Steering Dynamical Regimes of Diffusion Models by Breaking Detailed Balance

Anonymous Authors[1]

Abstract

Diffusion models pass through time-localized dynamical regimes in which samples first commit to semantic modes and may later collapse toward individual training examples. We study how these regimes change when the forward Ornstein–Uhlenbeck noising process is made non-reversible while preserving its invariant Gaussian measure. The drift is written as $\mathbf{A} = (\mathbf{I} + \mathbf{Q})\mathbf{U}$, where $\mathbf{U}$ fixes the stationary geometry and the anti-symmetric matrix $\mathbf{Q}$ creates probability currents. We derive a matrix criterion for the speciation time and a Random-Energy-Model criterion for collapse that depends on a Mahalanobis signal-to-noise spectrum. Experiments on Gaussian mixtures and trained DDPMs show that geometry-aware non-reversibility can move speciation earlier, whereas collapse timing is not generically moved earlier by breaking detailed balance.

1. Introduction

Generative diffusion models are often described as reverse-time stochastic processes driven by a learned score (Song et al., 2021). Recent statistical-physics analyses show that this reverse trajectory is not a featureless denoising path: it contains dynamical regimes associated with semantic mode selection, generalization, and memorization (Biroli et al., 2024; Biroli & Mézard, 2023; Ambrogioni, 2025). In particular, the *speciation* transition marks the time at which trajectories commit to a data mode, while the *collapse* transition marks the onset of a memorization-dominated regime in which individual training samples become the relevant attractors (Achilli et al., 2025; Pham et al., 2025; Ye et al., 2026). Understanding how these transition times depend on the forward generator is directly related to the workshop themes of memorization, generalization and dynamics.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the FoGen Workshop at ICML 2026. Do not distribute.

The usual linear forward process is reversible and isotropic. This is convenient, but it fixes a particular dynamical route toward the noise distribution. Linear nonequilibrium theory suggests a different design space: one can keep the invariant distribution fixed while adding rotational probability currents that break detailed balance (Ao, 2004; Kwon et al., 2005; Qian, 2013; Godrèche & Luck, 2019). For Gaussian diffusions such currents can improve asymptotic relaxation (Hwang et al., 1993; Ichiki & Ohzeki, 2013; Lelièvre et al., 2013), but it is not obvious whether this relaxation-rate improvement translates into useful changes in diffusion-model regime timing.

This paper asks how non-reversible forward dynamics reshape speciation and collapse. Our contributions are: (i) a non-reversible OU formulation that preserves the stationary noising distribution while changing transient dynamics; (ii) a general speciation criterion expressed as a matrix eigenvalue crossing; (iii) a collapse analysis separating the definition-level entropic criterion from an implicit REM criterion for general linear drifts; and (iv) Gaussian-mixture and real-data DDPM experiments showing that non-reversibility is strongly geometry-dependent. Suitable Lelièvre-style currents can move speciation earlier, but weak generic anti-symmetric perturbations leave real-data speciation and collapse largely unchanged.

This distinction is important for the language of "acceleration." In the stochastic-process literature, breaking detailed balance can accelerate convergence to equilibrium in a spectral-gap sense. The diffusion-model transitions studied here are different objects: they are finite-time dynamical regimes along the reverse generative trajectory. We therefore ask whether these regimes are *shifted in time*, rather than assuming that all desirable events are accelerated. The answer is asymmetric. Speciation can move earlier when the non-reversible current couples to the class-separation signal, while collapse is governed by an entropic or REM competition and does not generically move earlier.

2. Related Work and Positioning

Dynamical regimes and memorization. Biroli and collaborators introduced a statistical-physics view of diffusion

generation in high dimension, identifying speciation and collapse as distinct dynamical events (Biroli et al., 2024; Biroli & Mézard, 2023). Related work connects diffusion models to associative memory (Ambrogioni, 2024; Pham et al., 2025), manifold-driven memorization (Achilli et al., 2025; 2024), entropy signatures of class commitment (Handke et al., 2026), and training-time separations between memorization and generalization (Bonnaire et al., 2025; Ye et al., 2026). Our contribution is complementary: we do not study when a neural score network memorizes during optimization, but how a fixed generator's reverse-time regimes are displaced when the forward noising dynamics is changed.

Inductive bias through the forward generator. Several works show that diffusion models encode data geometry and intrinsic dimension through the noising and denoising process (Stanczuk et al., 2024; Kadkhodaie et al., 2024). Others propose alternative samplers, time parameterizations, or architectures to reduce sampling cost. Our setting is narrower but analytically controlled: for a fixed symmetric geometry $\mathbf{U}$, changing $\mathbf{Q}$ keeps the stationary noising distribution fixed and changes only the probability current. This isolates a clean form of dynamical inductive bias: the score-learning objective and model class are kept fixed in form, while the transient forward marginals and reverse linear drift are modified by $\mathbf{Q}$.

Non-reversible dynamics. Nonequilibrium decompositions of diffusion processes separate potential forces, probability currents, and noise (Ao, 2004; Kwon et al., 2005; Qian, 2013). For linear OU processes this decomposition is explicit, and the stationary covariance can be preserved while adding anti-symmetric currents (Kwon et al., 2011; Godrèche & Luck, 2019). Non-reversible perturbations are known to improve asymptotic relaxation rates for Gaussian diffusions (Hwang et al., 1993; Ichiki & Ohzeki, 2013; Lelièvre et al., 2013). We use this machinery as a controllable forward-process design, but we evaluate it through speciation and collapse rather than through equilibrium convergence alone.

Coupled OU dynamics and synchronization gaps. A closely related work is Albrychiewicz et al. (Albrychiewicz et al., 2026a), which studies multimodal diffusion through coupled OU dynamics and identifies a spectral hierarchy that creates a synchronization gap between relaxation channels. The intervention is different from ours. Their control parameter is an inter-modality coupling structure and strength, whereas we add an anti-symmetric component to the generator while preserving the invariant Gaussian measure.

For speciation, Albrychiewicz et al. characterize mode selection through a bifurcation of fixed points in the reverse-time deterministic drift. Starting from the fixed-point condition for the reverse drift, they reduce the problem to a scalar self-consistency relation, and the critical time is the corresponding pitchfork threshold. Here speciation is formulated instead as a curvature instability of the evolving log-density. This gives the matrix eigenvalue condition in (7), which also covers non-normal non-reversible drifts.

For collapse, Albrychiewicz et al. derive a semi-analytical condition for the collapse time in the coupled setting and observe numerically that the overall collapse onset is robust to the coupling. For $\mathbf{A} = (\mathbf{I} + \mathbf{Q})\mathbf{U}$, we keep the same entropic-volume definition but make the REM approximation explicit for a general stable linear drift. The resulting implicit criterion depends on the eigenvalues of $\mathbf{K}_t = e^{-\mathbf{A}^\top t}\boldsymbol{\Sigma}_{\text{sto}}^{-1}(t)e^{-\mathbf{A}t}$, not only on $\mathrm{Tr}(\mathbf{A})$. Thus $\mathrm{Tr}(\mathbf{QU}) = 0$ explains a volume-contraction robustness mechanism, but anti-symmetric perturbations can still shift the REM collapse time by reshaping this Mahalanobis signal-to-noise spectrum.

3. Non-Reversible Forward Dynamics

Let the data distribution be supported on samples $\mathbf{a}_\mu \in \mathbb{R}^d$. We replace the standard isotropic OU noising process by

$$d\mathbf{x}_t = -\mathbf{A}\mathbf{x}_t\, dt + \sqrt{2}\, d\mathbf{W}_t, \qquad \mathbf{A} = (\mathbf{I} + \mathbf{Q})\mathbf{U}, \quad (1)$$

where $\mathbf{U} = \mathbf{U}^\top > 0$ and $\mathbf{Q} = -\mathbf{Q}^\top$. The symmetric matrix $\mathbf{U}$ defines the invariant covariance, while $\mathbf{Q}$ changes transient currents. Indeed, the stationary covariance solves

$$\mathbf{A}\boldsymbol{\Sigma}_s + \boldsymbol{\Sigma}_s\mathbf{A}^\top = 2\mathbf{I}, \qquad \boldsymbol{\Sigma}_s = \mathbf{U}^{-1}. \quad (2)$$

Thus changing $\mathbf{Q}$ does not change the stationary Gaussian noising target.

The forward solution from an initial point $\mathbf{a}$ is

$$\mathbf{x}_t = e^{-\mathbf{A}t}\mathbf{a} + \sqrt{2}\int_0^t e^{-\mathbf{A}(t-s)} d\mathbf{W}_s, \quad (3)$$

with stochastic covariance

$$\boldsymbol{\Sigma}_{\text{sto}}(t) = 2\int_0^t e^{-\mathbf{A}s}e^{-\mathbf{A}^\top s}\, ds = \boldsymbol{\Sigma}_s - e^{-\mathbf{A}t}\boldsymbol{\Sigma}_s e^{-\mathbf{A}^\top t}. \quad (4)$$

Under time reversal, the learned reverse SDE has linear term $\mathbf{A}\mathbf{y}$ plus the usual score correction. Hence $\mathbf{Q}$ changes the reverse-time drift field while preserving the invariant Gaussian reference; the score correction is still the score of the forward marginal generated by the chosen drift.

More explicitly, if $\tau = t_f - t$ denotes reverse time, the denoising process has the form

$$d\mathbf{y}_\tau = \left[\mathbf{A}\mathbf{y}_\tau + 2\nabla_{\mathbf{y}} \log p_{t_f-\tau}(\mathbf{y}_\tau)\right] d\tau + \sqrt{2}\, d\mathbf{W}_\tau. \quad (5)$$

The extra non-reversible component is therefore the linear term $\mathbf{QU}\mathbf{y}_\tau$. In learned DDPMs we use the same network

architecture and training protocol across drifts; the forward linear operator changes, and the reverse dynamics uses the corresponding linear term and learned score. This separation is useful because it avoids conflating changes in dynamical regime timing with changes in model capacity.

We use two kinds of anti-symmetric perturbations. The first is a simple dense perturbation with $Q_{ij} = q$ for $i < j$ and $Q_{ji} = -q$. The second is the exponentially optimal construction of Lelièvre et al. (Lelièvre et al., 2013), which can equalize the real parts of the spectrum:

$$\max_{\mathbf{Q}^\top = -\mathbf{Q}} \min \Re\sigma((\mathbf{I}+\mathbf{Q})\mathbf{U}) = \frac{\mathrm{Tr}(\mathbf{U})}{d}. \tag{6}$$

This is an asymptotic relaxation statement, not by itself a theorem about speciation or collapse. The transition criteria below identify the additional objects that actually control those regimes.

The Lelièvre construction also clarifies what the simple perturbation cannot guarantee. A dense Simple $\mathbf{Q}$ may rotate coordinates, but it is not adapted to the data signal or to the slow eigenspaces of $\mathbf{U}$. By contrast, the Lelièvre perturbation is designed to redistribute decay rates across the spectrum. Even then, finite-time behavior can differ from the asymptotic spectral rate because non-normal drift matrices can introduce prefactors and transient amplification. For this reason, the spectral-gap optimum is best viewed as a controlled benchmark rather than as a universal prescription for all transition times.

4. Dynamical-Regime Criteria

4.1. Speciation

Speciation is the time at which the evolving distribution becomes locally unstable to symmetry breaking between data modes. A Landau expansion of $-\log P_t(\mathbf{x})$ gives the general criterion

$$\lambda_{\min}\Big(\widetilde{\mathbf{M}}(t_S)\Big) = 0, \qquad \widetilde{\mathbf{M}}(t) = \mathbf{\Sigma}_{\mathrm{sto}}(t) - e^{-\mathbf{A}t}\mathbf{\Sigma}_{\mathrm{B}}e^{-\mathbf{A}^\top t}. \tag{7}$$

Here $\mathbf{\Sigma}_{\mathrm{B}}$ is the symmetry-breaking part of the data covariance. For a symmetric two-component Gaussian mixture with means $\pm\mathbf{m}$ and within-component covariance $\sigma^2\mathbf{I}$, one has $\mathbf{\Sigma}_{\mathrm{B}} = \mathbf{m}\mathbf{m}^\top$. Equation (7) compares accumulated noise against the remaining propagated class-separation signal. When $\mathbf{A}$ and $\mathbf{\Sigma}_{\mathrm{B}}$ are simultaneously diagonalizable, it reduces to the scalar expression

$$t_S = \frac{\log(1+\Lambda d_\Lambda)}{2d_\Lambda}, \tag{8}$$

where Λ is the leading symmetry-breaking eigenvalue and d_Λ is the corresponding drift eigenvalue. The full matrix form is needed for non-normal non-reversible drifts.

The criterion also gives a simple interpretation of why non-reversibility can be effective for speciation. The stochastic covariance $\mathbf{\Sigma}_{\mathrm{sto}}(t)$ grows as the noising process accumulates uncertainty, while the signal covariance $e^{-\mathbf{A}t}\mathbf{\Sigma}_{\mathrm{B}}e^{-\mathbf{A}^\top t}$ decays as the class-separation directions are contracted. Changing $\mathbf{Q}$ changes both terms through $e^{-\mathbf{A}t}$ and the Lyapunov covariance in (4). Therefore the relevant question is not whether the invariant density changes, but whether the transient flow rotates or redistributes the symmetry-breaking signal relative to the noise geometry before the instability occurs.

In the isotropic reversible limit $\mathbf{A} = \mathbf{I}$, (8) reduces to $t_S = \frac{1}{2}\log(1+\Lambda)$, and for large signal strength this matches the familiar $\frac{1}{2}\log\Lambda$ scaling. In the anisotropic reversible case, only the decay rate along the leading signal direction matters. In the non-reversible case, the signal direction need not remain an eigenvector, so the full matrix crossing in (7) is the natural object.

The speciation figures use the cloning probability $\phi(t)$: two reverse trajectories are continued from the same noisy state at time t, and $\phi(t)$ is the probability that their terminal classes agree. Values near $1/2$ indicate no class commitment, while values near 1 indicate that the shared noisy state has already selected a class.

4.2. Collapse

Collapse is a different transition. The most general criterion is entropic: the effective volume of the true forward distribution becomes comparable to the volume of n separated Gaussian lumps,

$$s(t_C) = s^{\mathrm{sep}}(t_C) = \alpha + s_G(t_C), \qquad \alpha = \frac{\log n}{d}. \tag{9}$$

This is a definition-level criterion; it becomes predictive only after one approximates the true entropy density $s(t)$.

For analytical insight we use a Random Energy Model approximation. Define

$$\mathbf{M}_t = e^{-\mathbf{A}t}, \qquad \mathbf{\Gamma}_t = \mathbf{\Sigma}_{\mathrm{sto}}(t), \qquad \mathbf{K}_t = \mathbf{M}_t^\top \mathbf{\Gamma}_t^{-1}\mathbf{M}_t. \tag{10}$$

The eigenvalues $\kappa_j(t)$ of $\mathbf{K}_t$ measure the remaining separation between training means in the Mahalanobis geometry of the single-lump noise covariance. Under the REM approximation $\mathbf{a}_1 - \mathbf{a}_\mu \sim \mathcal{N}(0, 2\sigma_0^2\mathbf{I})$, collapse is determined implicitly by

$$2\alpha + 1 = \frac{1}{d}\sum_{j=1}^{d}\left[\log(1+2\sigma_0^2\kappa_j(t_C)) + \frac{1}{1+2\sigma_0^2\kappa_j(t_C)}\right]. \tag{11}$$

This is not a closed-form solution and not an unrestricted theorem for arbitrary data. Its role is to show how $\mathbf{Q}$ can

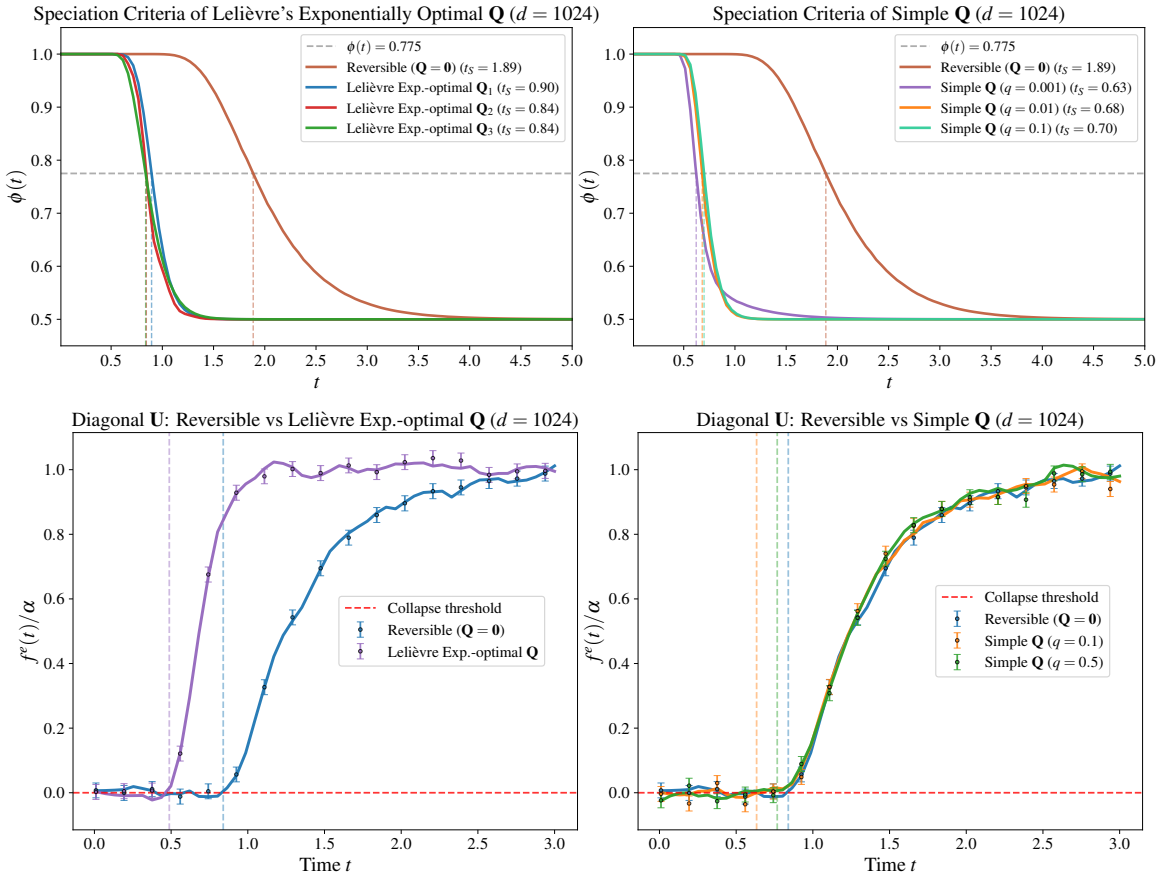

Figure 1. Gaussian mixtures show that non-reversibility can advance speciation without generically shifting collapse earlier. Top: the Lelièvre Exp.-optimal perturbations are constructed by the spectral-gap equalization principle in (6) (with the plotted $\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3$ corresponding to different auxiliary spectra), while Simple $\mathbf{Q}$ uses the dense anti-symmetric family $Q_{ij} = q$ for $i < j$ and $Q_{ji} = -q$; both can shift the cloning transition when they couple to the mixture signal. Bottom: the same reversible reference is compared with Lelièvre Exp.-optimal $\mathbf{Q}$ and selected Simple $\mathbf{Q}$ values, showing that the finite-dimensional entropy proxy can move when $\mathbf{Q}$ reshapes the Mahalanobis spectrum in (10), but the response is weaker and more diagnostic-dependent than for speciation.

enter collapse: the trace identity $\mathrm{Tr}(\mathbf{A}) = \mathrm{Tr}(\mathbf{U})$ fixes Euclidean volume contraction, but the REM criterion depends on the full spectrum of $\mathbf{K}_t$.

The diagonal reversible formula used in earlier analyses is recovered only after imposing $\mathbf{Q} = 0$ and $\mathbf{A} = \mathbf{U} = \mathrm{diag}(u_1, \ldots, u_d)$. In that case

$$\kappa_j(t) = \frac{u_j e^{-2u_j t}}{1 - e^{-2u_j t}}, \tag{12}$$

and (11) becomes the standard diagonal implicit REM equation. This restriction matters. The general entropic criterion (9) is a definition of collapse, while (11) is an approximation that becomes analytical through $\mathbf{K}_t$. Thus collapse can be $\mathbf{Q}$-dependent, but only through changes in the Mahalanobis signal-to-noise spectrum; it is not generically controlled by the same object as speciation.

5. Experiments

5.1. Gaussian Mixtures

We first use a two-component Gaussian mixture in $d = 1024$ with $\mathbf{U} = \mathrm{diag}(u_1, \ldots, u_d)$ and $u_j = 1 + 8(j-1)/(d-1)$. Speciation is measured by the trajectory-cloning probability $\phi(t)$. At forward time t, two independent reverse continuations are initialized from the same noisy state $\mathbf{y}_t$; if $c(\mathbf{x}_0)$ denotes the terminal component or class, then

$$\phi(t) = \mathbb{P}\Big[c(\mathbf{x}_0^{(1)}) = c(\mathbf{x}_0^{(2)}) \,\Big|\, \mathbf{x}_t^{(1)} = \mathbf{x}_t^{(2)} = \mathbf{y}_t\Big]. \tag{13}$$

Thus $\phi(t) \approx 1/2$ indicates no class commitment, while $\phi(t) \approx 1$ indicates that the shared noisy state already fixes the terminal component.

The Gaussian mixture is intentionally favorable to non-reversible control: the class-separation vector is spread across eigendirections of $\mathbf{U}$, so rotations can redistribute signal weight across slow and fast directions. This setting isolates the mechanism predicted by (7) before moving to

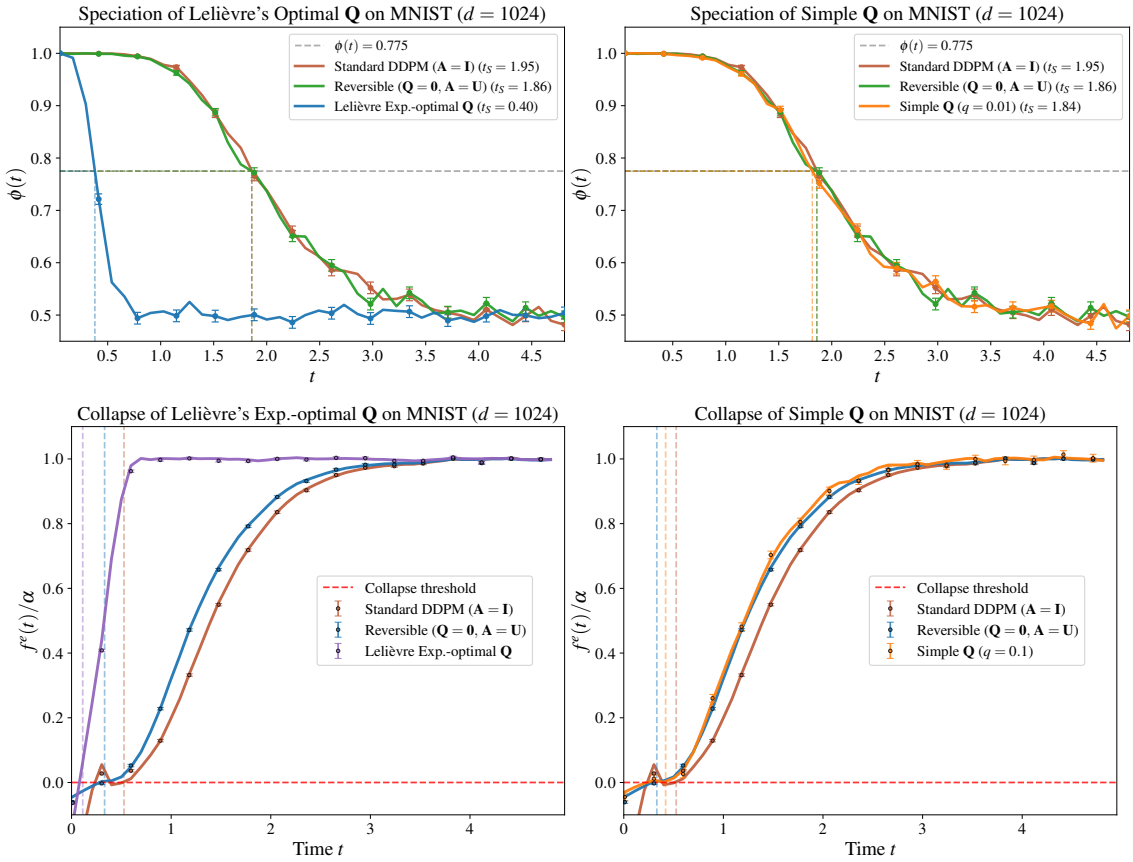

Figure 2. MNIST shows that geometry-aware non-reversibility mainly affects speciation. Top: the Lelièvre drift advances the predicted speciation transition from $t_S^{\mathrm{Rev}} = 1.86$ to $t_S^{\mathrm{NR}} = 0.40$, while Simple $\mathbf{Q}$ remains close to the reversible reference with $t_S^{\mathrm{SQ}} = 1.84$. Bottom: Simple $\mathbf{Q}$ produces only a modest, metric-dependent collapse displacement, while the Lelièvre drift moves the finite-dimensional markers more strongly.

real datasets, where the signal geometry is learned from data and can be much more concentrated.

The cloning probability is used because it measures a dynamical commitment rather than only a marginal density feature. This makes $\phi(t)$ a direct operational proxy for the regime-I to regime-II transition described by the curvature criterion.

Figure 1 shows that non-reversible drift can substantially advance speciation in physical time. This effect is not simply the asymptotic spectral gap in (6); it occurs when $\mathbf{Q}$ changes the propagated signal term in (7).

The same figure shows the corresponding collapse diagnostic. Weak simple perturbations induce modest shifts, while the Lelièvre construction can move the zero-crossing more substantially. This agrees with (11): collapse depends on $\mathbf{K}_t$, not merely on trace-level volume contraction. The normalized transition curves in Figure 4 further check that the matrix criterion predicts the actual transition scale.

The normalized curves are used only as a consistency check for t_S. They show that once each raw trajectory is measured in its own predicted instability time, the rapid change of $\phi(t)$ is aligned across the different $\mathbf{Q}$ families. The raw curves in Figure 1, however, remain the relevant comparison for regime timing, because the question is whether the physical reverse-time trajectory commits to a class earlier or later under a fixed noising schedule.

The contrast between the two rows of Figure 1 is the central theoretical message. The same anti-symmetric current that moves the speciation instability can have a smaller or diagnostic-dependent effect on collapse. This is why the paper separates the two criteria rather than summarizing both as relaxation acceleration.

5.2. Trained DDPMs on Image Data

We next train full DDPMs on binary image tasks using four forward drifts: standard DDPM ($\mathbf{A} = \mathbf{I}$), reversible anisotropic control ($\mathbf{Q} = 0, \mathbf{A} = \mathbf{U}$), Lelièvre Exp.-optimal $\mathbf{Q}$, and Simple $\mathbf{Q}$. The anisotropic operators are specified in the full PCA basis and rotated back to pixel space, so no dimensionality reduction is used in the model. All networks use the same U-Net architecture and are trained

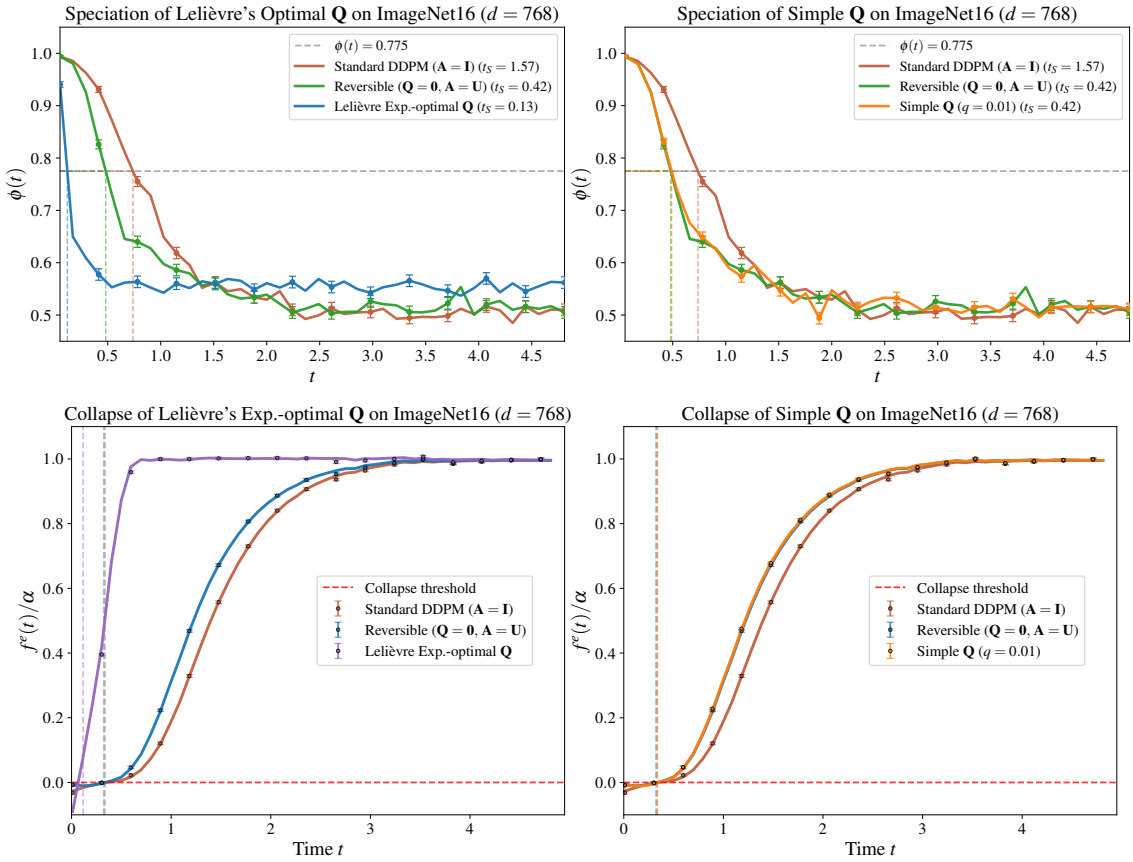

Figure 3. ImageNet16 confirms that real-data regime shifts require geometry-aware non-reversibility. Top: the Lelièvre drift advances the predicted speciation transition from $t_S^{\mathrm{Rev}} = 0.424$ to $t_S^{\mathrm{NR}} = 0.131$, while Simple **Q** with $q = 0.01$ remains essentially at $t_S^{\mathrm{SQ}} = 0.424$. Bottom: weak Simple **Q** leaves the collapse diagnostic essentially unchanged, while the Lelièvre drift moves the finite-dimensional marker much more strongly.

for 350,000 steps.

The real-data experiments are designed to test whether the synthetic Gaussian-mixture effect survives when the class-separation vector is determined by the dataset. We use MNIST 0 vs 7 and an ImageNet16 binary task. For speciation, we use the same trajectory-cloning protocol as above, with a ResNet-18 classifier to identify the terminal class. For collapse, we measure the empirical excess entropy density

$$f^e(t) = s^{\mathrm{sep}}(t) - s^e(t), \qquad s^{\mathrm{sep}}(t) = \frac{\log n}{d} + s_G(t), \tag{14}$$

where $s^e(t)$ is the entropy density of the empirical forward distribution and $s_G(t)$ is the entropy density of the corresponding single-Gaussian lump. We also use a nearest-neighbour proxy $\hat{t}_C$, defined by the last reverse-time point at which the nearest training index changes. These are operational diagnostics of the collapse criterion, not new definitions of memorization.

The four drifts play different roles in the comparison. The standard DDPM baseline fixes the familiar isotropic schedule. The reversible anisotropic control tests the effect of changing **U** without probability currents. The Simple **Q** perturbation asks whether generic detailed-balance breaking is enough. The Lelièvre perturbation is the geometry-aware benchmark, because it is constructed from the drift spectrum and therefore directly targets the slow directions of the anisotropic OU process.

Figure 2 shows that geometry-aware non-reversibility still works in a trained DDPM, but that the two dynamical regimes respond differently. For speciation, Lelièvre **Q** moves the cloning transition strongly while Simple **Q** remains close to the reversible reference. For collapse, Simple **Q** produces only a modest, metric-dependent displacement. In MNIST, about 98.6% of the class-separation energy lies in the leading principal component, which is also the slowest direction of the drift. A weak generic rotation has little signal mass to redistribute, so the cloning curve remains close to the reversible anisotropic model.

This overlap can be read directly from the signal term in (7). If $\mathbf{m}$ denotes the empirical class-separation vector and $\mathbf{v}_k$ are drift modes with coefficients α_k, then, up to non-normal

prefactors,

$$\|e^{-\mathbf{A}t}\mathbf{m}\|^2 \approx \sum_{k=1}^{d} |\alpha_k|^2 e^{-2\mathrm{Re}(\lambda_k)t}. \tag{15}$$

Simple $\mathbf{Q}$ can shift speciation only when it moves appreciable signal weight between directions with different decay rates. This is visible in Gaussian mixtures, where the signal is deliberately spread across eigendirections, but it is much weaker on real data when the class signal is already concentrated in a low-dimensional slow sector.

On ImageNet16, Figure 3 shows the same hierarchy as the theory predicts. The Lelièvre perturbation advances speciation, whereas Simple $\mathbf{Q}$ nearly overlaps with the reversible reference. The reason is again geometric: the class-separation signal is concentrated in a low-dimensional slow sector, with about 97.1% of its energy in the first 20 principal components. A weak generic anti-symmetric perturbation does not rotate enough signal weight out of this sector, while the Lelièvre construction directly reshapes the drift spectrum.

The ImageNet16 result is useful because it rules out a MNIST-specific explanation. Although the visual statistics and dimension differ, Simple $\mathbf{Q}$ again overlaps with the reversible reference. The common factor is not the dataset identity but the alignment between the perturbation and the data signal. This is the real-data counterpart of the signal-decay term in (7), summarized by (15).

Figure 3 also shows that the collapse-side result is more conservative. On ImageNet16, the entropy zero-crossings are approximately 0.323, 0.337, and 0.116 for the reversible, Simple $\mathbf{Q}$, and Lelièvre curves; the nearest-neighbour estimates are $\hat{t}_C = 0.214 \pm 0.141$, 0.214 ± 0.186, and 0.011 ± 0.013. Together with Figure 2, this supports the REM interpretation: collapse diagnostics can move when the Mahalanobis spectrum changes substantially, but breaking detailed balance alone does not imply an earlier collapse transition.

6. Discussion

We studied linear forward diffusions with drift $\mathbf{A} = (\mathbf{I} + \mathbf{Q})\mathbf{U}$, where $\mathbf{U}$ fixes the invariant Gaussian measure and the anti-symmetric part $\mathbf{Q}$ creates non-reversible probability currents. This separation lets us change transient relaxation without changing the stationary target. For the speciation transition, the effect is visible in the instability condition $\lambda_{\min}(\widetilde{\mathbf{M}}(t_S)) = 0$: changing $\mathbf{Q}$ changes both the propagated signal and the accumulated noise covariance, and can move the instability to earlier physical times. The Lelièvre construction gives the strongest shifts in our tests because it equalizes the real parts of the drift spectrum rather than only perturbing the slow modes weakly.

The collapse transition behaves differently. Its most general formulation is the entropic criterion $s(t_C) = \alpha + s_G(t_C)$, while analytical progress requires an approximation for the true entropy density $s(t)$. Within the REM approximation for a linear OU drift, the collapse time is controlled by the spectrum of $\mathbf{K}_t = e^{-\mathbf{A}^\top t}\mathbf{\Sigma}_{\mathrm{sto}}^{-1}(t)e^{-\mathbf{A}t}$. The trace identity $\mathrm{Tr}(\mathbf{A}) = \mathrm{Tr}(\mathbf{U})$ still shows that the Euclidean volume contraction of the cloud of means is fixed by the symmetric part, but this volume contraction alone does not determine the REM collapse criterion. Consequently, anti-symmetric perturbations can shift operational diagnostics such as zero-crossings of $f^e(t)$ or nearest-neighbour based estimates of $\hat{t}_C$, with the size of the shift depending on how strongly $\mathbf{Q}$ changes the Mahalanobis signal-to-noise spectrum.

This also separates our mechanism from the synchronization-gap picture in (Albrychiewicz et al., 2026a;b). Their gap comes from coupling-induced separation of relaxation channels, and recent work shows how such a gap can be realized mechanistically inside pre-trained Diffusion Transformers through attention-mediated interactions. Here the invariant measure is held fixed and the current is changed instead. Both mechanisms affect the timing of mode selection and memorization, but in our linear setting the collapse-side effect is mediated by the REM spectrum $\{\kappa_j(t)\}$ rather than by a scalar trace alone.

The recent literature also suggests several complementary directions. Entropy-based speciation diagnostics (Handke et al., 2026) could provide an alternative empirical observable for evaluating how non-reversible currents shift semantic commitment windows in trained models. Out-of-equilibrium pattern-formation analyses (Ambrogioni, 2026) and synchronization-gap studies in Diffusion Transformers (Albrychiewicz et al., 2026b) point toward architectural and spatially structured extensions of the present linear OU theory. Finally, training-time analyses of memorization and generalization (Bonnaire et al., 2025; Ye et al., 2026) address a different but related question: why learned score networks avoid or enter memorization during optimization, whereas our collapse criterion concerns the reverse-time dynamical regime induced by a fixed generator.

This interpretation also fixes the scope of the word "acceleration." The Lelièvre construction accelerates asymptotic relaxation of the forward OU process in a precise spectral sense, but our empirical claims are about displacement of finite-time reverse regimes. Earlier speciation can be useful because semantic commitment happens at a different noise level; earlier collapse is not necessarily desirable, since it corresponds to a memorization-dominated regime. Thus the relevant object is not a single speed-up factor, but the selective redistribution of speciation and collapse times.

The main limitation is that the analysis is still tied to linear forward processes and high-dimensional approximations.

Extending the same construction to nonlinear drifts or to learned, state-dependent generators would require controlling both the invariant measure and the induced score. A second open point is discretization: the continuous-time optimal $\mathbf{Q}$ need not give the best finite-step sampler once numerical stability and fixed compute are imposed. Finally, finite-time non-normal effects enter through prefactors as well as exponents, so optimizing the spectral gap alone may miss the relevant time scale for practical sampling.

References

Achilli, B., Ventura, E., Silvestri, G., Pham, B., Raya, G., Krotov, D., Lucibello, C., and Ambrogioni, L. Losing dimensions: Geometric memorization in generative diffusion. *arXiv*, 2024. URL https://arxiv.org/abs/2410.08727.

Achilli, B., Ambrogioni, L., Lucibello, C., Mézard, M., and Ventura, E. Memorization and generalization in generative diffusion under the manifold hypothesis. *J. Stat. Mech.: Theory Exp.*, 2025(7): 073401, 2025. doi: 10.1088/1742-5468/add4a6. URL https://iopscience.iop.org/article/10.1088/1742-5468/ade136/meta.

Albrychiewicz, E., Franco Valiente, A., and Chen, L.-C. Dynamical regimes of multimodal diffusion models. *arXiv*, 2026a. URL https://arxiv.org/abs/2602.04780.

Albrychiewicz, E., Franco Valiente, A., Chen, L.-C., and Zhao, V. Z. Interpreting the synchronization gap: The hidden mechanism inside diffusion transformers. *arXiv preprint arXiv:2603.20987*, 2026b. doi: 10.48550/arXiv.2603.20987. URL https://arxiv.org/abs/2603.20987.

Ambrogioni, L. In search of dispersed memories: Generative diffusion models are associative memory networks. *Entropy*, 26(5):381, 2024. doi: 10.3390/e26050381. URL https://doi.org/10.3390/e26050381.

Ambrogioni, L. The statistical thermodynamics of generative diffusion models: Phase transitions, symmetry breaking and critical instability. *Entropy*, 27(3):291, 2025. doi: 10.3390/e27030291. URL https://doi.org/10.3390/e27030291.

Ambrogioni, L. How out-of-equilibrium phase transitions can seed pattern formation in trained diffusion models. *arXiv preprint arXiv:2603.20092*, 2026. doi: 10.48550/arXiv.2603.20092. URL https://arxiv.org/abs/2603.20092.

Ao, P. Potential in stochastic differential equations: novel construction. *J. Phys. A*, 37:L25, 2004. URL http://iopscience.iop.org/0305-4470/37/3/L01.

Biroli, G. and Mézard, M. Generative diffusion in very large dimensions. *J. Stat. Mech.: Theory Exp.*, 2023(9): 093402, 2023. doi: 10.1088/1742-5468/acf0b8. URL https://iopscience.iop.org/article/10.1088/1742-5468/acf8ba/meta.

Biroli, G., Mézard, M., et al. Dynamical regimes of diffusion models. *Nat. Commun.*, 15:10187,

2024. doi: 10.1038/s41467-024-54414-7. URL https://www.nature.com/articles/s41467-024-54281-3.

Bonnaire, T., Urfin, R., Biroli, G., and Mezard, M. Why diffusion models don't memorize: The role of implicit dynamical regularization in training. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=BSZqpqgqM0.

Godrèche, C. and Luck, J.-M. Characterising the nonequilibrium stationary states of ornstein-uhlenbeck processes. *J. Phys. A: Math. Theor.*, 52(3):035002, 2019. doi: 10.1088/1751-8121/aaf190. URL https://doi.org/10.1088/1751-8121/aaf190.

Handke, F., Stančević, D., Koulischer, F., Demeester, T., and Ambrogioni, L. The entropic signature of class speciation in diffusion models. *arXiv preprint arXiv:2602.09651*, 2026. doi: 10.48550/arXiv.2602.09651. URL https://arxiv.org/abs/2602.09651.

Hwang, C.-R., Hwang-Ma, S.-Y., and Sheu, S.-J. Accelerating gaussian diffusions. *Ann. Appl. Probab.*, 3(3):897–913, 1993. doi: 10.1214/aoap/1177005371. URL https://doi.org/10.1214/aoap/1177005371.

Ichiki, A. and Ohzeki, M. Violation of detailed balance accelerates relaxation. *Phys. Rev. E*, 88(2):020101, 2013. doi: 10.1103/PhysRevE.88.020101. URL https://doi.org/10.1103/PhysRevE.88.020101.

Kadkhodaie, Z., Guth, F., Simoncelli, E. P., and Mallat, S. Generalization in diffusion models arises from geometry-adaptive harmonic representations. In *ICLR*, 2024. URL https://openreview.net/forum?id=ANvmVS2Yr0.

Kwon, C., Ao, P., and Thouless, D. J. Structure of stochastic dynamics near fixed points. *Proc. Natl. Acad. Sci. U.S.A.*, 102(37):13029–13033, 2005. doi: 10.1073/pnas.0506347102. URL https://doi.org/10.1073/pnas.0506347102.

Kwon, C., Noh, J. D., and Park, H. Non-equilibrium fluctuations for linear diffusion dynamics. *Phys. Rev. E*, 83(6):061145, 2011. doi: 10.1103/PhysRevE.83.061145. URL https://doi.org/10.1103/PhysRevE.83.061145.

Lelièvre, T., Nier, F., and Pavliotis, G. A. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *J. Stat. Phys.*, 152(2):237–274, 2013. doi: 10.1007/s10955-013-0782-9. URL https://link.springer.com/article/10.1007/s10955-013-0769-x.

Pham, B., Raya, G., Negri, M., Zaki, M. J., Ambrogioni, L., and Krotov, D. Memorization to generalization: Emergence of diffusion models from associative memory networks. In *New Frontiers in Associative Memories*, 2025. URL https://openreview.net/forum?id=IWZnhP3YgK.

Qian, H. A decomposition of irreversible diffusion processes without detailed balance. *J. Math. Phys.*, 54(5):053302, 2013. doi: 10.1063/1.4803847. URL https://doi.org/10.1063/1.4803847.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Stanczuk, J. P., Batzolis, G., Deveney, T., and Schönlieb, C.-B. Diffusion models encode the intrinsic dimension of data manifolds. In Salakhutdinov, R., Kolter, Z., Heller, K., Weller, A., Oliver, N., Scarlett, J., and Berkenkamp, F. (eds.), *ICML*, volume 235 of *Proceedings of Machine Learning Research*, pp. 46412–46440. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/stanczuk24a.html.

Ye, Z., Zhu, Q., Tao, M., and Chen, M. Provable separations between memorization and generalization in diffusion models. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=42gfTZzyvV.

A. Forward OU Algebra and Time Reversal

This section records the algebra behind the non-reversible OU process used in the main text. For a dataset of n points $\mathbf{a}_\mu \in \mathbb{R}^d$, the empirical data law is

$$P_0^e(\mathbf{a}) = \frac{1}{n}\sum_{\mu=1}^{n} \delta(\mathbf{a} - \mathbf{a}_\mu). \tag{16}$$

The generalized forward noising process is

$$d\mathbf{x}_t = -\mathbf{A}\mathbf{x}_t\, dt + \sqrt{2}\, d\mathbf{W}_t, \qquad \mathbf{A} = (\mathbf{I} + \mathbf{Q})\mathbf{U}, \qquad \mathbf{U} = \mathbf{U}^\top > 0, \quad \mathbf{Q} = -\mathbf{Q}^\top. \tag{17}$$

The explicit solution from $\mathbf{x}_0 = \mathbf{a}$ is

$$\mathbf{x}_t = e^{-\mathbf{A}t}\mathbf{a} + \sqrt{2}\int_0^t e^{-\mathbf{A}(t-s)} d\mathbf{W}_s. \tag{18}$$

The stationary covariance $\mathbf{\Sigma}_s$ solves the Lyapunov equation

$$\mathbf{A}\mathbf{\Sigma}_s + \mathbf{\Sigma}_s\mathbf{A}^\top = 2\mathbf{I}. \tag{19}$$

For the chosen decomposition, $\mathbf{\Sigma}_s = \mathbf{U}^{-1}$ because

$$(\mathbf{I} + \mathbf{Q})\mathbf{U}\mathbf{U}^{-1} + \mathbf{U}^{-1}\mathbf{U}(\mathbf{I} + \mathbf{Q})^\top = \mathbf{I} + \mathbf{Q} + \mathbf{I} - \mathbf{Q} = 2\mathbf{I}. \tag{20}$$

Thus the anti-symmetric perturbation changes the probability current but leaves the invariant Gaussian measure fixed.

The stochastic covariance in (18) is

$$\mathbf{\Sigma}_{\text{sto}}(t) = 2\int_0^t e^{-\mathbf{A}s} e^{-\mathbf{A}^\top s}\, ds. \tag{21}$$

It can also be written in closed Lyapunov form:

$$\mathbf{\Sigma}_{\text{sto}}(t) = \mathbf{\Sigma}_s - e^{-\mathbf{A}t}\mathbf{\Sigma}_s e^{-\mathbf{A}^\top t}. \tag{22}$$

Indeed, differentiating $e^{-\mathbf{A}t}\mathbf{\Sigma}_s e^{-\mathbf{A}^\top t}$ gives

$$\frac{d}{dt}\left(e^{-\mathbf{A}t}\mathbf{\Sigma}_s e^{-\mathbf{A}^\top t}\right) = -e^{-\mathbf{A}t}\left(\mathbf{A}\mathbf{\Sigma}_s + \mathbf{\Sigma}_s\mathbf{A}^\top\right) e^{-\mathbf{A}^\top t} = -2e^{-\mathbf{A}t} e^{-\mathbf{A}^\top t}, \tag{23}$$

and integration from 0 to t gives (22).

For time reversal, let $\tau = t_f - t$ and $\mathbf{y}_\tau = \mathbf{x}_{t_f - \tau}$. For a forward diffusion $d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, t)\, dt + \sqrt{2}\, d\mathbf{W}_t$, the reverse-time process satisfies

$$d\mathbf{y}_\tau = [-\mathbf{f}(\mathbf{y}_\tau, t) + 2\nabla_{\mathbf{y}} \log p_t(\mathbf{y}_\tau)]\, d\tau + \sqrt{2}\, d\mathbf{W}_\tau, \qquad t = t_f - \tau. \tag{24}$$

Specializing to $\mathbf{f}(\mathbf{x}, t) = -\mathbf{A}\mathbf{x}$ gives

$$d\mathbf{y}_\tau = \left[\mathbf{A}\mathbf{y}_\tau + 2\nabla_{\mathbf{y}} \log p_{t_f - \tau}(\mathbf{y}_\tau)\right] d\tau + \sqrt{2}\, d\mathbf{W}_\tau. \tag{25}$$

The only structural change relative to the reversible linear model is therefore the additional linear reverse drift $\mathbf{Q}\mathbf{U}\mathbf{y}_\tau$; the score term remains the score of the forward marginal generated by the chosen noising process.

B. Anti-Symmetric Perturbation Families

The optimized non-reversible perturbation used in the experiments follows the construction of Lelièvre et al. (Lelièvre et al., 2013). For a fixed positive definite $\mathbf{U}$, the spectral optimization problem is

$$\max_{\mathbf{Q}^\top=-\mathbf{Q}} \min \Re\sigma((\mathbf{I}+\mathbf{Q})\mathbf{U}). \tag{26}$$

The maximal achievable value is the mean curvature:

Proposition 1 (Maximal spectral gap (Lelièvre et al., 2013))**.** *For* $\mathbf{A}=(\mathbf{I}+\mathbf{Q})\mathbf{U}$ *with* $\mathbf{Q}^\top=-\mathbf{Q}$ *and* $\mathbf{U}=\mathbf{U}^\top>0$*,*

$$\max_{\mathbf{Q}^\top=-\mathbf{Q}} \min \Re\sigma(\mathbf{A}) = \frac{\mathrm{Tr}(\mathbf{U})}{d}. \tag{27}$$

At a high level, the constructive algorithm starts from an orthonormal basis and repeatedly identifies one vector whose quadratic level $(\psi, \mathbf{U}\psi)$ lies above $\mathrm{Tr}(\mathbf{U})/d$ and one whose level lies below it. A two-dimensional rotation within their span produces a new vector whose level is exactly $\mathrm{Tr}(\mathbf{U})/d$, followed by Gram–Schmidt orthonormalization of the remaining directions. Iterating this pair-rotation procedure yields an equilibrated basis from which an anti-symmetric perturbation attaining the maximal spectral gap can be constructed.

This construction is used only as a geometry-aware benchmark for transient regime control. The theorem concerns the asymptotic exponential relaxation rate of the forward OU process. Speciation and collapse are finite-time reverse-process regimes, so their behavior must still be evaluated through (7) and (11). In particular, non-normal drift matrices can introduce finite-time prefactors and transient amplification that are not captured by the spectral exponent alone.

The Simple perturbation family used for comparison is deliberately less adapted:

$$Q_{ij} = q \quad (i<j), \qquad Q_{ji} = -q, \qquad Q_{ii} = 0. \tag{28}$$

It breaks detailed balance and introduces rotational probability currents, but it does not target the slow eigenspaces of $\mathbf{U}$ or the empirical class-separation direction. Its role is to test whether generic anti-symmetric perturbations are sufficient, as opposed to perturbations designed from the drift geometry.

C. Detailed Speciation Derivation

C.1. Landau Expansion

This section gives the derivation behind (7). The forward marginal can be written as

$$P_t(\mathbf{x}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_{\mathrm{sto}}(t))}} \exp\left(-\frac{1}{2}\mathbf{x}^\top\boldsymbol{\Sigma}_{\mathrm{sto}}^{-1}(t)\mathbf{x} + g(\mathbf{x})\right), \tag{29}$$

where

$$g(\mathbf{x}) = \log \int d\mathbf{a}\, P_0(\mathbf{a}) \exp\Big[-\frac{1}{2}(e^{-\mathbf{A}t}\mathbf{a})^\top\boldsymbol{\Sigma}_{\mathrm{sto}}^{-1}(t)(e^{-\mathbf{A}t}\mathbf{a}) + \mathbf{x}^\top\boldsymbol{\Sigma}_{\mathrm{sto}}^{-1}(t)e^{-\mathbf{A}t}\mathbf{a}\Big]. \tag{30}$$

Let

$$\Phi(\mathbf{a};\mathbf{x},t) = -\frac{1}{2}\mathbf{a}^\top e^{-\mathbf{A}^\top t}\boldsymbol{\Sigma}_{\mathrm{sto}}^{-1}(t)e^{-\mathbf{A}t}\mathbf{a} + \mathbf{x}^\top\boldsymbol{\Sigma}_{\mathrm{sto}}^{-1}(t)e^{-\mathbf{A}t}\mathbf{a}. \tag{31}$$

When the propagated signal is small enough for a local expansion, $\exp(\Phi) = 1 + \Phi + \frac{1}{2}\Phi^2 + \cdots$. Integrating term by term with respect to P_0 gives the $\mathbf{x}$-dependent contributions

$$\langle\text{linear}\rangle = \mathbf{x}^\top\boldsymbol{\Sigma}_{\mathrm{sto}}^{-1}(t)e^{-\mathbf{A}t}\langle\mathbf{a}\rangle, \tag{32}$$

$$\langle\text{quadratic}\rangle = \frac{1}{2}\mathbf{x}^\top\boldsymbol{\Sigma}_{\mathrm{sto}}^{-1}(t)e^{-\mathbf{A}t}\langle\mathbf{a}\mathbf{a}^\top\rangle e^{-\mathbf{A}^\top t}\boldsymbol{\Sigma}_{\mathrm{sto}}^{-1}(t)\mathbf{x}. \tag{33}$$

Using $\log(1+z) = z - \frac{1}{2}z^2 + \cdots$, the second-order term from the square of the linear contribution subtracts

$$\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}_{\text{sto}}^{-1}(t) e^{-\mathbf{A}t} \langle \mathbf{a} \rangle \langle \mathbf{a} \rangle^\top e^{-\mathbf{A}^\top t} \boldsymbol{\Sigma}_{\text{sto}}^{-1}(t) \mathbf{x}. \tag{34}$$

Thus the total quadratic form in $-\log P_t(\mathbf{x})$ has curvature

$$\mathbf{M}(t) = \boldsymbol{\Sigma}_{\text{sto}}^{-1}(t) - \boldsymbol{\Sigma}_{\text{sto}}^{-1}(t) e^{-\mathbf{A}t} \boldsymbol{\Sigma}_{\text{data}} e^{-\mathbf{A}^\top t} \boldsymbol{\Sigma}_{\text{sto}}^{-1}(t), \tag{35}$$

where

$$\boldsymbol{\Sigma}_{\text{data}} = \langle \mathbf{a}\mathbf{a}^\top \rangle - \langle \mathbf{a} \rangle \langle \mathbf{a} \rangle^\top. \tag{36}$$

For speciation, one must isolate the covariance component that changes the topology of the density rather than the component that only broadens a single Gaussian. For a symmetric two-component Gaussian mixture with means $\pm\mathbf{m}$ and isotropic within-component covariance $\sigma^2\mathbf{I}$,

$$\boldsymbol{\Sigma}_{\text{data}} = \sigma^2 \mathbf{I} + \mathbf{m}\mathbf{m}^\top, \qquad \boldsymbol{\Sigma}_{\text{B}} = \mathbf{m}\mathbf{m}^\top. \tag{37}$$

The isotropic part renormalizes the local width; the symmetry-breaking part $\boldsymbol{\Sigma}_{\text{B}}$ drives the order-parameter instability. Replacing $\boldsymbol{\Sigma}_{\text{data}}$ in (35) by $\boldsymbol{\Sigma}_{\text{B}}$, the stability loss is equivalent to

$$\lambda_{\min}\Big(\boldsymbol{\Sigma}_{\text{sto}}(t) - e^{-\mathbf{A}t} \boldsymbol{\Sigma}_{\text{B}} e^{-\mathbf{A}^\top t} \Big) = 0. \tag{38}$$

The equivalence follows from a congruence transformation by the positive-definite matrix $\boldsymbol{\Sigma}_{\text{sto}}^{-1/2}(t)$.

C.2. Simultaneously Diagonalizable Limit

Assume that $\mathbf{A}$ and $\boldsymbol{\Sigma}_{\text{B}}$ are simultaneously orthogonally diagonalizable:

$$\mathbf{P}^\top \mathbf{A} \mathbf{P} = \text{diag}(d_1, \ldots, d_d), \qquad \mathbf{P}^\top \boldsymbol{\Sigma}_{\text{B}} \mathbf{P} = \text{diag}(c_1, \ldots, c_d), \qquad d_i > 0. \tag{39}$$

In this basis,

$$\boldsymbol{\Sigma}_{\text{sto}}(t) = \mathbf{P}\, \text{diag}\left(\frac{1 - e^{-2d_1 t}}{d_1}, \ldots, \frac{1 - e^{-2d_d t}}{d_d} \right) \mathbf{P}^\top \tag{40}$$

and

$$e^{-\mathbf{A}t} \boldsymbol{\Sigma}_{\text{B}} e^{-\mathbf{A}^\top t} = \mathbf{P}\, \text{diag}\big(c_1 e^{-2d_1 t}, \ldots, c_d e^{-2d_d t} \big)\, \mathbf{P}^\top. \tag{41}$$

The scalar stability entries are therefore

$$\widetilde{m}_i(t) = \frac{1 - e^{-2d_i t}}{d_i} - c_i e^{-2d_i t}. \tag{42}$$

If $\Lambda = \max_i c_i$ is the leading symmetry-breaking eigenvalue and d_Λ is the drift eigenvalue in that direction, setting $\widetilde{m}_\Lambda(t_S) = 0$ gives

$$\frac{1 - e^{-2d_\Lambda t_S}}{d_\Lambda} = \Lambda e^{-2d_\Lambda t_S}, \qquad t_S = \frac{\log(1 + \Lambda d_\Lambda)}{2d_\Lambda}. \tag{43}$$

The isotropic reversible expression follows by setting $d_\Lambda = 1$.

D. Gaussian-Mixture Cloning Observable

This section records the analytic form of the cloning observable used to validate the speciation criterion in the Gaussian-mixture experiments. Consider two equally weighted Gaussian clusters with means $\pm\mathbf{m}$ and initial within-cluster covariance $\sigma^2\mathbf{I}$. Under the linear forward process, the marginal at time t can be written as

$$P_t(\mathbf{y}) = \frac{1}{2}G(\mathbf{y}, \boldsymbol{\mu}(t), \boldsymbol{\Sigma}_G(t)) + \frac{1}{2}G(\mathbf{y}, -\boldsymbol{\mu}(t), \boldsymbol{\Sigma}_G(t))\,, \tag{44}$$

where

$$\boldsymbol{\mu}(t) = e^{-\mathbf{A}t}\mathbf{m}, \qquad \boldsymbol{\Sigma}_G(t) = \boldsymbol{\Sigma}_{\text{sto}}(t) + \sigma^2 e^{-\mathbf{A}t}e^{-\mathbf{A}^\top t}. \tag{45}$$

Here $G(\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the d-dimensional Gaussian density with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

Given a noisy state $\mathbf{y}$ at time t, the posterior probabilities of the two components determine whether two independent reverse clones are likely to terminate in the same component. Integrating over $\mathbf{y}$ gives

$$\begin{aligned}\phi(t) &= \frac{1}{2}\int d\mathbf{y}\, \frac{G(\mathbf{y}, \boldsymbol{\mu}(t), \boldsymbol{\Sigma}_G(t))^2 + G(\mathbf{y}, -\boldsymbol{\mu}(t), \boldsymbol{\Sigma}_G(t))^2}{G(\mathbf{y}, \boldsymbol{\mu}(t), \boldsymbol{\Sigma}_G(t)) + G(\mathbf{y}, -\boldsymbol{\mu}(t), \boldsymbol{\Sigma}_G(t))} \\ &= 1 - \int d\mathbf{y}\, \frac{G(\mathbf{y}, \boldsymbol{\mu}(t), \boldsymbol{\Sigma}_G(t))\, G(\mathbf{y}, -\boldsymbol{\mu}(t), \boldsymbol{\Sigma}_G(t))}{G(\mathbf{y}, \boldsymbol{\mu}(t), \boldsymbol{\Sigma}_G(t)) + G(\mathbf{y}, -\boldsymbol{\mu}(t), \boldsymbol{\Sigma}_G(t))}.\end{aligned} \tag{46}$$

Substituting the Gaussian density and collecting terms gives the equivalent form

$$\phi(t) = 1 - \frac{1}{2}\exp\left(-\frac{1}{2}\boldsymbol{\mu}(t)^\top \boldsymbol{\Sigma}_G(t)^{-1}\boldsymbol{\mu}(t)\right)\int \frac{d\mathbf{y}}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_G(t))}}\frac{\exp\left(-\frac{1}{2}\mathbf{y}^\top \boldsymbol{\Sigma}_G(t)^{-1}\mathbf{y}\right)}{\cosh(\mathbf{y}^\top \boldsymbol{\Sigma}_G(t)^{-1}\boldsymbol{\mu}(t))}. \tag{47}$$

For numerical evaluation, the Gaussian measure in (47) can be whitened by $\mathbf{z} = \boldsymbol{\Sigma}_G(t)^{-1/2}\mathbf{y}$. In the isotropic reversible case, all directions orthogonal to $\mathbf{m}$ factorize and the expression reduces to the standard one-dimensional integral. For the non-reversible drifts used here, $\boldsymbol{\Sigma}_G(t)$ is generally non-diagonal, so $\phi(t)$ is measured by cloned reverse simulations and compared against the matrix-predicted t_S.

E. Random Energy Model Analysis for Collapse

E.1. General Linear Drift

Let

$$\mathbf{M}_t = e^{-\mathbf{A}t}, \qquad \boldsymbol{\Gamma}_t = \boldsymbol{\Sigma}_{\text{sto}}(t). \tag{48}$$

The empirical forward density is the Gaussian mixture

$$P_t^e(\mathbf{x}) \propto \sum_{\mu=1}^{n} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{M}_t\mathbf{a}_\mu)^\top \boldsymbol{\Gamma}_t^{-1}(\mathbf{x} - \mathbf{M}_t\mathbf{a}_\mu)\right]. \tag{49}$$

We probe the density near the forward image of a reference point:

$$\mathbf{x} = \mathbf{M}_t\mathbf{a}_1 + \mathbf{z}, \qquad \mathbf{z} \sim \mathcal{N}(0, \boldsymbol{\Gamma}_t). \tag{50}$$

For a background point $\mu \neq 1$, define $\boldsymbol{\delta}_\mu = \mathbf{a}_1 - \mathbf{a}_\mu$. Its energy is

$$E_\mu(t) = \frac{1}{2}\left(\mathbf{M}_t\boldsymbol{\delta}_\mu + \mathbf{z}\right)^\top \boldsymbol{\Gamma}_t^{-1}\left(\mathbf{M}_t\boldsymbol{\delta}_\mu + \mathbf{z}\right). \tag{51}$$

The REM approximation treats the background differences as independent Gaussian vectors, $\boldsymbol{\delta}_\mu \sim \mathcal{N}(0, 2\sigma_0^2\mathbf{I})$. Introduce

$$\mathbf{K}_t = \mathbf{M}_t^\top \boldsymbol{\Gamma}_t^{-1}\mathbf{M}_t, \qquad \mathbf{h}_t = \mathbf{M}_t^\top \boldsymbol{\Gamma}_t^{-1}\mathbf{z}. \tag{52}$$

Then

$$E_\mu(t) = \frac{1}{2}\boldsymbol{\delta}_\mu^\top \mathbf{K}_t \boldsymbol{\delta}_\mu + \boldsymbol{\delta}_\mu^\top \mathbf{h}_t + \frac{1}{2}\mathbf{z}^\top \boldsymbol{\Gamma}_t^{-1}\mathbf{z}. \tag{53}$$

The Gaussian integral over $\boldsymbol{\delta}_\mu$ gives

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\delta}} \exp[-\beta E_\mu(t)] = \det\big(\mathbf{I} + 2\beta\sigma_0^2\mathbf{K}_t\big)^{-1/2} \\ \times \exp\left[-\frac{\beta}{2}\mathbf{z}^\top\boldsymbol{\Gamma}_t^{-1}\mathbf{z} + \beta^2\sigma_0^2\mathbf{h}_t^\top\left(\mathbf{I} + 2\beta\sigma_0^2\mathbf{K}_t\right)^{-1}\mathbf{h}_t\right].\end{aligned} \tag{54}$$

Since $\mathbf{z} \sim \mathcal{N}(0, \boldsymbol{\Gamma}_t)$, $d^{-1}\mathbf{z}^\top\boldsymbol{\Gamma}_t^{-1}\mathbf{z} \to 1$ in high dimension. Moreover, $\mathbf{h}_t$ has covariance $\mathbf{K}_t$. In the eigenbasis of $\mathbf{K}_t$, concentration gives

$$\frac{1}{d}\mathbf{h}_t^\top F(\mathbf{K}_t)\mathbf{h}_t \to \frac{1}{d}\mathrm{Tr}[\mathbf{K}_t F(\mathbf{K}_t)] \tag{55}$$

for bounded spectral observables F. If $\kappa_j(t)$ are the eigenvalues of $\mathbf{K}_t$, the background free-energy density is

$$g_t(\beta) = -\frac{1}{2d}\sum_{j=1}^{d}\left[\log\left(1 + 2\beta\sigma_0^2\kappa_j(t)\right) + \frac{\beta}{1 + 2\beta\sigma_0^2\kappa_j(t)}\right]. \tag{56}$$

The correct data point contributes $-E_1(t)/d \approx -1/2$. The background contribution contains the pattern entropy $\alpha = (\log n)/d$ and is $\alpha + g_t(1)$. Equating the reference and background free-energy densities gives

$$2\alpha + 1 = \frac{1}{d}\sum_{j=1}^{d}\left[\log\left(1 + 2\sigma_0^2\kappa_j(t_C)\right) + \frac{1}{1 + 2\sigma_0^2\kappa_j(t_C)}\right], \tag{57}$$

which is the implicit REM criterion used in the main text.

The trace identity explains only a restricted robustness mechanism. For $\mathbf{A} = (\mathbf{I} + \mathbf{Q})\mathbf{U}$,

$$\mathrm{Tr}(\mathbf{A}) = \mathrm{Tr}(\mathbf{U}) + \mathrm{Tr}(\mathbf{QU}) = \mathrm{Tr}(\mathbf{U}), \tag{58}$$

because $\mathbf{U}$ is symmetric and $\mathbf{Q}$ is anti-symmetric. Therefore $\det(e^{-\mathbf{A}t}) = \exp[-t\mathrm{Tr}(\mathbf{U})]$ is independent of $\mathbf{Q}$. However, the REM criterion depends on the full spectrum of $\mathbf{K}_t = \mathbf{M}_t^\top\boldsymbol{\Gamma}_t^{-1}\mathbf{M}_t$. The trace identity therefore does not imply $\mathbf{Q}$-independence of t_C.

E.2. Diagonal Reversible Limit

Restrict to $\mathbf{Q} = 0$ and $\mathbf{A} = \mathbf{U} = \mathrm{diag}(u_1, \ldots, u_d)$. Then

$$\mathbf{M}_t = \mathrm{diag}(e^{-u_1 t}, \ldots, e^{-u_d t}), \qquad \boldsymbol{\Gamma}_t = \mathrm{diag}\left(\frac{1 - e^{-2u_1 t}}{u_1}, \ldots, \frac{1 - e^{-2u_d t}}{u_d}\right). \tag{59}$$

Thus

$$\kappa_j(t) = \frac{u_j e^{-2u_j t}}{1 - e^{-2u_j t}}. \tag{60}$$

Substitution into the general REM criterion gives

$$2\alpha + 1 = \frac{1}{d}\sum_{j=1}^{d}\left[\log\left(1 + 2\sigma_0^2\frac{u_j e^{-2u_j t_C}}{1 - e^{-2u_j t_C}}\right) + \frac{1}{1 + 2\sigma_0^2\frac{u_j e^{-2u_j t_C}}{1 - e^{-2u_j t_C}}}\right]. \tag{61}$$

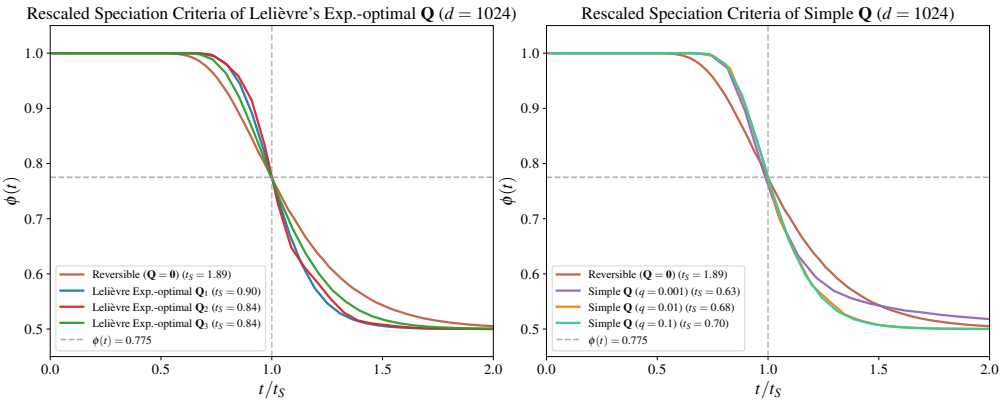

Figure 4. The predicted Gaussian-mixture speciation time aligns the transition window across controls. Rescaling time by the t_S obtained from (7) collapses the rapid change in $\phi(t)$ for both Lelièvre and Simple $\mathbf{Q}$ families.

The same expression can be checked component by component. Let $\lambda_j(t) = u_j/(1-e^{-2u_j t})$ and write $z_j \sim \mathcal{N}(0, \lambda_j(t)^{-1})$. The one-dimensional contribution to the background moment is

$$\begin{aligned} M_j(\beta) &= \mathbb{E}_{\delta_j} \exp\left[-\frac{\beta}{2}\lambda_j(t)\left(e^{-u_j t}\delta_j + z_j\right)^2\right] \\ &= \left(1 + 2\sigma_0^2\beta\lambda_j(t)e^{-2u_j t}\right)^{-1/2} \exp\left[-\frac{\beta\lambda_j(t)z_j^2}{2\left(1 + 2\sigma_0^2\beta\lambda_j(t)e^{-2u_j t}\right)}\right], \end{aligned} \tag{62}$$

where $\delta_j \sim \mathcal{N}(0, 2\sigma_0^2)$. Replacing $\lambda_j(t)z_j^2$ by its typical value 1 in high dimension gives $g_t(\beta) = d^{-1}\sum_j \log M_j(\beta)$, which is the diagonal specialization of (56). This is the diagonal reversible implicit analytical criterion. It is not a closed-form solution for t_C; t_C still appears inside the equation and is obtained by root finding.

F. Experimental and Implementation Details

F.1. Gaussian-Mixture Setup

The Gaussian-mixture experiments use two equally weighted clusters with means $\pm\mathbf{m}$ and within-cluster covariance $\sigma^2\mathbf{I}$. In the paper figures, $d = 1024$, $\mathbf{m} = \mathbf{1}$, and the initial covariance is chosen so that the symmetry-breaking signal is spread across eigendirections rather than concentrated in one coordinate. The diagonal potential profile is

$$\mathbf{U} = \mathrm{diag}(u_1, \ldots, u_d), \qquad u_j = 1 + \frac{8(j-1)}{d-1}. \tag{63}$$

This profile provides a controlled range of reversible relaxation rates.

The Simple perturbation family is

$$Q_{ij} = q \quad (i < j), \qquad Q_{ji} = -q, \qquad Q_{ii} = 0. \tag{64}$$

The Lelièvre curves use three auxiliary spectra in the construction algorithm. The labels $\mathbf{Q}_1$, $\mathbf{Q}_2$, and $\mathbf{Q}_3$ correspond respectively to an unshifted linear spectrum, the shifted spectrum used in the Remark 7 construction of Lelièvre et al., and a geometrically spaced spectrum. These families separate generic rotational perturbations from a geometry-aware perturbation designed to redistribute relaxation rates.

Speciation is evaluated by the trajectory-cloning probability $\phi(t)$. For each forward time t, a noisy state is sampled from the forward process, two reverse trajectories are initialized from that same state, and $\phi(t)$ is the fraction of pairs that terminate in the same mixture component. The vertical dashed lines in the speciation plots are the theoretical t_S values obtained by solving (7). The normalized plot in Figure 4 uses the same trajectories but displays $\phi(t)$ against t/t_S.

The Gaussian-mixture collapse diagnostic uses the same $d = 1024$ and the same diagonal $\mathbf{U}$. The reversible reference is compared with Simple $\mathbf{Q}$ at $q \in \{0.1, 0.5\}$ and with the Lelièvre Exp.-optimal perturbation. The plotted entropy quantity is a finite-dimensional proxy for the entropic criterion; the reported crossing is the last zero-crossing with $t < 1.0$, matching the plotting rule used for the paper-facing collapse figures.

F.2. Real-Data DDPM Details

The real-data experiments train DDPMs on binary image tasks with the same network architecture and training budget across drifts. The anisotropic drift matrices are constructed in the full PCA basis of the training set and then rotated back to pixel space before training. The model therefore operates in the original image coordinates; PCA is used only to define direction-dependent OU rates. All reported models are trained for $350{,}000$ steps. The Simple $\mathbf{Q}$ speciation runs use $q = 0.01$; the MNIST collapse diagnostic uses $q = 0.1$, and the ImageNet16 collapse diagnostic uses $q = 0.01$.

For MNIST, the binary task is digits 0 versus 7 with $n = 12{,}188$ and $d = 1024$. For ImageNet16, the task uses $n = 2000$ and $d = 768$. Speciation is evaluated by the same cloning protocol as in the Gaussian mixture: two reverse trajectories are initialized from a common noisy state, and their terminal classes are assigned by a ResNet-18 classifier. The empirical crossing is read using the cloning-probability threshold shown in the plotted figures.

Collapse is evaluated by two diagnostics. The first is the empirical excess entropy density

$$f^e(t) = s^{\mathrm{sep}}(t) - s^e(t), \qquad s^{\mathrm{sep}}(t) = \alpha + s_G(t), \tag{65}$$

where $s^e(t)$ is the entropy density of the empirical forward distribution and $s_G(t)$ is the entropy density of the corresponding single-Gaussian reference. The second is the nearest-neighbour switching proxy $\hat{t}_C$, defined as the last reverse-time point at which the nearest training index changes along a trajectory. These diagnostics are used only to operationalize the collapse criterion in finite-dimensional trained models.

For MNIST, the entropy zero-crossings with $t < 1$ are approximately 0.332, 0.420, and 0.112 for the reversible, Simple $\mathbf{Q}$, and Lelièvre curves; the corresponding nearest-neighbour estimates are

$$\hat{t}_C = 0.237 \pm 0.241, \quad 0.189 \pm 0.214, \quad 0.0084 \pm 0.0106. \tag{66}$$

For ImageNet16, the entropy zero-crossings are approximately 0.323, 0.337, and 0.116; the nearest-neighbour estimates are

$$\hat{t}_C = 0.214 \pm 0.141, \quad 0.214 \pm 0.186, \quad 0.011 \pm 0.013. \tag{67}$$

These values are diagnostics of finite-dimensional collapse behavior, not replacements for the entropic definition of t_C.