The Spectral Zeta Function of Markov Generators

Universal Diagnostics for Complex Dynamical Systems

One function. Four observables. Any generator.

Tamás Nagy, Ph.D.

tamas@thelatent.space

Draft

Abstract

We introduce the spectral zeta function $\zeta_L(s) = \mathrm{Tr}(L^{-s})$ of the generator L of a continuous-time Markov process as a universal diagnostic for complex dynamical systems. From a single analytic function of one complex variable s, we extract four computable observables: (i) the mean first-passage time ($s = 1$), (ii) pathway entropy (a Shannon entropy over mode relaxation times), (iii) effective spectral dimension (via the Weyl law for eigenvalue asymptotics), and (iv) exponential purity of the relaxation (via the heat kernel trace). These four quantities characterize, respectively, the timescale, diversity, dimensionality, and simplicity of the system's dominant dynamics — from a single matrix.

We prove the formulas in general, establish their relationships, and validate all four on the Fokker–Planck generators of 14 protein domains from molecular dynamics (mdCATH), where the correlations reach $r = 0.91$ (Weyl law) and $r = 0.88$ (mean time formula). We then show how the same diagnostics apply without modification to credit portfolio dynamics, plasma confinement, epidemic spreading, quantum decoherence, and chemical reaction networks. The spectral zeta function provides a domain-agnostic language for comparing dynamical complexity across fields.

---

1. Introduction

Complex systems across science share a common mathematical substrate: a state space, a dynamics, and a generator. In continuous-time Markov processes, the generator L is a matrix (or operator) whose eigenvalues $\{\lambda_k\}$ encode the complete kinetic behavior of the system. The spectral gap $|\lambda_1|$ — the smallest non-zero eigenvalue magnitude — determines the slowest relaxation timescale and has been studied extensively in probability theory, statistical mechanics, and machine learning.

But the spectral gap is one number. The full eigenvalue spectrum $\{\lambda_k\}_{k=1}^{N}$ contains far more information: the distribution of relaxation timescales, the effective dimensionality of the dynamics, the diversity of transition pathways, and the degree to which the system's long-time behavior is dominated by a single mode.

We propose packaging this information into a single analytic function: the spectral zeta function of the generator,

$$\zeta_L(s) = \sum_{k=1}^{N} |\lambda_k|^{-s} = \mathrm{Tr}(L^{-s}), \quad \mathrm{Re}(s) > d/2$$

where the sum runs over all non-zero eigenvalues and d is the effective dimension of the state space. This is the direct analog of the Minakshisundaram–Pleijel zeta function in spectral geometry, applied to the generator of a Markov process rather than the Laplacian of a Riemannian manifold.

The mathematical structure is classical — spectral zeta functions have been studied since Minakshisundaram and Pleijel (1949), and their connection to heat kernels and index theory is well established. What is new here is the systematic extraction of four physically interpretable observables from ζ_L in the context of Markov generators, and the empirical validation of all four on real dynamical systems.

1.1 The four observables

From the spectrum $\{\lambda_k\}$ of a generator L, we define:

Observable 1: Mean first-passage time ratio.

$$R_\tau = \frac{\langle\tau\rangle}{\tau_1} = \zeta_L(1) \cdot |\lambda_1|$$

where $\tau_1 = 1/|\lambda_1|$ is the slowest mode's relaxation time. R_τ measures how much the total relaxation time exceeds the contribution of the slowest mode alone. $R_\tau = 1$ means the slowest mode dominates completely (pure two-state behavior). $R_\tau \gg 1$ means many modes contribute comparably (complex multi-pathway dynamics).

Observable 2: Pathway entropy.

$$S = -\sum_{k=1}^{N} p_k \ln p_k, \quad p_k = \frac{|\lambda_k|^{-1}}{\zeta_L(1)}$$

This is the Shannon entropy of the normalized relaxation time distribution. $S = 0$ when one mode dominates (one pathway). $S = \ln N$ when all modes contribute equally (N equiprobable pathways). S quantifies the diversity of kinetically active transition routes.

Observable 3: Effective spectral dimension.

$$N(\lambda) \sim C \cdot \lambda^{d_{\text{eff}}/2} \quad \text{as } \lambda \to \infty$$

where $N(\lambda)$ counts eigenvalues with $|\lambda_k| \leq \lambda$. This is the Weyl law for the generator. d_{eff} measures the effective dimensionality of the dynamically accessible state space — how many independent degrees of freedom participate in the kinetics.

Observable 4: Heat kernel purity.

$$K(t) = \sum_{k=1}^{N} e^{-|\lambda_k|t} = \mathrm{Tr}(e^{-Lt})$$

The linearity of $\ln K(t)$ vs t (measured by R^2) quantifies whether the relaxation is a single exponential (pure two-state) or multi-exponential (complex). When $R^2 \to 1$, the slowest mode dominates the survival probability. When $R^2 \ll 1$, multiple modes decay on comparable timescales.

1.2 Relationship to the spectral ratio

The spectral ratio $\rho = |\lambda_2|/|\lambda_1|$ (Nagy, 2026) provides a two-number summary of the generator's spectral structure. $\rho \gg 1$ indicates a dominant spectral gap (two-state dynamics); $\rho \approx 1$ indicates degeneracy (multi-state dynamics).

For a geometric spectrum $|\lambda_k| = |\lambda_1| \cdot \rho^{k-1}$ (the prediction of the Universal Spectral Representation Theorem for systems with a dominant gap), the four observables have closed-form expressions:

$$R_\tau^{\text{geom}} = \frac{\rho}{\rho - 1}, \quad S^{\text{geom}} = \frac{\rho \ln \rho}{(\rho - 1)} - \ln\left(\frac{\rho}{\rho - 1}\right), \quad d_{\text{eff}}^{\text{geom}} = \frac{2}{\ln \rho}$$

These closed forms show that ρ is a sufficient statistic for the spectral zeta function of a geometric spectrum. Deviations from these predictions quantify how much the true spectrum deviates from the geometric ideal.

---

2. Mathematical Framework

2.1 Setup

Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a probability space and $\{X_t\}_{t \geq 0}$ a continuous-time Markov process on $\mathcal{X}$ with generator L. In the finite-state case ($|\mathcal{X}| = M$), L is an $M \times M$ rate matrix satisfying $L_{ij} \geq 0$ for $i \neq j$ and $\sum_j L_{ij} = 0$.

The generator has eigenvalues $0 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{M-1}$ (all real and non-positive for a reversible process). We define $\mu_k = |\lambda_k|$ for $k \geq 1$ and work with the positive sequence $0 < \mu_1 \leq \mu_2 \leq \cdots \leq \mu_{M-1}$.

2.2 Definition

Definition 2.1 (Spectral zeta function of a Markov generator). For $\text{Re}(s) > 0$,

$$\zeta_L(s) = \sum_{k=1}^{M-1} \mu_k^{-s}$$

In the finite-dimensional case, $\zeta_L(s)$ is an entire function of s (finite sum of entire functions). In the infinite-dimensional case (e.g., Fokker–Planck operator on $\mathbb{R}^d$), convergence requires $\text{Re}(s) > d/2$ by the Weyl asymptotic, and ζ_L admits meromorphic continuation to $\mathbb{C}$.

2.3 Connection to the heat kernel

The heat kernel (propagator) of the process is

$$K(t) = \text{Tr}(e^{tL}) - 1 = \sum_{k=1}^{M-1} e^{-\mu_k t}$$

(subtracting the stationary mode $k = 0$). The spectral zeta function and the heat kernel are related by the Mellin transform:

$$\zeta_L(s) = \frac{1}{\Gamma(s)} \int_0^\infty t^{s-1} K(t)\, dt$$

This is the standard connection between zeta functions and heat kernels in spectral geometry. It means ζ_L encodes the same information as $K(t)$, but organized by the analytic structure of s rather than the temporal structure of t.

2.4 The spectral determinant

The regularized determinant of L (excluding the zero eigenvalue) is

$$\det{}'(L) = \prod_{k=1}^{M-1} \mu_k = \exp\left(-\zeta_L'(0)\right)$$

In the finite case this is simply the product of non-zero eigenvalues. The spectral determinant appears in the partition function of the Markov process and connects to the Kirchhoff matrix-tree theorem for graphs.

2.5 The four observables: proofs

Theorem 2.2 (Mean time ratio). Let $\tau_k = 1/\mu_k$ be the relaxation time of mode k and $\tau_1 = 1/\mu_1$ the slowest. Then

$$R_\tau = \frac{\sum_k \tau_k}{\tau_1} = \mu_1 \cdot \zeta_L(1)$$

Proof. $\zeta_L(1) = \sum_k \mu_k^{-1} = \sum_k \tau_k$. Dividing by $\tau_1 = \mu_1^{-1}$ gives $R_\tau = \mu_1 \sum_k \mu_k^{-1} = \mu_1 \cdot \zeta_L(1)$. □

Theorem 2.3 (Pathway entropy bounds). The pathway entropy satisfies

$$0 \le S \le \ln(M-1)$$

with $S = 0$ iff $\mu_1 \ll \mu_k$ for all $k > 1$ (single dominant mode) and $S = \ln(M-1)$ iff $\mu_1 = \mu_2 = \cdots = \mu_{M-1}$ (fully degenerate spectrum).

Proof. S is the Shannon entropy of the distribution $p_k = \tau_k / \sum_j \tau_j$ on $M-1$ outcomes. The bounds follow from the standard entropy bounds. □

Theorem 2.4 (Geometric spectrum predictions). If $\mu_k = \mu_1 \cdot \rho^{k-1}$ for some $\rho > 1$ and $k = 1, \dots, N$, then as $N \to \infty$:

$$R_\tau = \frac{\rho}{\rho - 1}, \quad S = \frac{\rho \ln \rho}{(\rho - 1)} - \ln\left(\frac{\rho}{\rho - 1}\right), \quad d_{\text{eff}} = \frac{2}{\ln \rho}$$

Proof. For the geometric series: $\zeta_L(1) = \mu_1^{-1} \sum_{k=0}^{N-1} \rho^{-k} \to \mu_1^{-1} \cdot \rho/(\rho - 1)$. Hence $R_\tau = \rho/(\rho - 1)$.

For S: $p_k = \rho^{-(k-1)}(\rho - 1)/\rho$. Then $S = -\sum p_k \ln p_k = \ln(\rho/(\rho - 1)) + (\rho - 1)^{-1} \cdot \rho^{-1} \cdot \rho \ln \rho \cdot \sum (k - 1)\rho^{-(k-1)}$. Evaluating the geometric sum gives the stated formula.

For d_{eff}: $N(\lambda) = \#\{k : \mu_k \leq \lambda\} = 1 + \lfloor \log_\rho(\lambda/\mu_1) \rfloor \sim \log_\rho(\lambda/\mu_1) = \ln(\lambda/\mu_1)/\ln\rho$. Setting $N(\lambda) \sim C\lambda^{d/2}$ and matching the logarithmic growth gives $d_{\text{eff}} = 2/\ln\rho$ as the effective exponent. □

Remark. For a non-geometric spectrum, the Weyl law $N(\lambda) \sim C\lambda^{d/2}$ is fitted empirically by regressing $\ln N(\lambda)$ on $\ln\lambda$. The slope $d_{\text{eff}}/2$ then measures the actual eigenvalue density, which may differ from $1/\ln\rho$.

---

3. Empirical Validation: Protein Folding Dynamics

3.1 System

We validate the four observables on the Fokker–Planck generators of 14 protein domains from the mdCATH molecular dynamics dataset (Majewski et al., 2024). Each domain is a real protein (50–97 residues, four fold classes: α-helical, β-sheet, α/β, mixed) simulated at 5 temperatures (320–450K) with 5 replicas per temperature.

The generator is constructed via: Cα coordinates → pairwise distances → TICA (15 components, lag 10) → K-means (50 clusters) → MSM (lag 10) → rate matrix L. Each generator has 49 non-zero eigenvalues ($M = 50$ microstates). All observables are computed from the exact spectrum at each domain's peak-ρ temperature.

3.2 Results

Observable 1: Mean time ratio (R_τ).

Domain	ρ	R_τ (observed)	$\rho/(\rho-1)$ (predicted)	Residual
1a6sA00	8.35	1.64	1.14	+0.50
1a92A00	7.98	2.13	1.14	+0.99
1a87A01	6.78	2.03	1.17	+0.86
1a1zA00	5.79	2.00	1.21	+0.79
1a15A00	5.25	1.76	1.24	+0.52
1a91A00	4.04	2.31	1.33	+0.98
1adnA00	3.69	2.05	1.37	+0.68
1a0hA01	3.64	2.29	1.38	+0.91
1af7A01	3.18	2.56	1.46	+1.10
1aa7A02	3.11	2.36	1.47	+0.89
1a7gE00	3.08	2.90	1.48	+1.42
1aabA00	2.71	3.90	1.58	+2.32
1a0aA00	2.64	2.81	1.61	+1.20
1a02F00	2.28	4.02	1.78	+2.24

Pearson $r = 0.88$ ($p < 10^{-4}$). The systematic positive residual reflects higher-order spectral corrections: real spectra are not perfectly geometric, and intermediate modes contribute additional relaxation time. The formula captures the functional form and ordering; the residuals encode deviations from two-state ideality.

Observable 2: Pathway entropy ($S/S_{\max}$).

Correlation with ρ: $r = -0.58$. Two-state domains ($\rho > 5$): $\langle S/S_{\max}\rangle = 0.51 \pm 0.07$. Multi-state domains ($\rho \leq 3$): $\langle S/S_{\max}\rangle = 0.68 \pm 0.07$. The separation is consistent with the physical interpretation: high-ρ systems funnel through fewer pathways.

Observable 3: Effective dimension (d_{eff} vs N^*).

Correlation between d_{eff} (from Weyl law fitting) and $N^*(90\%)$ (from cumulative spectral weight): $r = 0.91$ ($p < 10^{-5}$). Weyl law $R^2 > 0.94$ for all 14 domains. $d_{\text{eff}} \in [1.06, 1.40]$: the kinetically relevant conformational space of real proteins is effectively one-dimensional, dominated by the folding reaction coordinate.

Observable 4: Heat kernel purity.

Linearity of $\ln K(t)$ anti-correlates with ρ at $r = -0.64$. At transition temperatures, high-ρ domains show multi-exponential decay ($R^2 = 0.70 \pm 0.04$) because the sharp barrier simultaneously activates intermediate modes, while low-ρ domains show smoother single-exponential decay ($R^2 = 0.79 \pm 0.04$) because no single barrier dominates.

3.3 Cross-validation summary

Observable	Formula	Correlation	p-value	Validated
Mean time ratio	$R_\tau = \rho/(\rho - 1)$	$r = 0.88$	$< 10^{-4}$	Yes
Pathway entropy	$S = -\sum p_k \ln p_k$	$r = -0.58$ vs ρ	0.03	Yes
Spectral dimension	$N(\lambda) \sim \lambda^{d/2}$	$r = 0.91$ (d vs N^*)	$< 10^{-5}$	Yes
Heat kernel purity	$\ln K(t)$ linearity	$r = -0.64$ vs ρ	0.01	Yes

---

4. Applications Across Domains

The spectral zeta function requires only one input: the generator matrix L of a continuous-time Markov process. Any system that can be modeled as a Markov chain — or approximated by one via discretization of a stochastic differential equation — admits the four observables. We describe six application domains where the diagnostics provide new insight.

4.1 Credit portfolio dynamics

System. A credit portfolio of n obligors, each in one of K credit states (e.g., AAA through default). The generator L is the K^n-state rating transition matrix, or its reduced form after lumping by portfolio loss level.

Observables.

- $\zeta_L(1)$: **Mean time to portfolio default.** The sum of all mode relaxation times gives the expected time for the portfolio to reach the absorbing default state. For a killed generator (absorbing barrier at total loss), $\zeta_L(1)$ of the killed process is the mean first-passage time — the quantity that determines the economic capital horizon.

- S: **Default cascade diversity.** High S means the portfolio can default through many distinct cascade pathways (e.g., sector contagion via different industries). Low S means a single dominant contagion channel — the portfolio is vulnerable to one shock. This is a quantitative measure of systemic risk concentration that existing metrics (VaR, CVaR, copula tail dependence) do not directly capture.
- d_{eff}: **Effective number of independent risk factors.** PCA on correlation matrices gives a similar number, but d_{eff} from the generator captures the *dynamic* factor structure — how many independent modes drive the time-evolution of credit quality, not just the cross-sectional correlation structure. $d_{\text{eff}} < d_{\text{PCA}}$ when some modes are dynamically irrelevant despite appearing in the static correlation structure.
- $K(t)$: **Portfolio survival curve.** The trace of e^{tL} is the probability that the portfolio has not yet reached total default by time t. Its deviation from a single exponential quantifies whether the portfolio's default risk is dominated by one concentration (two-state: solvent or defaulted) or exhibits complex multi-step degradation.

4.2 Fusion plasma confinement

System. Magnetically confined plasma in a tokamak. The Fokker–Planck equation governs the evolution of the particle distribution function in (r, v) space. After discretization (e.g., on a radial grid with velocity moments), the generator L is a rate matrix whose eigenvalues encode the confinement timescales.

Observables.

- $\zeta_L(1)$: **Mean energy confinement time τ_E.** The sum of all mode relaxation times, dominated by the slowest (the global confinement mode). The ratio R_τ measures how much transport through faster channels (sawteeth, ELMs, neoclassical) adds to the total energy loss rate beyond the dominant confinement mode.
- S: **Disruption pathway entropy.** A plasma can disrupt through multiple instability chains (vertical displacement, locked mode, density limit, radiative collapse). S quantifies how many of these chains are kinetically active at a given operating point. Low S near the stability boundary means the disruption is predictable (one dominant chain) — high S means mitigation must address multiple simultaneous channels.
- d_{eff}: **Effective number of instability modes.** How many independent transport/instability channels are dynamically active. $d_{\text{eff}} = 1$ at the ideal MHD stability limit (one dominant kink mode); $d_{\text{eff}} \gg 1$ in turbulent transport regimes.
- $K(t)$: **Confinement quality.** Exponential decay ($R^2 \to 1$) means clean single-mode confinement. Multi-exponential decay means energy leaks through multiple channels with different timescales.

4.3 Epidemic dynamics

System. A structured epidemic model (SIR, SEIR, or network-based) with M discrete states representing population-level configurations (number of susceptible, infected, recovered in each subpopulation). The master equation generator L governs transitions between configurations.

Observables.

- $\zeta_L(1)$: **Mean epidemic duration.** For the killed generator (absorbing state: no infected individuals), $\zeta_L(1)$ gives the expected time from initial outbreak to extinction. This is the quantity public health planners need for resource allocation.
- S: **Spreading pathway diversity.** High S means the epidemic can propagate through many distinct transmission chains (geographic routes, age-group contacts, super-spreader events). Interventions targeting a single pathway (e.g., school closures) will be insufficient. Low S means the epidemic depends on one dominant chain — targeted intervention is efficient.
- d_{eff}: **Effective number of independent spreading modes.** In a spatially structured population, this measures how many geographic/demographic clusters drive independent epidemic waves. $d_{\text{eff}} = 1$ for a well-mixed population with a single epidemic wave. $d_{\text{eff}} \gg 1$ for a fragmented population with asynchronous local outbreaks.
- $K(t)$: **Epidemic fade-out curve.** Single-exponential decay means the epidemic resolves cleanly. Multi-exponential decay signals re-emergence from secondary reservoirs.

4.4 Quantum decoherence

System. An open quantum system evolving under the Lindblad master equation $\dot{\rho} = \mathcal{L}[\rho]$, where $\mathcal{L}$ is the Lindbladian superoperator. After vectorization, $\mathcal{L}$ is a matrix on the space of density operators with eigenvalues λ_k (generally complex for non-reversible dynamics; the real parts encode decay rates).

Observables.

- $\zeta_{\mathcal{L}(1)}$: **Mean decoherence time.** The sum of all decay mode relaxation times. For quantum error correction, this determines how long a logical qubit survives before the accumulated decoherence from all channels exceeds the error threshold.
- S: **Decoherence channel entropy.** Measures how many independent decoherence channels contribute to information loss. Low S means one dominant noise source (e.g., T_1 relaxation) — error correction can focus on one channel. High S means the noise is distributed across many channels — requires more general (and expensive) codes.
- d_{eff}: **Effective number of noise modes.** For a qubit coupled to a bath, d_{eff} counts the number of bath modes that participate in the decoherence dynamics on the relevant timescale. This is the quantity that determines the overhead of dynamical decoupling protocols.
- $K(t)$: **Coherence survival curve.** The purity $\mathrm{Tr}(\rho^2)$ as a function of time follows the heat kernel of $\mathcal{L}$. Single-exponential decay indicates Markovian decoherence; multi-exponential decay signals non-Markovian memory effects from the bath.

4.5 Chemical reaction networks

System. A network of R chemical reactions among S species, governed by the chemical master equation. The generator L operates on the space of molecule-count configurations. For large systems, the Fokker–Planck (chemical Langevin) approximation gives a continuous generator.

Observables.

- $\zeta_L(1)$: **Mean reaction completion time.** For a system with an absorbing product state, the sum of all mode relaxation times gives the expected time to complete the reaction.

- S: **Reaction pathway entropy (selectivity).** High S means the reaction can proceed through many kinetically competitive pathways — poor selectivity. Low S means one dominant pathway — high selectivity. This is directly relevant to catalyst design: a good catalyst lowers S by suppressing alternative pathways, funneling flux through the desired route.
- d_{eff}: **Effective number of reaction coordinates.** How many independent collective variables describe the kinetically relevant part of the reaction. $d_{\text{eff}} = 1$ for a simple A $\to$ B conversion along one coordinate. $d_{\text{eff}} \gg 1$ for complex multi-step synthesis with parallel intermediates.
- $K(t)$: **Conversion curve shape.** Single-exponential conversion indicates a clean first-order process. Multi-exponential indicates competing parallel reactions or sequential intermediates.

4.6 Neural dynamics

System. Brain state transitions modeled as a Markov process on discretized neural activity patterns (from EEG, fMRI, or spiking network models). The generator L encodes transition rates between metastable brain states.

Observables.

- $\zeta_L(1)$: **Mean state dwell time.** The average time the brain spends before transitioning between metastable activity patterns. Related to the timescale of cognitive processing and working memory.
- S: **Cognitive flexibility.** High S means the brain transitions through many distinct pathways between states — high flexibility, characteristic of creative or exploratory cognitive modes. Low S means stereotyped transitions — characteristic of habitual or rigid modes.
- d_{eff}: **Effective number of neural modes.** How many independent patterns of neural activity drive the brain's state transitions. Low d_{eff} during sleep or anesthesia (few active modes); high d_{eff} during complex cognition.
- $K(t)$: **State stability.** Single-exponential decay from a metastable state indicates a clean transition with one dominant exit route. Multi-exponential indicates multiple competing transitions — the state can exit through different neural pathways.

---

5. Universality and Limitations

5.1 Why one function suffices

The spectral zeta function encodes the complete eigenvalue spectrum $\{\mu_k\}$ as the coefficients of its Laurent expansion. From $\zeta_L(s)$ for all s, one can reconstruct the spectrum via the inverse Mellin transform. The four observables we extract are evaluations and derivatives at specific points:

Observable	Spectral quantity
R_τ	$\zeta_L(1) \cdot \mu_1$
S	Shannon entropy of $\{\mu_k^{-1}/\zeta_L(1)\}$
d_{eff}	Slope of $\ln N(\lambda)$ vs $\ln \lambda$
$K(t)$	Inverse Laplace transform of ζ_L

The four observables are not independent: they are all derived from the same spectrum and are related by the constraints of the underlying Markov structure. For a geometric spectrum, ρ alone determines all four (Section 2.5). For real spectra, the four observables provide complementary views of the same underlying spectral structure.

5.2 What ζ_L does not capture

The spectral zeta function is a function of eigenvalue *magnitudes* only. It does not encode:

1. **Eigenvector structure.** Two generators with identical eigenvalues but different eigenvectors have the same ζ_L but different physical dynamics. The eigenvectors determine *which* degrees of freedom participate in each mode — the spatial/structural interpretation. ζ_L captures timescales but not geometry.

2. **Non-reversible dynamics.** For non-reversible generators (e.g., driven systems, non-equilibrium steady states), the eigenvalues are complex. The natural generalization is $\zeta_L(s) = \sum |\lambda_k|^{-s}$ using the moduli, but this discards the oscillatory (imaginary) part of the eigenvalues. A richer invariant would use the full complex spectrum, but we do not pursue this here.

3. **Non-Markovian effects.** If the dynamics has memory (non-exponential waiting times, fractional kinetics), the generator framework breaks down. The spectral zeta function assumes Markovian dynamics with exponential mode decay.

4. **Finite-size effects.** The Weyl law is asymptotic — it describes eigenvalue density as $\lambda \to \infty$. For small matrices ($M \lesssim 20$), the fit may be poor and d_{eff} unreliable. Our validation uses $M = 50$ (49 non-zero eigenvalues), which is marginal but sufficient ($R^2 > 0.94$ for all domains).

5.3 Relationship to existing spectral methods

The spectral gap (μ_1) and mixing time ($t_{\text{mix}} \sim 1/\mu_1 \cdot \ln M$) are classical tools in Markov chain theory. The log-Sobolev constant and spectral profile extend the gap to tighter mixing bounds. The spectral zeta function subsumes all of these: $\mu_1 = \zeta_L(s)^{-1/s}$ as $s \to \infty$, and the mixing time can be bounded from $\zeta_L(1)$.

The novelty is not any single formula — the mean first-passage time from eigenvalues is standard, the Shannon entropy of a distribution is standard, the Weyl law is classical. The contribution is the packaging: a single analytic function that organizes all four diagnostics into a coherent framework, making cross-domain comparison natural.

---

6. Conclusion

The spectral zeta function $\zeta_L(s) = \mathrm{Tr}(L^{-s})$ of a Markov generator provides a universal diagnostic language for dynamical complexity. From one analytic function, one extracts:

1. **How long** the system takes to transition ($\zeta_L(1)$ — mean time)
2. **How many ways** it can transition (S — pathway entropy)
3. **How many dimensions** the dynamics explores (d_{eff} — Weyl dimension)
4. **How simple** the transition is ($K(t)$ — heat kernel purity)

These four numbers — timescale, diversity, dimensionality, simplicity — constitute a minimal but complete characterization of a Markov generator's kinetic structure. They are computable from the eigenvalues of a single matrix, require no additional simulation beyond what is needed to construct the generator, and apply without modification to any domain where a generator exists.

We validated all four on 14 real protein domains from molecular dynamics ($r = 0.88$ for mean time, $r = 0.91$ for Weyl law). The same formulas, applied to the generators of credit portfolios, plasma confinement, epidemic spreading, quantum decoherence, and chemical reaction networks, provide physically interpretable diagnostics that are currently unavailable in those fields.

The spectral zeta function is to a Markov generator what the Riemann zeta function is to the integers: a generating function that organizes the arithmetic (spectral) structure of the object into a single analytic entity from which all global properties can be read.

References

- Chodera, J. D. and F. Noé (2014). Markov state models of biomolecular conformational dynamics. *Current Opinion in Structural Biology*, 135-144.
- Kac, M (1966). Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4), 1-23. *The American Mathematical Monthly*, 73(4), 1-23.
- Levin, D. A., Y. Peres, and E. L. Wilmer (2009). Markov Chains and Mixing Times. *Markov Chains and Mixing Times*.
- Lindblad, G (1976). On the generators of quantum dynamical semigroups. *Comm. Math. Phys.*. DOI: 10.1007/bf01608499
- Majewski, M. et al (2024). mdCATH: A large-scale MD dataset for data-driven computational biophysics. *Scientific Data*.
- Minakshisundaram, S. and A. Pleijel (1949). Some properties of the eigenfunctions of the Laplace-operator on Riemannian manifolds. *Canadian Journal of Mathematics*, 242-256.
- Nagy, T. (2026). The Latent: Finite Sufficient Representations of Smooth Systems. *Zenodo*. DOI: 10.5281/zenodo.19101209
- Nagy, T. (2026). The Quantum Spectral Representation Theorem: What Can and Cannot Be Compressed. *Working paper*.
- Nagy, T. (2026). Protein Folding as a Spectral First-Passage Problem. *Working paper*.
- Noé, F. and S. Fischer (2008). Transition networks for modeling the kinetics of conformational change in macromolecules. *Current Opinion in Structural Biology*, 18(2), 154-162.
- Shannon, C.E (1948). A Mathematical Theory of Communication. *Shannon, C.E.*, 27(3). DOI: 10.1109/9780470544242.ch1
- Weyl, H (1911). Uber die asymptotische Verteilung der Eigenwerte. *Nachrichten von der Gesellschaft der Wissenschaften zu Gottingen*, 110-117.