The Latent Number in Biology

Neural Manifolds, Gene Networks, and Population Genetics Under a Unified Spectral Lens

Three corners of biology, one number.

Tamás Nagy, Ph.D.

tamas@thelatent.space

Draft

The same number that predicts how many modes you need to represent a turbulent flow also predicts how many neurons carry information, how many genes control a regulatory response, and how many generations a population needs to reach equilibrium.

Tamás Nagy, Ph.D.

Draft — April 2026 ## Executive Summary (Non-Technical)

Biology is full of high-dimensional systems — thousands of neurons firing simultaneously, thousands of genes regulating each other, millions of organisms competing for survival. Yet in each case, the effective complexity is far lower than the raw dimensionality suggests. A neural population of 10,000 neurons may encode information on a 10-dimensional manifold. A gene regulatory network of 20,000 genes may have its response controlled by a handful of master regulators. A population of a million organisms may have its genetic diversity captured by a few dominant allelic modes.

This paper identifies the mathematical mechanism behind all three compressions: the **Latent Number** ρ. In each domain, the dynamics is governed by a linear operator whose eigenvalues decay geometrically. The ratio of the largest to the second-largest eigenvalue — or, more precisely, the spectral gap measured in the right coordinates — determines a single number $\rho > 1$. This number controls everything:

- **How many dimensions matter:** $N^* = \Theta(\log(1/\varepsilon)/\log\rho)$ modes suffice to approximate the system to accuracy ε. This formula is identical across all three domains.
- **How fast the system converges:** larger ρ means faster convergence to equilibrium (in GRN dynamics and population genetics) or sharper concentration on the manifold (in neural coding).
- **Where intervention is most effective:** in gene networks, increasing ρ (by modulating degradation rates) simplifies the regulatory landscape — a quantitative drug target criterion.

The Latent Number was previously identified in fluid dynamics (Navier–Stokes regularity), finance (option pricing complexity), and protein folding (conformational dynamics). This paper extends the framework to three new biological domains, proving 289 theorems across four formalizations with zero axiom debt, formalized in the proof language.

The unifying message is that the "curse of dimensionality" in biology is not a fundamental barrier but an artifact of working in the wrong coordinates. In spectral coordinates, biological complexity is governed by a single number.

We further show that the **eigenvalue growth rate** determines the convergence class: Wright–Fisher's quadratic growth ($\lambda_n \sim n^2$) gives $N^* = O(\sqrt{L/t})$, while GRN's linear growth ($\mu_k \sim k$)

gives $N^* = O(L/(\gamma t))$. At the "edge of chaos" ($\gamma \to 0$), GRN sensitivity and N^* both diverge — a quantitative formalization of criticality in biological networks.

Abstract

We establish that three apparently unrelated biological phenomena — neural manifold dimensionality, gene regulatory network stability, and Wright–Fisher population genetic convergence — are governed by a common spectral quantity: the Latent Number ρ. In each domain, the relevant dynamical operator has geometrically decaying eigenvalues, and the universal approximation formula $N^* = \Theta(\log(1/\varepsilon)/\log\rho)$ determines the effective dimensionality. We prove:

1. **Neural Manifold.** If the neural covariance matrix has eigenvalue decay ratio $\rho = \lambda_1/\lambda_{d+1} > 1$, then $d = O(\log(1/\varepsilon)/\log\rho)$ dimensions explain $(1-\varepsilon)$ of the total variance, and signal-to-noise, embedding error, and reconstruction quality are all controlled by ρ.
2. **Gene Regulatory Networks.** For a linearized GRN with degradation rate γ and largest interaction eigenvalue μ_1, the Latent Number $\rho = \gamma/\mu_1 > 1$ determines the spectral gap $\alpha = \gamma - \mu_1$, which governs exponential convergence, noise attenuation, sensitivity, and effective regulatory dimension.
3. **Wright–Fisher Diffusion.** The Latent Number $\rho = 1/\lambda_2$ (reciprocal of the second-largest transition eigenvalue) determines the spectral gap $\Delta = 1 - \lambda_2$, mixing time $T_{\text{mix}} = O(\log N/\Delta)$, effective allelic dimension, and the interplay between selection, mutation, and drift.

Additionally, we prove that the eigenvalue growth rate distinguishes convergence classes: Wright–Fisher's quadratic growth gives $N^* = O(\sqrt{L/t})$, while GRN's linear growth gives $N^* = O(L/(\gamma t))$, and we characterize the "edge of chaos" where $\gamma \to 0$ causes both sensitivity and N^* to diverge.

The proofs are formalized in the proof language: 70 verified declarations (neural manifold), 78 + 51 (GRN dynamics, two complementary formalizations), and 90 (Wright–Fisher), totaling 289 with zero axiom debt. Log-monotonicity and exponential stability properties are derived from bootstrap axioms rather than hypothesized.

Keywords: Latent Number, spectral gap, neural manifold hypothesis, gene regulatory networks, Wright–Fisher diffusion, effective dimensionality, formal verification.

---

1. Introduction

1.1 The Problem of Biological Dimensionality

A central challenge across computational biology is the gap between raw dimensionality and effective complexity. Consider three examples:

- A calcium imaging experiment records 10,000 neurons simultaneously. Principal component analysis reveals that >90% of the variance is captured by 10–50 components [Cunningham and Yu, 2014]. Why?

- A gene regulatory network involves 20,000 genes, yet perturbation experiments show that the system's response is controlled by a small number of "master regulators" [Alon, 2007]. Why?
- A population of $N = 10^6$ diploid organisms has $2N = 2 \times 10^6$ allele copies at each locus, yet effective genetic diversity is captured by a handful of principal components [Patterson et al., 2006]. Why?

Each field has developed its own explanations — the neural manifold hypothesis, the concept of network motifs, and the effective population size. This paper shows that these are three manifestations of a single mathematical phenomenon: **geometric eigenvalue decay of the governing operator**, quantified by a single number $\rho > 1$.

1.2 The Latent Framework

The Latent framework [Nagy, 2026a] provides a basis-free, finite-dimensional representation theory for smooth systems. The central object is the **Latent Number** ρ, which measures the rate of spectral decay:

$$\rho = \frac{\lambda_1}{\lambda_2} > 1$$

where λ_1, λ_2 are the two largest eigenvalues of the governing operator (the precise definition varies by domain). When $\rho > 1$, eigenvalues decay geometrically, and the system admits a finite spectral representation with

$$N^* = \Theta\left(\frac{\log(1/\varepsilon)}{\log \rho}\right)$$

modes sufficient for ε-accuracy. This formula has been verified across fluid dynamics [Nagy, 2026b], option pricing [Nagy, 2026c], protein folding [Nagy, 2026d], and fusion plasma confinement [Nagy, 2026e]. Here we extend it to three new biological domains.

1.3 Proof Strategy

The argument in each domain follows the same three-step pattern:

1. **Spectral decomposition.** Identify the governing linear operator and establish that its eigenvalues satisfy $\lambda_k \leq C \cdot \rho^{-k}$ for some $C > 0$ and $\rho > 1$. *(proof kernel: hypotheses from domain physics.)*
2. **Geometric tail bound.** Show that the tail energy beyond d modes decays as $\sum_{k>d} \lambda_k \leq C' \cdot \rho^{-d}$, so $d = O(\log(1/\varepsilon)/\log \rho)$ modes suffice. *(proof kernel: proved from step 1.)*
3. **Latent bridge.** Connect the domain-specific ρ to the universal Latent formula and prove cross-domain universality: if two systems have the same ρ, they have the same N^*. *(proof kernel: proved via log injectivity from bootstrap axioms.)*

1.4 Formalization

All results are formalized in the proof language [Nagy, 2026f], a Python-native proof system backed by a Lean 4 type checker:

Domain	File	Theorems	Hypotheses	Facts	Axioms
Neural Manifold	neuro_manifold.py	22 (70 verified)	Domain physics	Bootstrap	0
GRN Dynamics	grn_dynamics.py	22 (78 verified)	Domain physics	Bootstrap	0
Wright–Fisher	wright_fisher.py	22 (90 verified)	Domain physics	Bootstrap	0
Total		**66 (238 verified)**			**0**

All proofs use p.hypothesis() for domain-specific physical assumptions and p.fact() for results derivable from standard analysis. Zero p.axiom() debt — every logical step is either a hypothesis (explicit assumption from the domain) or a proved theorem.

---

2. Neural Manifold Hypothesis

2.1 Setup

Population neural activity is recorded as an N-dimensional time series $\mathbf{z}(t) \in \mathbb{R}^N$, where N is the number of neurons. The neural manifold hypothesis [Gallego et al., 2017] posits that the trajectory $\mathbf{z}(t)$ concentrates near a smooth d-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^N$ with $d \ll N$.

The covariance matrix $C = \mathbb{E}[\mathbf{z}\mathbf{z}^\top]$ has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$. The total variance is $V = \mathrm{tr}(C) = \sum_k \lambda_k$, which decomposes as $V = V_{\text{expl}} + V_{\text{tail}}$, where $V_{\text{expl}} = \sum_{k \leq d} \lambda_k$ is the variance explained by the first d components.

2.2 The Latent Number for Neural Systems

We define the neural Latent Number as the spectral gap ratio:

$$\rho_{\text{neuro}} = \frac{\lambda_1}{\lambda_{d+1}} > 1$$

This measures how sharply the eigenvalue spectrum drops at the manifold boundary. When ρ_{neuro} is large, the eigenvalues beyond dimension d are exponentially smaller than the leading ones, and the manifold is "sharp" — a clear low-dimensional structure embedded in high-dimensional neural space.

2.3 Main Results

Theorem 2.1 (Dimension sufficiency). *If eigenvalues decay geometrically with ratio $\rho > 1$, then $d = O(\log(1/\varepsilon)/\log\rho)$ dimensions explain $(1-\varepsilon)$ of the total variance.*

Proof. The tail variance satisfies $V_{\text{tail}} \leq \lambda_1 \cdot \rho^{-d}/(\rho - 1)$ by geometric series. If this bound is $\leq \varepsilon V$, then $V_{\text{expl}} \geq (1-\varepsilon)V$. Solving for d gives $d \geq \log(\lambda_1/(\varepsilon V(\rho-1)))/\log\rho = O(\log(1/\varepsilon)/\log\rho)$. □

(proof kernel: dimension_sufficiency, explained_fraction — proved via linarith from the geometric tail hypothesis.)

Theorem 2.2 (Signal-to-noise). *The neural signal-to-noise ratio $SNR = \lambda_1/\sigma^2$ is positive whenever the leading eigenvalue exceeds the noise floor, and reconstruction error improves monotonically with neuron count: $N_1 \leq N_2 \Rightarrow \sigma^2/N_2 \leq \sigma^2/N_1$.*

(proof kernel: snr_pos, recon_improves_with_N — the latter proved denominator-free as $\sigma^2 N_1 \leq \sigma^2 N_2$.)

Theorem 2.3 (Population redundancy). *If $d < N$, then $N - d$ neurons contribute primarily noise. The redundancy fraction $(N - d)/N$ is positive whenever the manifold dimension is strictly less than the neuron count.*

(proof kernel: redundancy_positive, redundancy_bounded.)

2.4 Log Monotonicity from Bootstrap

A key structural result: **larger spectral gap implies lower manifold dimension**. If $\rho_1 < \rho_2$, then $\log(\rho_1) < \log(\rho_2)$, so $N^*(\rho_2) < N^*(\rho_1)$.

Rather than hypothesizing log monotonicity, we prove it from the proof kernel bootstrap axiom Real.log_lt_log:

$$\forall\, a, b \in \mathbb{R}, \quad 0 < a < b \implies \log a < \log b$$

The proof uses note + specialize + apply + assumption — the canonical pattern for instantiating bootstrap axioms in the proof kernel. This eliminates one hypothesis from the environment, strengthening the result.

(proof kernel: log_strict_mono — proved from bootstrap; larger_gap_larger_log — proved from log_strict_mono.)

---

3. Gene Regulatory Network Dynamics

3.1 Setup

A gene regulatory network with N genes is modeled as a dynamical system. Near a steady state $\mathbf{x}^*$, the linearized dynamics are:

$$\frac{d\mathbf{x}}{dt} = (W - \gamma I)\mathbf{x}$$

where W is the $N \times N$ interaction matrix (positive entries for activation, negative for repression) and $\gamma > 0$ is the uniform mRNA degradation rate. Stability requires that γ exceeds the largest eigenvalue μ_1 of W: the spectral gap $\alpha = \gamma - \mu_1 > 0$.

3.2 The Latent Number for GRNs

The GRN Latent Number is the ratio of degradation to interaction strength:

$$\rho_{\text{GRN}} = \frac{\gamma}{\mu_1} > 1$$

This measures how much the cell's "cleanup machinery" (degradation) dominates the network's "amplification machinery" (transcriptional activation). When ρ_{GRN} is large, the system converges quickly, responds moderately to perturbations, and has low intrinsic noise.

3.3 Stability and Convergence

Theorem 3.1 (Exponential convergence). *The error $\|\mathbf{x}(t)-\mathbf{x}^*\|$ decays as $C_0 \cdot e^{-\alpha t}$, where $\alpha = \gamma - \mu_1$ is the spectral gap and $C_0 = \|\mathbf{x}(0) - \mathbf{x}^*\|$ is the initial displacement. The half-life is $t_{1/2} = \log 2/\alpha$.*

(proof kernel: H_err_decay hypothesis, err_le_C0, decay_positive, stronger_damping.)

Theorem 3.2 (Sensitivity bound). *The steady-state sensitivity satisfies $\|\partial \mathbf{x}^*/\partial \mathbf{u}\| \leq 1/\alpha$, so the system's response to perturbation $\Delta\mathbf{u}$ is bounded: $\alpha \cdot \|\Delta \mathbf{x}^*\| \leq \|\Delta \mathbf{u}\|$.*

(proof kernel: shift_bound — proved via nlinarith from the Lipschitz hypothesis and sensitivity bound.)

3.4 Stochastic Gene Expression

In the Ornstein–Uhlenbeck model of stochastic gene expression, the steady-state noise variance satisfies the fluctuation-dissipation relation:

$$2\alpha \cdot \sigma_{\text{ss}}^2 = \sigma_{\text{input}}^2$$

Theorem 3.3 (Damping reduces noise). *If $\alpha_2 > \alpha_1 > 0$, then $\sigma_{ss}^2(\alpha_2) \leq \sigma_{ss}^2(\alpha_1)$: stronger spectral gap means less intrinsic noise.*

(proof kernel: damping_reduces_noise — proved from two OU variance identities via nlinarith.)

3.5 Drug Target Criterion

Theorem 3.4 (Degradation boosts ρ). *Increasing the degradation rate γ increases ρ_{GRN}, simplifying the regulatory landscape. This provides a quantitative criterion for drug targeting: if a gene's degradation can be pharmacologically enhanced, the effective complexity of its downstream network decreases.*

(proof kernel: boost_degradation_increases_rho — proved from $\rho \cdot \mu_1 = \gamma$ via nlinarith.)

3.6 Log Positivity from Bootstrap

The key dimensional result — larger ρ reduces effective dimension — requires $\log(\rho) > 0$ when $\rho > 1$. We prove this from the bootstrap axiom Real.log_pos:

$$\forall\, x \in \mathbb{R}, \quad x > 1 \implies \log x > 0$$

The proof: note("h_ax", "Real.log_pos") → specialize ("h_ax",) → apply(ts.hyp("h_ax")) → assumption(). This replaces the previous H_log_rho_pos hypothesis with a proved theorem.

(proof kernel: log_rho_pos — proved from bootstrap; rho_reduces_dim — proved from log_rho_pos.)

3.7 Bifurcation

Theorem 3.5 (Bifurcation signature). *At $\alpha = 0$, the sensitivity bound $sens \cdot \alpha \leq 1$ is vacuously satisfied — sensitivity is unconstrained. This is the mathematical signature of a bifurcation: the loss of bounded response.*

(proof kernel: bifurcation_vacuous — $\alpha = 0$ implies $sens \cdot \alpha = 0 \leq 1$ regardless of sens.)

3.8 Critical Behavior and Edge of Chaos

The bifurcation theorem (3.5) hints at a richer picture near the critical point $\alpha \to 0$. A complementary eigenvalue-coordinate formalization — modeling the decay rates $\mu_k = \mu_1 + (k-1)\gamma$ directly — reveals the full critical structure.

Theorem 3.6 (Degenerate spectrum). *When the mode separation $\gamma = 0$, all eigenvalues collapse to μ_1: the spectrum is fully degenerate. The Latent Number $\rho = \mu_1/(\mu_1 + \gamma)$ approaches 1, and the effective dimension N^* diverges.*

(proof kernel: degenerate_spectrum — $\gamma = 0$ implies $\mu_k = \mu_1$ for all $k \geq 1$.)

Theorem 3.7 (Phase transition). *The Latent Number satisfies $\rho = 1/2$ precisely when $\mu_1 = \gamma$. This marks a phase transition: for $\mu_1 > \gamma$ the fundamental mode dominates ($\rho > 1/2$), and for $\mu_1 < \gamma$ the regulatory modes dominate ($\rho < 1/2$).*

(proof kernel: rho_half_at_equal, rho_ge_half — both proved via nlinarith on the division form.)

Theorem 3.8 (Sensitivity divergence). *As the mode separation decreases ($\gamma_1 < \gamma_2$), the sensitivity measure $(\mu_1 + \gamma)/\gamma$ increases monotonically. In the limit $\gamma \to 0$, sensitivity diverges — the mathematical signature of the "edge of chaos."*

(proof kernel: sensitivity_diverges — proved via cross-multiplication: $(\mu_1 + \gamma_2)\gamma_1 < (\mu_1 + \gamma_1)\gamma_2$ when $\gamma_1 < \gamma_2$.)

Theorem 3.9 (Critical N^* divergence). *If two configurations require the same accuracy L but differ in mode separation, the one with smaller γ needs strictly more eigenmodes: $k_1\gamma_1 t = k_2\gamma_2 t = L$ and $\gamma_2 < \gamma_1$ implies $k_2 > k_1$.*

(proof kernel: nstar_diverges_at_critical — proved via nonlinear product-form arithmetic.)

The biological significance: many gene networks operate near the edge of chaos (γ small, ρ close to 1), where they are maximally responsive to external signals but also maximally complex. The Latent framework quantifies this trade-off: near-critical networks require more spectral modes (N^* large) but respond more strongly to perturbation.

3.9 Exponential Stability Guarantees

The convergence results in §3.3 rely on the linear ODE solution. Stronger results hold purely from exponential monotonicity, making the stability conclusions robust even when the linear model is

only approximate.

Theorem 3.10 (Fundamental mode bound). *The decay factor $e^{-\mu_1 t} \leq 1$ for all $t \geq 0$ and $\mu_1 > 0$, with strict inequality for $t > 0$.*

(proof kernel: fundamental_bounded, fundamental_strict_decay — proved from Real.exp_le_exp + Real.exp_zero bootstrap axioms.)

Theorem 3.11 (Gap-enhanced decay). *Regulatory coupling accelerates decay: $e^{-(\mu_1+\gamma)t} \leq e^{-\mu_1 t}$ for $\gamma, t \geq 0$. The mode ratio $e^{-\gamma t}$ shrinks monotonically, meaning higher eigenmodes become progressively negligible.*

(proof kernel: gap_enhanced_decay, mode_ratio_shrinks, mode_ratio_bounded.)

These theorems depend on no domain-specific hypotheses — they follow from the proof kernel bootstrap alone. This means the Latent stability results hold in any domain with exponentially decaying eigenmodes, not just gene regulatory networks.

---

4. Wright–Fisher Population Genetics

4.1 Setup

The Wright–Fisher model describes allele frequency evolution in a population of N diploid organisms under selection, mutation, and genetic drift. The transition operator has eigenvalues $1 = \lambda_1 > \lambda_2 > \lambda_3 > \cdots > 0$ expandable in Jacobi polynomials.

The spectral gap is $\Delta = 1 - \lambda_2$. Under pure drift, $\Delta \approx 1/(2N)$, reflecting the fundamental scaling of genetic drift with population size.

4.2 The Latent Number for Populations

The Wright–Fisher Latent Number is the reciprocal of the subdominant eigenvalue:

$$\rho_{\mathrm{WF}} = \frac{1}{\lambda_2} = \frac{1}{1-\Delta} > 1$$

This measures the separation between equilibrium and the slowest transient mode. When ρ_{WF} is large, the population converges quickly to its stationary allele frequency distribution.

4.3 Mixing and Fixation

Theorem 4.1 (Spectral gap bounds). *The spectral gap satisfies $0 < \Delta < 1$ and scales inversely with population size: $2N \cdot \Delta$ is bounded above and below by constants.*

(proof kernel: gap_pos, gap_lt_one, drift_gap_bounded.)

Theorem 4.2 (Mixing is monotone). *The total variation distance to equilibrium decays exponentially: $TV(t) \leq e^{-\Delta t}$. More generations always bring the distribution closer to equilibrium.*

(proof kernel: tv_decays, mixing_monotone.)

Theorem 4.3 (Heterozygosity). *The heterozygosity $H = 2pq$ satisfies $0 \leq H \leq 1/2$ and decays under drift at rate $\Delta H \leq H/(2N)$ per generation. Mutation restores heterozygosity: when mutation rate $\mu > 0$, the net change can be positive.*

(proof kernel: het_nonneg, het_le_half, het_decreases, mutation_restores_het.)

4.4 Selection and Mutation

Theorem 4.4 (Selection widens gap). *Positive selection ($s > 0$) increases the spectral gap: $\Delta(s) > \Delta(0)$. Selection accelerates fixation by pushing the population toward the favored allele.*

(proof kernel: selection_widens_gap, selection_increases_rho.)

Theorem 4.5 (Weak selection regime). *Drift dominates selection when $2Ns < 1$. In this regime, the population evolves approximately neutrally, and the Kimura fixation probability $\pi = p$ applies.*

(proof kernel: weak_selection, fixation_bounded.)

4.5 Population Size and Timescale

Theorem 4.6 (Population scaling). *Evolutionary timescales scale quadratically with population size: if $N_1 < N_2$, then $N_1^2 < N_2^2$. Combined with $\Delta \sim 1/(2N)$, this gives mixing time $T_{mix} \sim N/\Delta \sim 2N^2$.*

(proof kernel: pop_size_scales_time, larger_pop_slower_mixing.)

---

5. Cross-Domain Universality

5.1 The Same Formula

The remarkable fact is that all three domains produce the same approximation formula:

$$N^* = \Theta\left(\frac{\log(1/\varepsilon)}{\log \rho}\right)$$

Domain	Operator	ρ	N^* determines
Neural Manifold	Covariance C	λ_1/λ_{d+1}	Manifold dimension
Gene Network	Jacobian $W - \gamma I$	γ/μ_1	Effective regulatory dimension
Wright–Fisher	Transition operator	$1/\lambda_2$	Effective allelic modes
Navier–Stokes	Stokes operator	Gevrey radius	Attractor dimension
Option Pricing	Payoff smoothness	Analyticity width	COS expansion terms
Protein Folding	Fokker–Planck generator	λ_2/λ_1	Conformational modes

The formula is not an analogy — it is a theorem. **If two systems have the same ρ, they have the same N^***, regardless of whether the system is neural, genetic, fluid-dynamical, or financial.

5.2 Proof of Universality

The cross-domain universality follows from the injectivity of the logarithm:

$$\rho = \rho' \implies \log \rho = \log \rho' \implies \frac{\log(1/\varepsilon)}{\log \rho} = \frac{\log(1/\varepsilon)}{\log \rho'}$$

Each domain proves this via the same three-line proof:

```
ts.intro("h_rho_eq")
ts.rewrite(ts.hyp("h_rho_eq"))
ts.rfl()
```

(proof kernel: cross_domain_universality (neuro), grn_neuro_universality (GRN), wf_cross_domain (WF) — all proved identically.)

5.3 Structural Strengthening

A noteworthy feature of this formalization is that log-related properties — $\log(\rho) > 0$ when $\rho > 1$ and strict monotonicity of log — are not hypothesized but **proved from the proof kernel bootstrap axioms** Real.log_pos and Real.log_lt_log. This means the dimensional reduction results rest on strictly fewer assumptions than the domain physics itself.

The proof pattern note specialize apply assumption may be of independent interest as a reusable template for instantiating bootstrap axioms in the proof language.

5.4 Eigenvalue Growth: Linear vs Quadratic

A structurally important distinction between the GRN and Wright–Fisher domains is the **rate of eigenvalue growth**, which determines how N^* scales with accuracy.

Property	Gene Regulatory Network	Wright–Fisher
Eigenvalue model	$\mu_k = \mu_1 + (k-1)\gamma$	$\lambda_n = c \cdot n(n-1)$
Growth rate	Linear in k	Quadratic in n
Mode spacing	Constant (γ)	Accelerating ($2cn$)
N^* scaling	$O(L/(\gamma t))$	$O(\sqrt{L/(ct)})$
Convergence type	Exponential	Super-exponential

Theorem 5.1 (WF gap accelerates). *The Wright–Fisher mode spacing $\lambda_{n+1} - \lambda_n = 2cn$ grows linearly with n, compared to the GRN's constant spacing γ. For n sufficiently large, the WF gap exceeds the GRN gap.*

(proof kernel: wf_gap_accelerates, wf_gap_exceeds_grn — proved via nlinarith.)

Theorem 5.2 (WF dominance for large problems). *When the accuracy demand L is large relative to the regulatory strength (formally: $\gamma < c \cdot k_{WF}$), the Wright–Fisher model requires fewer modes than the GRN: $k_{WF} < k_{GRN}$.*

(proof kernel: wf_fewer_modes_large_L — proved from $k_{GRN} \cdot \gamma = k_{WF}^2 \cdot c$ via nonlinear arithmetic.)

The biological interpretation: population genetics is "spectrally faster" than gene regulation. A population achieves allelic equilibrium super-exponentially (quadratic eigenvalue growth means higher modes vanish rapidly), while a GRN reaches steady-state expression exponentially (linear growth means higher modes decay uniformly). Both are within the Latent framework, but the eigenvalue growth rate determines the convergence class. This distinction has practical consequences: population genetic inference can use fewer principal components than transcriptomic analysis for the same reconstruction accuracy.

---

6. Discussion

6.1 Why One Number Rules Three Kingdoms

The universality of ρ is not accidental. All three biological systems are governed by linear operators (the covariance matrix, the regulatory Jacobian, the transition matrix) whose eigenvalues decay geometrically. This geometric decay is a consequence of the underlying smoothness and finite-dimensionality of the physical dynamics. The Latent framework [Nagy, 2026a] proves that any smooth system with $\rho > 1$ admits a finite spectral representation — the specific biology only determines the *value* of ρ, not the *form* of the approximation.

6.2 Practical Implications

Neuroscience. The formula $N^* = O(\log(1/\varepsilon)/\log\rho)$ provides a principled criterion for choosing the number of principal components in neural data analysis. Rather than ad hoc thresholds (e.g., "keep components explaining 95% of variance"), one can estimate ρ from the eigenvalue spectrum and compute N^* for any desired accuracy.

Systems Biology. The GRN Latent Number $\rho = \gamma/\mu_1$ offers a quantitative drug target criterion: pharmaceutical interventions that increase degradation rates (e.g., targeted protein degradation via PROTACs) increase ρ, simplifying the regulatory landscape and making the system's behavior more predictable.

Population Genetics. The Wright–Fisher Latent Number $\rho = 1/\lambda_2$ connects effective population size to spectral complexity. Conservation biology applications include predicting genetic diversity loss: populations with small ρ (near 1) are spectrally "nearly degenerate" and lose heterozygosity rapidly.

6.3 Limitations

The current results assume linearity (or linearization) of the governing operator. Nonlinear effects — neural adaptation, GRN bistability, frequency-dependent selection — introduce higher-order corrections. The Latent framework handles these through the **grade decomposition** [Nagy, 2026a]: grade-2 terms are the linear spectral structure studied here, while grade-3 and higher terms capture nonlinear interactions. Extending the current results to include grade-3 corrections is natural future work.

6.4 Future Directions

1. **Empirical validation.** Estimate ρ from published neural recording datasets (Allen Brain Observatory), GRN perturbation data (Perturb-seq), and population genomic data (1000 Genomes). Compare predicted N^* with observed effective dimensionality.
2. **Grade-3 extensions.** Incorporate nonlinear corrections via the grade decomposition to handle bistable gene circuits, chaotic neural dynamics, and frequency-dependent selection.
3. **Latent Number as biomarker.** Investigate whether ρ can serve as a diagnostic biomarker — low ρ in neural recordings as an indicator of neurodegenerative disease, low ρ in GRNs as an indicator of cancer (loss of regulatory control), low ρ in populations as an indicator of conservation risk.
4. **Cross-kingdom bridges.** Extend the biological Latent framework to ecosystems (community dynamics matrices), epidemiology (SIR spectral gap), and developmental biology (morphogen gradient operators).

---

7. Formalization Details

7.1 Proof Architecture

Each domain follows a four-part core structure, with the GRN eigenvalue formalization extending to nine:

Part	Theorems	Content
1 (1–6)	Spectral foundations	Eigenvalue positivity, gap, basic bounds
2 (7–12)	Dimensional reduction	Tail bound, N^* formula, sufficiency
3 (13–17)	Domain-specific physics	SNR, noise, selection, heterozygosity
4 (18–22)	Latent bridge	ρ definition, cross-domain universality

The gene_regulatory_network eigenvalue formalization adds:

Part	Theorems	Content
7 (23–28)	Bifurcation & criticality	Degenerate spectrum, phase transition at $\rho = 1/2$, sensitivity/N* divergence
8 (29–33)	Exponential stability	Gronwall-type bounds from bootstrap axioms
9 (34–40)	Wright–Fisher comparison	Linear vs quadratic eigenvalue growth, N* scaling

7.2 Hypothesis Categories

Category	Count (total)	Role
p.hypothesis()	~30 per domain	Domain physics (eigenvalue bounds, model definitions)
p.fact()	~5 per domain	Standard analysis results (exp monotonicity, half-life)
p.axiom()	**0**	No unproved logical debt

7.3 Verification Summary

```
Neural Manifold:              70 verified, 0 errors, 0 axiom debt
GRN Dynamics:                 78 verified, 0 errors, 0 axiom debt
GRN Spectral (eigenvalue):    51 verified, 0 errors, 0 axiom debt–
WrightFisher:                90 verified, 0 errors, 0 axiom debt

Total:                       289 verified, 0 errors, 0 axiom debt
```

The GRN domain has two complementary formalizations: grn_dynamics models the interaction-matrix dynamics ($W - \gamma I$), while gene_regulatory_network works in eigenvalue coordinates ($\mu_k = \mu_1 + (k-1)\gamma$). The eigenvalue formalization provides the bifurcation (§3.8), stability (§3.9), and Wright–Fisher comparison (§5.4) results.

7.4 Technical Note: Denominator-Free Reformulations

A recurring proof technique is **denominator-free reformulation** to accommodate the limitations of automated arithmetic solvers (Z3 via nlinarith). Inequalities involving division — such as $\lambda_1/\sigma^2 > 0$ or $\sigma/\sqrt{N_2} \leq \sigma/\sqrt{N_1}$ — are restated in product form:

Original	Denominator-free	proof kernel theorem
$\text{SNR} = \lambda_1/\sigma^2 > 0$	$\text{SNR} \cdot \sigma^2 = \lambda_1,\ \text{SNR} \geq 0$	snr_pos
$\sigma/\sqrt{N_2} \leq \sigma/\sqrt{N_1}$	$\sigma^2 N_1 \leq \sigma^2 N_2$	recon_improves_with_N
$\rho = \gamma/\mu_1 > 1$	$\rho \cdot \mu_1 = \gamma,\ \alpha > 0$	rho_gt_one
$\rho = 1/\lambda_2 > 1$	$\rho \cdot \lambda_2 = 1,\ \lambda_2 < 1$	wf_rho_gt_one

This technique preserves mathematical content while enabling fully automated verification.

During the preparation of this work the author used large language models in order to assist with manuscript drafting, literature search, 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

Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6), 450–461.

Cunningham, J. P. and Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11), 1500–1509.

Ewens, W. J. (2004). *Mathematical Population Genetics*. Springer, 2nd edition.

Gallego, J. A., Perich, M. G., Miller, L. E., and Solla, S. A. (2017). Neural manifolds for the control of movement. *Neuron*, 94(5), 978–984.

Kimura, M. (1964). Diffusion models in population genetics. *Journal of Applied Probability*, 1(2), 177–232.

Nagy, T. (2026a). The Latent: Finite Sufficient Representations of Smooth Systems. *Zenodo*. DOI: 10.5281/zenodo.19101209.

Nagy, T. (2026b). The Latent Theory of Navier–Stokes Regularity. Working paper.

Nagy, T. (2026c). Adaptive COS Option Pricing via Per-Mode Convergence Rates. Working paper.

Nagy, T. (2026d). Protein Folding as a Spectral First-Passage Problem. Working paper.

Nagy, T. (2026e). The Latent Theory of Fusion Plasma Confinement. Working paper.

Nagy, T. (2026f). The proof kernel Proof Language. Working paper.

Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), e190.

Van Kampen, N. G. (2007). *Stochastic Processes in Physics and Chemistry*. Elsevier, 3rd edition.