphors without importing new analytic machinery.

4. Numerical Validation

We constructed three synthetic one-dimensional potentials: a two-state landscape, a three-state lineage tree, and a reprogrammed variant with lowered saddle between previously separated attractors. The accompanying script elysium/ fields /bio_waddington/numerical_validation.py scans each discretized potential for basins and saddles, checks positivity of dominant barrier gaps, evaluates the analytic Kramers factor on the two-state baseline, and computes the §4 spectral diagnostics below (reproducible at grid size $N = 500$ with the defaults in that file).

Measurement note. The quantities ρ and N^* in Table 1 are computed from the discrete Fourier power spectrum of each sampled potential (ratio of the two leading power components for ρ; smallest rank capturing 90% of power for N^*, relative to grid size N). They operationalize "compressibility" and "low effective dimension" for these synthetic curves. They are **not** direct Monte Carlo estimates

of the full σ_0^2/σ_*^2 Latent statistic in §2.2, which would require an explicit latent embedding Φ and reference ensemble.

Landscape	ρ (spectral proxy)	N^*/N	Dominant barrier ΔV (grid scale)
Two-state	2.33	0.60%	0.086
Three-state	1.41	0.80%	4.27
Reprogrammed	2.36	0.60%	1.12

Across scenarios, the spectral proxy satisfies $\rho \in [1.4, 2.4]$ and $N^*/N < 1\%$ under the default grid in numerical_validation.py, reflecting strong compressibility of these calibrated one-dimensional potentials. Reprogramming lowers the dominant barrier gap in the paired two-well construction (Test 5 in that script), consistent with the qualitative Kramers prediction that smaller ΔV increases $\exp(-\Delta V/D)$.

Test harness (15/15). All fifteen checks are implemented in numerical_validation.py: three basin-count tests; three dominant-barrier positivity tests; two analytic Kramers-factor checks on the two-state landscape; three spectral-compression tests ($\rho > 1$); one reprogramming barrier-drop test; and three "small N^*/N" tests. They are **not** Monte Carlo path estimates, embedding-rotation checks, or regression clones of individual platonic.py lemmas.

Sensitivity to diffusion. Extending the harness to scan D (or to simulate sample paths) is straightforward but **not** part of the current script; qualitative Kramers scaling is illustrated only through the explicit $\exp(-\Delta V/D)$ evaluation at fixed $D = 1$ on the two-state baseline.

Parameter ethics. All synthetic potentials use globally convex tails to ensure integrability of Boltzmann factors in stationary measures. This is a modeling choice that aids theorem hypotheses but does not claim realism for unbounded gene expression coordinates.

5. Cross-Domain Connections

Protein folding. Folding funnels are historical prototypes for Waddington diagrams. The same ρ–N^* language applies: native basins are sharp minima with small N^* relative to conformational cardinality.

Fitness landscapes. Cell-fate decisions and evolutionary walks both involve potential-like surfaces over discrete or continuous genotypes. Bridge lemmas align barrier crossing in V with adaptive-walk trapping statistics under appropriate discretizations.

Neural manifolds (preview). Low-dimensional neural dynamics sometimes exhibit multistable attractors reminiscent of lineage choices. Connecting those models to the Latent potential picture here would require an explicit embedding and barrier identification; we defer that to dedicated neuroscience topics.

6. Discussion

The Latent Waddington construction makes canalization measurable: small N^*/N is the quantitative echo of Waddington's "creodes." The reprogramming experiments show that barrier surgery

moves transition rates in the direction Kramers theory demands, supporting the use of Latent potentials as more than visualization devices.

Limitations include the reliance on synthetic potentials and the need for careful identification of Φ from finite single-cell samples. Real systems exhibit history dependence, cell–cell coupling, and active processes outside gradient-flow models.

Future work will fit Latent potentials from RNA velocity fields and lineage tracing, incorporate cell-cycle phase as an explicit slow variable, and test ρ shifts during directed differentiation time courses.

From a modeling standpoint, the most productive next step is mixed observability: some experiments provide dense time courses (optimal for drift estimation), others provide only snapshots (better for ρ via covariance). The Latent framework treats these as two views of the same compressed coordinates and is designed to merge them under explicit uncertainty quantification rather than ad hoc stitching.

Non-claims. We do not yet claim identification of V from snapshot data alone without dynamical information; the present numerics assume known potentials for validation.

Open paths. RNA velocity provides a natural drift field that may replace pure gradient flow; extending the Latent theorems to non-gradient systems with skew-symmetric components is ongoing. Hybrid models (gradient plus curl) are expected to increase N^* modestly while leaving ρ informative as a marginal compressibility score.

During the preparation of this work the author used large language models to assist with manuscript drafting, literature alignment, and coding assistance. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

1. Waddington, C. H. *The Strategy of the Genes.* Allen & Unwin, 1957.
2. Huang, S. "Reprogramming cell fates: reconciling rarity with robustness." *BioEssays*, 31(5):546–560, 2009.
3. Moris, N. et al. "Transition states and cell fate decisions in epigenetic landscapes." *Nat. Rev. Genet.*, 2016.
4. Kramers, H. A. "Brownian motion in a field of force." *Physica*, 1940.
5. Bryngelson, J. D. et al. "Funnels, pathways, and the energy landscape of protein folding." *Proteins*, 1995.
6. Kauffman, S. A. *The Origins of Order.* Oxford University Press, 1993.
7. La Manno, G. et al. "RNA velocity of single cells." *Nature*, 2018.
8. Lange, M. et al. "CellRank for directed single-cell fate mapping." *Nat. Methods*, 2022.
9. Freidlin, M. I. & Wentzell, A. D. *Random Perturbations of Dynamical Systems.* Springer, 2012.
10. Villani, C. *Hypocoercivity.* American Mathematical Society, 2009.
11. Berglund, N. & Gentz, B. *Noise-Induced Phenomena in Slow-Fast Dynamical Systems.* Springer, 2006.

12. Takahashi, K. & Yamanaka, S. “Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.” *Cell*, 2006.