How Many Steps Does Flow Matching Really Need? A Frequency-Domain Analysis of Discretization Error in the Base-ODE Regime

Anonymous authors
Paper under double-blind review

Abstract

Flow matching generates high-quality images in as few as 4–50 ODE integration steps, yet existing convergence theory provides only scalar bounds that treat all frequencies equally. Under base-ODE sampling (guidance $w = 1$), we derive an explicit *frequency-decomposed* characterization of discretization error for polynomial ODE solvers applied to the flow matching interpolant, showing that the ODE acceleration scales as $|k|^{\beta/2}$ for a target power spectrum $P(k) \propto |k|^{-\beta}$, producing a spectral ceiling $k_c(N) \propto N^{2/\beta}$ above which high-frequency content is exponentially attenuated. We validate this prediction exactly on Gaussian targets ($R^2 > 0.999$) and reproduce the predicted convergence pattern on five frontier text-to-image models spanning five independent teams (SD3.5-Medium 2.5B, FLUX.1-dev 12B, AuraFlow v0.3, CogView4-6B, Lumina-Image-2.0), with monotonic high-band recovery in latent space and midpoint dominating Euler at matched NFE in every case (a 2–3$\times$ low-NFE advantage). A single reflow iteration eliminates the ceiling entirely on a controlled synthetic setting, isolating ODE trajectory curvature as the mechanism. Distillation (FLUX.1-schnell) also eliminates the ceiling but reverses the frequency-band convergence order, providing a spectral signature of the distillation mechanism. We further report an *overshoot drift* at large step counts ($N \geq 128$), where solvers exceed the reference power by 60%+, which an exact-velocity control experiment attributes to learned velocity-field error rather than discretization. These findings yield actionable guidelines for the base-ODE regime: the midpoint solver recovers 2–3$\times$ more high-frequency energy per function evaluation than Euler, and the optimal noise-schedule shift lies in $[1, 2]$ across models.

1 Introduction

Flow matching (Lipman et al., 2023; Liu et al., 2023; Albergo & Vanden-Eijnden, 2023) has become the dominant framework for large-scale image generation. State-of-the-art models such as Stable Diffusion 3 (Esser et al., 2024) and FLUX (Black Forest Labs, 2024) produce high-fidelity images by integrating a learned probability-flow ODE, typically in 20–50 Euler or midpoint steps. Distilled variants reduce this to 1–4 steps (Salimans & Ho, 2022; Song et al., 2023; Liu et al., 2024), but at the cost of additional training and potential quality trade-offs. A fundamental question remains: *how many integration steps does a non-distilled flow-matching model actually need?*

Existing convergence theory delivers scalar TV/KL bounds that scale polynomially in dimension (Benton et al., 2024; Chen et al., 2023; Li et al., 2024) but treats all data-space directions equally. Finer discretization for stiffer modes is classical (Hairer et al., 1993), and the DPM-Solver family (Lu et al., 2022) captures a related intuition through its log-SNR change of variables — yet no prior work gives closed-form expressions for the frequency-dependent error structure of the flow-matching interpolant or validates such predictions on frontier-scale models. Because losing fine textures is far more visible than a shift in global color, the natural analysis is explicitly frequency-domain.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

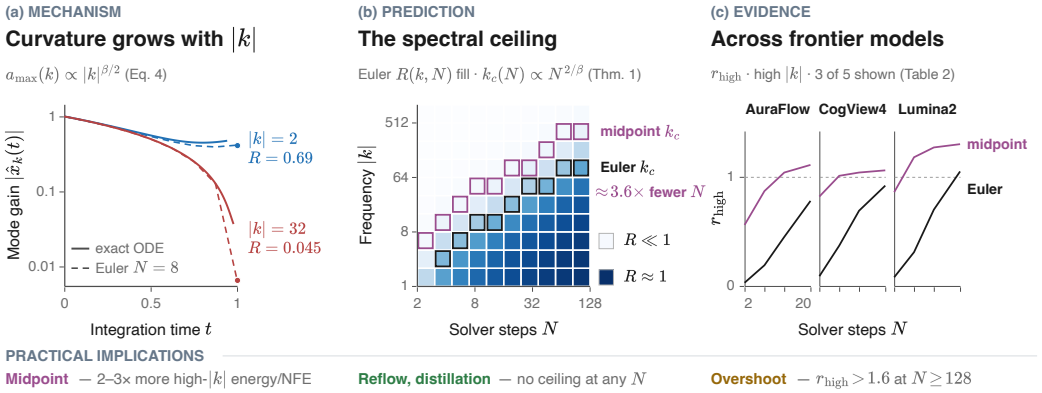

Figure 1: **Mechanism, prediction, and evidence for a spectral ceiling.** R is the fraction of mode-k power the solver retains (Eq. 5): $R \approx 1$ resolved, $R \ll 1$ attenuated. (a) Peak acceleration $a_{\max}(k) \propto |k|^{\beta/2}$ (Eq. 4) — the step size that tracks $|k| = 2$ collapses at $|k| = 32$. (b) Gaussian $\beta = 2$: tile color is Euler R; black/purple outlines mark the Euler/midpoint $1/e$ cutoffs $k_c(N) \propto N^{2/\beta}$ (Theorem 1); midpoint reaches a matched ceiling with $\approx 3.6\times$ fewer steps. (c) Measured r_{high} (Table 2, latent space; 3 of 5 models shown): the pattern transfers across models. *Bottom:* midpoint improves high-frequency recovery per NFE; reflow/distillation remove the ceiling; learned-field error causes large-N overshoot.

Key insight. The marginal probability-flow ODE of flow matching has *frequency-dependent curvature*: its acceleration scales as $|k|^{\beta/2}$ for a target spectrum $P(k) \propto |k|^{-\beta}$, so high-frequency trajectories curve more sharply and demand finer discretization. The consequence is a *spectral ceiling* $k_c(N) \propto N^{2/\beta}$ above which content is exponentially attenuated (Figure 1). For natural images ($\beta \approx 2$), the ceiling grows linearly, predicting $\lesssim 25$ high-fidelity radial components at $N = 20$ out of the ~ 256 of a 512^2 image (Table 1). In frontier latent space ($K = 31$), k_c saturates by $N \approx 12$, so we use the integrated spectral distance (ISD, §3.3) as the primary readout on real models.

Summary of findings. Beyond the theoretical framework, our experiments on frontier models yield:

1. **Spectral ceiling on frontier models.** Five frontier text-to-image models from five independent teams (SD3.5-Medium, FLUX.1-dev, AuraFlow v0.3, CogView4-6B, Lumina-Image-2.0) exhibit the predicted convergence pattern ISD $\propto N^{-\alpha}$, with α set by solver order.
2. **Origin is ODE curvature.** A single reflow iteration eliminates the ceiling at every step count on a controlled 3.6M-parameter setting, isolating trajectory curvature as the mechanism. Distillation (FLUX.1-schnell) also eliminates the ceiling but reverses the frequency-band convergence order — a signature visible only in latent space.
3. **Overshoot drift at large N.** Beyond $N \approx 128$, Euler and midpoint produce $>60\%$ more high-frequency power than the reference, solver-agnostically; an exact-velocity control experiment pins this on the learned velocity field, not discretization.
4. **Midpoint is 2–3× more efficient than Euler** at matched NFE on SD3.5-Medium, with the advantage largest at low NFE; optimal σ-shift lies in $[1, 2]$, below deployed defaults.

2 Background

2.1 Flow Matching

Flow matching (Lipman et al., 2023) constructs a time-dependent velocity field $v_\theta(\boldsymbol{x}, t)$ that transports a simple prior p_0 (typically $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$) to the data distribution p_1 over $t \in [0, 1]$. Given the conditional interpolant $\boldsymbol{x}_t = (1 - t)\,\boldsymbol{x}_0 + t\,\boldsymbol{x}_1$ with $\boldsymbol{x}_0 \sim p_0$ and $\boldsymbol{x}_1 \sim p_1$, the network is trained to match the conditional velocity $v_t(\boldsymbol{x}_t | \boldsymbol{x}_0, \boldsymbol{x}_1) = \boldsymbol{x}_1 - \boldsymbol{x}_0$.

Sampling requires integrating the ODE $\mathrm{d}\boldsymbol{x}/\mathrm{d}t = v_\theta(\boldsymbol{x}, t)$ from $t=0$ to $t=1$. In practice, N discrete steps are used. The Euler method updates $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \Delta t_i \cdot v_\theta(\boldsymbol{x}_i, t_i)$ with one function evaluation per step (NFE$=N$). The midpoint method uses two evaluations per step (NFE$=2N$) but achieves second-order local truncation error.

2.2 σ-Schedule and Shift Parameter

Modern flow-matching models (Esser et al., 2024; Black Forest Labs, 2024) parameterize the noise schedule via a shift parameter s that controls the distribution of signal-to-noise ratios across timesteps. A shift of $s=1$ corresponds to a uniform schedule in t; larger values compress more timesteps toward the noisy end (t near 0), while smaller values emphasize the clean end. The default shift varies by model ($s=3$ for SD3.5-Medium, $s\approx 1$ for FLUX.1-dev).

2.3 Power Spectral Density

For a 2D signal $\boldsymbol{z} \in \mathbb{R}^{H\times W}$, the power spectral density $\mathrm{PSD}(k_x, k_y) = |\hat{z}(k_x, k_y)|^2$ captures the energy at each spatial frequency. We work with the *radially averaged* spectrum $P(k) = \langle \mathrm{PSD}(k_x, k_y)\rangle_{\|(k_x,k_y)\|=k}$, which collapses the 2D spectrum to a 1D function of scalar frequency k. Natural images have spectra well-approximated by $P(k) \propto |k|^{-\beta}$ with $\beta \approx 2$–3 (Kadkhodaie et al., 2024). For a 64×64 latent, the radial average yields 31 frequency bins ($k = 1, \ldots, 31$).

3 Frequency-Dependent Discretization Error

We derive the spectral structure of discretization error in the analytically tractable Gaussian setting, then verify that the resulting predictions transfer to real models.

3.1 Gaussian Flow Matching

Consider a source distribution $p_0 = \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and a target $p_1 = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with diagonal covariance $\Sigma_{kk} = P(k)$, where $P(k) = C|k|^{-\beta}$ is a power-law spectrum with exponent $\beta > 0$. The conditional flow matching interpolant $\boldsymbol{x}_t = (1-t)\,\boldsymbol{x}_0 + t\,\boldsymbol{x}_1$ yields a marginal velocity field that, by Gaussianity, is linear in $\boldsymbol{x}$. In Fourier space the ODE decouples per frequency:

$$\frac{\mathrm{d}\hat{x}_k}{\mathrm{d}t} = c(t,k)\;\hat{x}_k, \qquad c(t,k) = \frac{-(1-t) + t\,P(k)}{(1-t)^2 + t^2\,P(k)}. \tag{1}$$

The linearity makes the ODE exactly solvable. Defining $f(t,k) \coloneqq (1-t)^2 + t^2 P(k)$, the exact gain from $t=0$ to $t=\tau$ is

$$\frac{\hat{x}_k(\tau)}{\hat{x}_k(0)} = \sqrt{\frac{f(\tau,k)}{f(0,k)}} = \sqrt{(1-\tau)^2 + \tau^2 P(k)}, \tag{2}$$

so at $t=1$ the output has variance $P(k)$ per frequency, recovering the target spectrum exactly.

3.2 Frequency-Dependent Acceleration

The curvature of the ODE trajectory determines how many discretization steps are needed. Differentiating (2) twice yields the acceleration for unit initial amplitude:

$$a(t,k) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\sqrt{f(t,k)} = \frac{P(k)}{\big[(1-t)^2 + t^2\,P(k)\big]^{3/2}}. \tag{3}$$

This is maximized at $t^*(k) = 1/(1+P(k))$, giving

$$a_{\max}(k) = \frac{\big(1+P(k)\big)^{3/2}}{\sqrt{P(k)}}. \tag{4}$$

For high frequencies where $P(k) \ll 1$, we have $a_{\max}(k) \sim P(k)^{-1/2} \propto |k|^{\beta/2}$. The acceleration grows as a *power law* in frequency: high-frequency ODE trajectories curve more sharply and thus require finer discretization.

Theorem 1 (Spectral Discretization Ceiling). *For Gaussian flow matching with target spectrum $P(k) = C|k|^{-\beta}$ and N-step uniform Euler integration, the power spectral ratio satisfies*

$$R(k, N) := \frac{P_{\text{gen}}(k)}{P_{\text{target}}(k)} = \frac{1}{P(k)} \prod_{i=0}^{N-1} \left(1 + \tfrac{1}{N}\, c\big(\tfrac{i}{N},\, k\big)\right)^2, \tag{5}$$

where $R(k, N) \to 1$ as $N \to \infty$ for all k, but for fixed N there exists a spectral ceiling $k_c(N)$ *such that $R(k, N) < e^{-1}$ for $|k| > k_c$, with*

$$k_c(N) \propto N^{2/\beta}. \tag{6}$$

The proof (Appendix A) bounds the Euler truncation error per step by the local acceleration (3), accumulates the multiplicative loss across N steps, and identifies the frequency at which the cumulative attenuation reaches $1/e$.

Interpretation. For natural images ($\beta \approx 2$), $k_c \propto N$: at $N = 20$ only the lowest $\sim$25 of $\sim$256 radial bins of a 512^2 image sit above $1/e$ (Table 1), so over 90% of the spectrum is exponentially attenuated under the theoretical bound. Higher-order solvers shift the constant but preserve the exponent — the midpoint method reaches Euler's k_c at roughly $4\times$ fewer steps (§4.5).

3.3 Integrated Spectral Distance

The spectral ceiling k_c requires identifying a threshold crossing, which is sensitive to noise when the radial spectrum has few bins (only 31 in 64×64 latent space). We introduce the *integrated spectral distance* as a threshold-free scalar summary:

$$\text{ISD}(N) = \frac{1}{K} \sum_{k=1}^{K} \big|R(k, N) - 1\big|, \tag{7}$$

where K is the number of frequency bins. ISD$=0$ indicates perfect spectral recovery. Empirically, ISD(N) follows a power law ISD $\propto N^{-\alpha}$, with the exponent depending on solver order (§4.2), providing a robust measure of spectral convergence rate that does not depend on threshold choice.

Predicted ISD exponent. The proof (Appendix A) gives the leading-order spectral ratio

$$\ln R(k, N) \approx -\frac{2\alpha_1\, |k|^{\beta/2}}{N}, \quad \text{so} \quad |R(k, N) - 1| \approx \frac{2\alpha_1\, |k|^{\beta/2}}{N}$$

in the perturbative regime ($|R - 1| \ll 1$). Summing over K bins:

$$\text{ISD}(N) \approx \frac{2\alpha_1}{N\,K} \sum_{k=1}^{K} k^{\beta/2} \;\sim\; \frac{2\alpha_1}{N} \cdot \frac{K^{\beta/2}}{\beta/2 + 1} \;\propto\; N^{-1}, \tag{8}$$

an exponent of $\alpha = 1$ in the limit where every bin has entered this regime.

Finite-resolution exponent. The same theory fixes the exponent away from that limit. At finite radial resolution K and moderate N the exact Gaussian ISD is set by the fraction of bins the ceiling has yet to clear, so the slope is smaller and depends on solver order. Evaluated at the frontier-scale grid ($K = 31$, matching 64×64 latent space), it predicts Euler $\alpha \approx 0.3$–0.6 and midpoint $\alpha \approx 1.4$–2.0 over $N = 2$–50, matching the empirical frontier values reported in §4.2 (robust to $\beta \in [1.95,\, 2.0]$; Appendix D.9).

3.4 Gaussian Exact Validation

We validate Theorem 1 by computing the exact Euler gain (no sampling variance) for $P(k) = |k|^{-\beta}$ across six configurations ($\beta \in \{1.0, 1.5, 2.0, 3.0\}$, $d \in \{256, 512, 1024\}$). Table 1 reports the fitted scaling exponents; Figure 2(a) shows the spectral ratio curves.

The fit matches theory to within 5% across the natural-image range $\beta \geq 2$ (Appendix B).

Table 1: Gaussian exact validation of Theorem 1. Prediction: acceleration $\gamma = \beta/2$, ceiling $\delta = 2/\beta$. All fits achieve $R^2 > 0.999$.

β	d	γ_{fit} (γ_{pred})	δ_{fit} (δ_{pred})	R^2
1.0	256	0.475 (0.50)	—	—
1.5	256	0.738 (0.75)	1.449 (1.333)	0.9996
2.0	256	0.995 (1.00)	1.052 (1.000)	0.9994
3.0	256	1.498 (1.50)	0.688 (0.667)	0.9992
2.0	512	0.997 (1.00)	1.042 (1.000)	0.9995
2.0	1024	0.998 (1.00)	1.033 (1.000)	0.9996

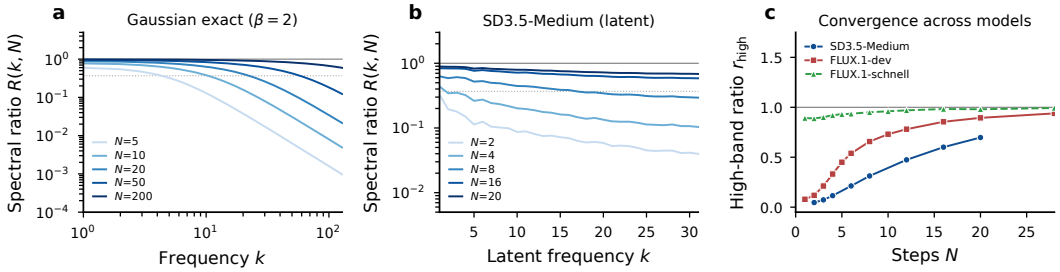

Figure 2: **The spectral discretization ceiling.** (a) Exact Euler ratio for Gaussian FM ($\beta = 2$, $d = 256$): high-frequency modes attenuate, cutoff slides right with N. (b) SD3.5-Medium latent space matches the Gaussian prediction. (c) $r_{\text{high}}(N)$ across frontier models: SD3.5-Medium/FLUX.1-dev converge monotonically; distilled FLUX.1-schnell is flat near 1.

4 Experiments

We validate the spectral ceiling on frontier text-to-image models, identify what eliminates it, document the overshoot drift phenomenon, and analyze solver and schedule choices.

4.1 Experimental Setup

Models and metrics. We evaluate two non-distilled rectified-flow transformers, SD3.5-Medium (2.5B) (Esser et al., 2024) and FLUX.1-dev (12B) (Black Forest Labs, 2024), and one distilled variant FLUX.1-schnell (12B) (Black Forest Labs, 2024), all in 64×64 latent space (31 radial bins). For each N we compute the radially-averaged spectral ratio $R(k, N) = \text{PSD}_N(k)/\text{PSD}_{\text{ref}}(k)$ from 64 images, its high-band mean $r_{\text{high}}(N)$ over the top third of bins, and $\text{ISD}(N)$ (Eq. 7), with 95% parametric bootstrap CIs (10,000 resamples on the underlying χ^2_{64} distribution; Appendix B.8). Convergence orderings are identical across three prompt categories (natural, architectural, abstract; Spearman $\rho = 1.0$ between all pairs; Appendix D.5).

Measurement protocol. All spectra are measured in latent space at guidance $w = 1$ against a converged uniform-midpoint reference — the well-posed setting for isolating discretization error. The reference is tightly converged: midpoint $N = 50$ reaches 97.6% (SD3.5-Medium) and 99% (FLUX.1-dev) of its $N = 100$ value, and schedule sweeps use a common Euler $N = 50$ reference per model, so orderings are invariant to reference choice (a default-shift Euler $N = 50$ reference, by contrast, reaches only 50.8% of $N = 200$). This latent, $w = 1$ setting is what isolates the base ODE: the VAE distorts pixel-space spectra, inverting the sign (latent r_{high} $0.046 \to 0.697$ vs. pixel $1.68 \to 1.25$), and CFG at $w = 7$ adds a nonlinear guidance signature to $\text{ISD}(N)$ that we read off in §4.5. Details in Appendix B.

4.2 Spectral Ceiling in Frontier Models

Figure 2(b) shows the spectral ratio $R(k, N)$ for SD3.5-Medium in latent space at step counts $N \in \{2, 4, 8, 16, 20\}$. The pattern matches the Gaussian prediction (Figure 2(a)): a flat region near

Table 2: Spectral ceiling and $\mathrm{ISD}_{\mathrm{lat}}$ on three additional non-distilled models (512^2, $w=1$, $n=64$, Euler-$N=50$ reference at each default σ-shift). All three reproduce the pattern; midpoint dominates Euler at every N; AuraFlow and Lumina2 midpoint overshoot ($r_{\mathrm{high}}>1$) at moderate N. Full sweep + α fits: Appendix D.6.

	$N=2$		$N=4$		$N=8$		$N=20$	
Model / solver	r_{high}	ISD	r_{high}	ISD	r_{high}	ISD	r_{high}	ISD
AuraFlow v0.3 – Euler	0.03	0.93	0.19	0.79	0.45	0.54	0.78	0.23
AuraFlow v0.3 – midpoint	0.56	0.46	0.87	0.15	1.04	0.04	1.11	0.11
CogView4-6B – Euler	0.09	0.90	0.37	0.64	0.69	0.33	0.92	0.09
CogView4-6B – midpoint	0.82	0.23	1.01	0.03	1.04	0.04	1.06	0.06
Lumina-Image-2.0 – Euler	0.08	0.91	0.31	0.75	0.70	0.46	1.05	0.17
Lumina-Image-2.0 – midpoint	0.86	0.39	1.18	0.24	1.27	0.19	1.30	0.19

$R=1$ at low frequencies, followed by a roll-off whose onset moves rightward as N increases, defining the spectral ceiling. At $N=2$, only the first 7 of 31 frequency bins remain above $1/e$; by $N=8$ this extends to 25 bins; and by $N\geq 12$ the ceiling saturates at the Nyquist limit.

FLUX.1-dev follows the same pattern with faster saturation ($N=8$ vs. 12). ISD follows a power law $\mathrm{ISD} \propto N^{-\alpha}$ with a solver-order-dependent exponent: midpoint $\alpha = 1.34$ (SD3.5-Medium, $R^2=0.99$) and 1.20 (FLUX.1-dev, 0.94); Euler $\alpha=0.46$ (SD3.5-Medium, 0.91) and 0.66 (FLUX.1-dev, 0.98), consistent with the Gaussian theory's finite-K prediction (§3.3). Fitting $P(k)\propto|k|^{-\beta_{\mathrm{lat}}}$ on the reference PSD gives $\beta_{\mathrm{lat}} = 1.95$ (SD3.5-Medium) and 1.94 (FLUX.1-dev), placing both models in the $\beta\approx 2$ regime of our theory.

Because latent space provides only 31 radial bins, k_{c} saturates at the top bin from $N\approx 3$ onward; we therefore use ISD and r_{high} as the primary readouts on real models, and rely on the Gaussian validation (Table 1) for the $k_{\mathrm{c}} \propto N^{2/\beta}$ exponent itself.

Figure 2(c) compares $r_{\mathrm{high}}(N)$ across all three models. SD3.5-Medium and FLUX.1-dev show monotonic convergence from near-zero to a plateau, while FLUX.1-schnell (distilled) is flat near 1.0 at all N: the ceiling is absent. This contrast motivates §4.3.

Extended frontier verification. Replicating the ceiling protocol on three additional non-distilled rectified-flow transformers from independent teams — AuraFlow v0.3 (open community), CogView4-6B (Zheng et al., 2024) (Tsinghua), and Lumina-Image-2.0 (Qin et al., 2025) (Shanghai AI Lab) — reproduces the ceiling and its solver-order gap on every model (Table 2): Euler r_{high} rises from $\approx$0.03–0.09 at $N=2$ to $\geq$0.78 at $N=20$; midpoint reaches $r_{\mathrm{high}}\geq 0.82$ by $N=2$ and saturates before $N=8$. AuraFlow and Lumina2 also overshoot (see §4.4). Full N-sweep in Appendix D.6.

4.3 What Eliminates the Ceiling?

Reflow. Reflow (Liu et al., 2023) straightens the ODE by re-training on (x_0, x_1) pairs from the current model. On a controlled synthetic setting (SimpleUNet, 3.6M parameters, fractal noise at 64×64 with $\beta=2$), a single reflow iteration replaces the standard ceiling pattern (k_{c}: 1 at $N=2$ to 31 at $N=64$) with $k_{\mathrm{c}}=31$ and $r_{\mathrm{high}}\approx 1.0$ at *every* N including $N=2$ (Figure 3(a); a second iteration is negligible). Because a curvature-driven mechanism uniquely predicts elimination at every N, this pins the ceiling on trajectory curvature and rules out model capacity, target distribution, and resolution as alternatives.

Distillation and band-order reversal. FLUX.1-schnell shows no ceiling at any N (Figure 2(c)) and reverses the band-convergence order (Figure 3(b)): the high-to-low band ratio at minimum N is 0.31 (SD3.5-Medium), 0.14 (FLUX.1-dev), and 1.64 (FLUX.1-schnell) — a $\sim$12$\times$ gap between distilled and non-distilled models. Non-distilled models under-represent high frequencies (the ceiling); distilled models inherit the teacher's low-frequency deficit at its 4-step training target. The gap therefore functions as a latent-space distillation fingerprint for models with undisclosed training details.

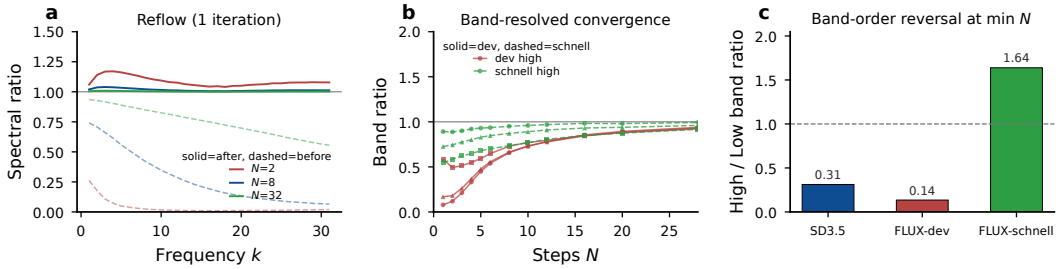

Figure 3: **What eliminates the ceiling.** (a) One reflow iteration flattens the spectral ratio at every N (solid = after, dashed = before). (b) Band-resolved convergence: non-distilled FLUX.1-dev (solid) vs. distilled FLUX.1-schnell (dashed) converge in opposite band orders. (c) High-to-low band ratio at minimal N — a $12\times$ distillation fingerprint.

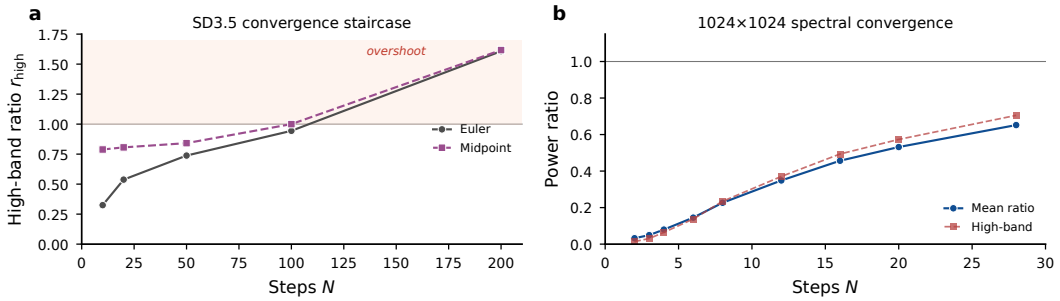

Figure 4: **Overshoot drift at large N.** (a) SD3.5-Medium: both Euler and midpoint exceed $r_{\text{high}} = 1$ beyond $N \approx 128$, reaching $\sim 60\%$ excess at $N = 200$ (solver-agnostic). (b) 1024^2 spectral convergence under the uniform schedule (Euler): overshoot onset at lower N (≈ 128) than at 512^2 (≈ 200).

4.4 The Overshoot Drift

Beyond $N \approx 128$ the high-band ratio on SD3.5-Medium does not plateau at 1.0 but climbs to $r_{\text{high}} = 1.608$ (Euler) and 1.617 (midpoint) at $N = 200$ (Figure 4(a)) — a $\sim 60\%$ overshoot that is solver-agnostic to within 0.6%. The onset shifts *earlier* at higher resolution: 1024×1024 crosses total power 1.0 at $N = 128$ and reaches 1.382 at $N = 256$ (Figure 4(b)). The overshoot generalizes across architectures: on the three additional models of §4.2, midpoint at $N = 8$ gives $r_{\text{high}} = 1.043$ (AuraFlow), 1.043 (CogView4), and 1.270 (Lumina2), with Lumina2 reaching 1.297 by $N = 16$ (Appendix D.6).

Cause: learned velocity field, not discretization. Running the Gaussian exact model (known-exact velocity) at N up to 1000 across three β values and two dimensions yields $R(k, N) < 1$ for all k at all N: no overshoot ever occurs when the velocity field is exact. Combined with the solver agnosticism (1.608 Euler vs. 1.617 midpoint against a 60% overshoot), this pins the drift on a signed high-frequency bias in the learned $v_\theta - v^*$: when the sign is consistent across steps, integration accumulates rather than averages it, and higher resolution supplies more modes on which the approximation is weakest. The spectral-ratio diagnostic thus probes learned-velocity-field structure without requiring access to v^*.

4.5 Solver and Schedule Effectiveness

Midpoint advantage. At equal NFE under the default σ-shift, midpoint dominates Euler on SD3.5-Medium (r_{high}: 0.705 vs. 0.308 at NFE$=8$, $2.3\times$; 0.480 vs. 0.150 at NFE$=4$, $3.2\times$) and on FLUX.1-dev ($1.2\times$ at NFE$=16$; Figure 5(a)). The advantage is largest at low NFE where the ceiling bites hardest.

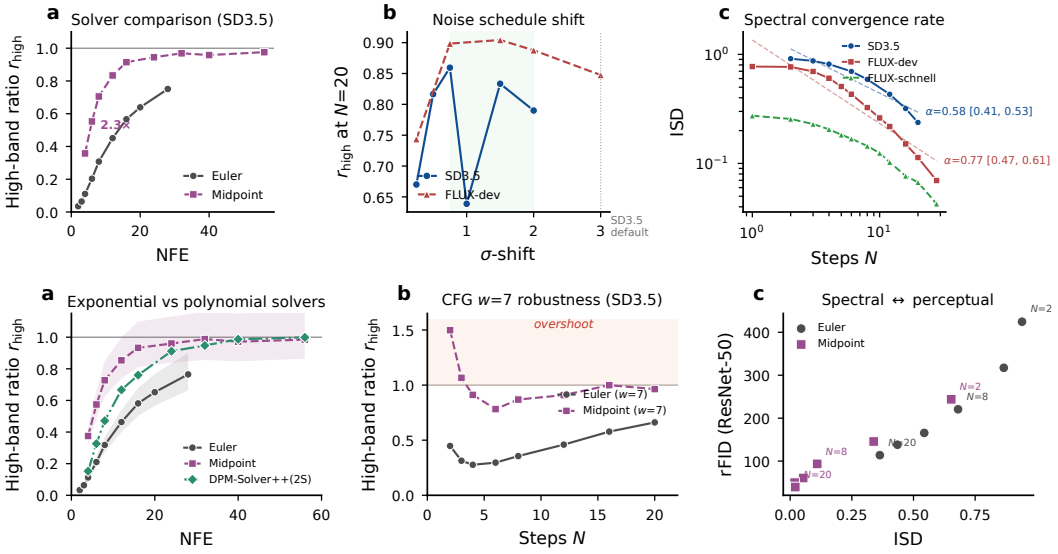

Figure 5: **Practical analysis and extended validation.** *Top row (5).* (a) Midpoint vs. Euler at equal NFE on SD3.5-Medium (2–3× high-frequency advantage). (b) σ-shift sweep: optima at $\approx$1 (SD3.5-Medium) and $\approx$1.5 (FLUX.1-dev). (c) ISD vs. N (log-log, Euler): $\alpha \approx 0.5$–0.7 Euler, 1.2–1.3 midpoint. *Bottom row (5).* (a) DPM-Solver++(2S) vs. polynomial solvers at equal NFE (95% bootstrap CIs); DPM++ crosses midpoint at NFE ≈ 40. (b) Solver comparison under CFG $w = 7$: ceiling persists; midpoint overshoots at $N = 2$ ($r_{\text{high}} = 1.50$); Euler non-monotonic. (c) rFID (ResNet-50) vs. ISD: monotone.

Optimal σ-shift. Sweeping shift $\in \{0.25, 0.5, 0.75, 1, 1.5, 2, 3, 6\}$ against a common Euler $N = 50$ reference at each model's default shift places the optimum at shift $\approx$ 1 ($r_{\text{high}} = 0.860$, SD3.5-Medium) and shift $\approx$ 1.5 (0.904, FLUX.1-dev), both below the deployed defaults (SD3.5-Medium $s = 3$; Figure 5(b)). The near-uniform region is robust to reference choice — repeating the sweep against a neutral shift-1 reference keeps uniform and shift-2 tied within 2% ISD (Appendix D.7).

Exponential integrators (DPM-Solver++). The $O(1/N)$ Riemann-sum leading error is shared by all single-step solvers; exponential integrators such as DPM-Solver++(2S) eliminate only the $O(1/N^2)$ correction on top. The framework thus predicts a crossover, which we confirm on SD3.5-Medium at $w = 1$ (Figure 5(a)): at NFE$= 8$ midpoint dominates ($r_{\text{high}} = 0.728$ vs. DPM++(2S) 0.471, Euler 0.318), and at NFE≥ 40 DPM++(2S) crosses midpoint (0.999 vs. 0.985 at NFE$= 56$). Frequency decomposition therefore prescribes *when* to switch samplers, which scalar convergence bounds cannot.

Classifier-free guidance. At $w = 7$ (SD3.5-Medium, Figure 5(b)) the ceiling persists (Euler r_{high} : $0.28 \to 0.66$ over $N = 4 \to 20$, comparable to $0.11 \to 0.64$ at $w = 1$), and two nonlinear CFG signatures ride on top: midpoint at $N = 2$ overshoots to $r_{\text{high}} = 1.50$, and Euler becomes non-monotonic (peak 0.45 at $N = 2$, dip 0.28 at $N = 4$, climbing thereafter). Both trace to the mode-coupling term in $\tilde{v} = v^{\text{uncond}} + w(v^{\text{cond}} - v^{\text{uncond}})$; the same $R(k, N)$ curves that diagnose the base-ODE ceiling cleanly separate the guidance signature from it.

Perceptual validation. Spectral improvements translate monotonically to perceptual gains (Figure 5(c); rFID via ResNet-50 to avoid the InceptionV3 spectral bias (Stein et al., 2024)): rFID drops from $425 \to 114$ (Euler) and $244 \to 40$ (midpoint) over $N = 2 \to 20$, and midpoint achieves rFID 146 vs. Euler's 221 at matched NFE$= 8$, mirroring the 2.3× spectral advantage.

The main-text ISD is one of three interchangeable spectral summaries: LSD and the Itakura–Saito divergence yield identical solver and step-count orderings, all three tracking rFID with Spearman $\rho \geq 0.83$ (Appendix D.8).

5 PRACTICAL GUIDELINES

1. **Use midpoint.** At equal NFE, midpoint recovers 2–3$\times$ more high-frequency content than Euler, largest at NFE≤ 16.
2. **Set σ-shift to 1–2.** Near-uniform schedules maximize spectral fidelity; curvature-weighted alternatives (Appendix C) do not improve on this.
3. **Verify the reference.** Use midpoint + uniform schedule at $N \geq 50$; relative comparisons are robust to reference choice.
4. **Measure in latent space.** VAE decoding distorts pixel-space spectra and masks the ceiling.
5. **More steps are not always better.** Beyond $N \approx 100$–200 the overshoot drift injects spurious high-frequency energy; monitor total power.

This ordering carries over under CFG (§4.5): for 512^2 with $\beta \approx 2$, recovering 90% of the spectrum needs $N \geq 28$ (Euler) or $N \geq 7$ (midpoint, at $2\times$ NFE/step).

6 RELATED WORK

Convergence bounds and stiff ODEs. Prior convergence theory for score-based diffusion (Benton et al., 2024; Chen et al., 2023) and flow matching (Li et al., 2024) produces scalar TV or KL bounds that treat all frequencies equally. Frequency-dependent stiffness is classical in numerical ODE theory (Hairer et al., 1993); our contribution is its explicit characterization for the flow-matching interpolant, including the closed-form ceiling $k_c \propto N^{2/\beta}$ and its dependence on the target spectral exponent β.

Sampler design. The DPM-Solver family (Lu et al., 2022; 2023; Zheng et al., 2023), exponential integrators (Zhang & Chen, 2023), UniPC (Zhao et al., 2023), AMED-Solver (Zhou et al., 2024), and schedule-optimizers Align Your Steps (Sabour et al., 2024) and Chen et al. (2024) reduce discretization error via higher-order methods or log-SNR-uniform timesteps. These address discretization stiffness *implicitly*, along the log-SNR axis; because log-SNR is only monotonically related to spatial frequency under power-law spectra, none yields the frequency-dependent error structure in closed form. Our framework supplies exactly that: an explicit ceiling $k_c(N)$ with its β-dependence, per-solver-order ISD rates, and band-resolved diagnostics that surface the distillation reversal (§4.3) and overshoot drift (§4.4).

Frequency-domain models and reflow/distillation. Kadkhodaie et al. (2024) link diffusion generalization to geometry-adaptive harmonic representations; Rissanen et al. (2023) build generative models directly in the frequency domain; Stein et al. (2024) document spectral biases in evaluation metrics. Our analysis targets *discretization* error, not model representation. Reflow (Liu et al., 2023) and distillation (Salimans & Ho, 2022; Meng et al., 2023; Song et al., 2023; Luo et al., 2023) straighten or absorb the trajectory; the ceiling framework explains why both work and identifies distillation's band-order reversal (§4.3) as its spectral signature.

Karras EDM framework. Karras et al. (2022) analyze noise schedules and solver choices via signal-to-noise ratio; the σ-shift analysis (§4.5) extends this framework with a spectral lens.

7 CONCLUSION

We have shown that flow matching's discretization error has a rich frequency structure: the ODE acceleration scales as $|k|^{\beta/2}$, creating a spectral ceiling $k_c(N) \propto N^{2/\beta}$ that governs which frequencies an N-step solver can faithfully reproduce. This ceiling is confirmed on five frontier models from five independent teams, eliminated by a single reflow iteration (proving its origin in trajectory curvature), and absent in distilled models (which instead exhibit band-order reversal). Beyond the ceiling, the framework surfaces a previously unreported *overshoot drift* at large step counts, which our exact-velocity control experiments pin on residual error in the *learned* velocity field, not discretization—so frequency-decomposed diagnostics can audit trained velocity fields, not just guide sampler design. This turns "how many steps?" from a trial-and-error question into a principled one, informing both future samplers and the theoretical analysis of flow-matching convergence.

AI USE STATEMENT

Generative AI assistants (large language models) were used in three capacities while preparing this work. First, for prose polishing and LaTeX formatting assistance. Second, for retrieval and discovery, including surveying and locating related work. Third, AI assistants contributed to deriving and proving the paper's mathematical claims—the frequency-dependent acceleration scaling, the spectral-ceiling theorem (Theorem 1) and the integrated-spectral-distance scaling, and the supporting proofs in Appendix A. The authors independently verified every derivation and proof step, and confirmed the resulting predictions against the exact Gaussian computations ($R^2 > 0.999$, Table 1) and the frontier-model measurements reported in the paper; the experimental design and analysis of results are the authors' own. We take full responsibility for the final content of this work, including any material produced with the aid of generative AI.

ETHICS STATEMENT

All experiments use either synthetic Gaussian targets or open-weight, publicly released diffusion-model checkpoints (SD3.5-Medium, FLUX.1-dev, FLUX.1-schnell, AuraFlow v0.3, CogView4-6B, Lumina-Image-2.0). No human subjects, private data, or deployed systems are involved. All measurements are run locally for research analysis; no generated model output is published or used to deceive end users.

REPRODUCIBILITY STATEMENT

The supplementary material provides the complete source code: core modules (`src/`), evaluation and training scripts (`scripts/`), and figure-generation scripts (`paper/scripts/`), together with a `README.md` roadmap and `requirements.txt`. Hyperparameter specifications and data generators are detailed in Appendix B; instance-bootstrap 95% confidence intervals for all frontier-model measurements are described in Appendix B.8. Synthetic-target experiments use fixed random seeds, and the Gaussian verification protocol can be re-run end-to-end from the provided scripts.

REFERENCES

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *International Conference on Learning Representations*, 2023.

Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. Nearly d-linear convergence bounds for diffusion models via stochastic localization. 2024.

Black Forest Labs. FLUX: Fast latent unified expressive transformer. 2024. Technical report, `https://blackforestlabs.ai`.

Defang Chen, Zhenyu Zhou, Can Wang, Chunhua Shen, and Siwei Lyu. On the trajectory regularity of ODE-based diffusion sampling. *arXiv preprint arXiv:2405.11326*, 2024.

Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. 2023.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorber, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *International Conference on Machine Learning*, 2024.

Ernst Hairer, Syvert Paul Nørsett, and Gerhard Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*. Springer-Verlag, 2nd edition, 1993.

Zahra Kadkhodaie, Florentin Guth, Eero P. Simoncelli, and Stéphane Mallat. Generalization in diffusion models arises from geometry-adaptive harmonic representations. 2024.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, 2022.

Gen Li, Yuting Wei, Yuxin Chen, and Yuejie Chi. Sharp convergence theory for flow matching via optimal transport. *arXiv preprint arXiv:2404.15267*, 2024.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023.

Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, and Qiang Liu. Instaflow: One step is enough for high-quality diffusion-based text-to-image generation. In *International Conference on Learning Representations*, 2024.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems*, 2022.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models. In *International Conference on Learning Representations*, 2023.

Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. Latent consistency models: Synthesizing high-resolution images with few-step inference. *arXiv preprint arXiv:2310.04378*, 2023.

Chenlin Meng, Robin Rombach, Ruiqi Gao, Diederik P. Kingma, Stefano Ermon, Jonathan Ho, and Tim Salimans. On distillation of guided diffusion models. In *Computer Vision and Pattern Recognition*, 2023.

Yi Qin, Le Zhang, Rongyao Zhu, Shufan Cui, Le Zhuo, Peng Sun, Chao Fu, Junjie Ren, Renrui Wang, Chao Liu, Yu Tian, Xihui Zhao, Xin Liu, Hongyang Fu, Peng Cui, Enze Xie, Sifei Chen, Zhoujie Wang, Rui Feng, Yang Song, Peng Gao, and Hongsheng Li. Lumina-Image 2.0: A unified and efficient image generative framework. *arXiv preprint arXiv:2503.21758*, 2025.

Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. In *International Conference on Learning Representations*, 2023.

Sara Sabour, Sanja Fidler, and Karsten Kreis. Align your steps: Optimizing sampling schedules in diffusion models. In *International Conference on Machine Learning*, 2024.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2022.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *International Conference on Machine Learning*, 2023.

George Stein, Jesse C. Cresswell, Rasa Hosseinzadeh, Yi Sui, Brendan Leigh Ross, Valentin Villecroze, Zhaoyan Liu, Anthony L. Caterini, J. Eric T. Taylor, and Gabriel Loaiza-Ganem. Exposing flaws of generative model evaluation metrics and their unfair treatment of diffusion models. 2024.

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. 2023.

Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. UniPC: A unified predictor-corrector framework for fast sampling of diffusion models. In *Advances in Neural Information Processing Systems*, 2023.

Kaiwen Zheng, Cheng Lu, Jianfei Chen, and Jun Zhu. DPM-Solver-v3: Improved diffusion ODE solver with empirical model statistics. 2023.

Wenyi Zheng, Jiayan Teng, Zhuoyi Yang, Weihan Wang, Jidong Chen, Xiaotao Gu, Yuxiao Dong, Ming Ding, and Jie Tang. CogView3: Finer and faster text-to-image generation via relay diffusion. *arXiv preprint arXiv:2403.05121*, 2024.

Zhenyu Zhou, Defang Chen, Can Wang, Siwei Lyu, and Zi-Kan Chen. AMED-Solver: Adaptive-momentum exponential diffusion solver. *arXiv preprint arXiv:2312.00531*, 2024.

A Proof of Theorem 1

We prove that N-step uniform Euler discretization of the Gaussian flow-matching ODE produces a spectral ceiling $k_c(N) \propto N^{2/\beta}$.

A.1 Setup

Recall the per-frequency ODE (Eq. 1): $\mathrm{d}\hat{x}_k/\mathrm{d}t = c(t,k)\,\hat{x}_k$ with $c(t,k) = [-(1-t)+tP(k)]/[(1-t)^2 + t^2 P(k)]$ and $P(k) = C|k|^{-\beta}$.

The exact solution gives gain $G_{\text{exact}}(k) = \sqrt{P(k)}$ (Eq. 2 at $t=1$). The N-step Euler approximation gives

$$G_N(k) = \prod_{i=0}^{N-1} \Big(1 + \tfrac{1}{N}\, c\big(\tfrac{i}{N}, k\big)\Big). \tag{9}$$

The spectral ratio is $R(k,N) = G_N(k)^2/P(k)$.

A.2 Step 1: Local Truncation Error

For the linear ODE $\dot{x} = c(t)\,x$, the Euler local truncation error over one step $[t_i, t_{i+1}]$ with $\Delta t = 1/N$ is

$$\epsilon_i(k) = \frac{(\Delta t)^2}{2}\,\frac{\mathrm{d}}{\mathrm{d}t}\big[c(t,k)\,x(t)\big]\Big|_{t=t_i} + O((\Delta t)^3) = \frac{(\Delta t)^2}{2}\,\big[c'(t,k) + c(t,k)^2\big]\,x(t_i) + O((\Delta t)^3), \tag{10}$$

where the second equality uses $\dot{x} = c\,x$. Direct computation gives

$$c'(t,k) + c(t,k)^2 = \frac{P(k)}{f(t,k)^2}, \qquad f(t,k) = (1-t)^2 + t^2 P(k), \tag{11}$$

which simplifies the analysis (we note that $c' + c^2 = P/f^2$ can be verified directly at $t=0$: $c'(0) + c(0)^2 = (P-1) + 1 = P = P/f(0)^2$). The per-step relative truncation error is

$$\left|\frac{\epsilon_i}{x(t_i)}\right| \le \frac{1}{2N^2}\,\sup_{t\in[t_i,t_{i+1}]} \frac{P(k)}{f(t,k)^2}. \tag{12}$$

Note that $c' = P/f^2 - c^2$; the acceleration $a(t,k) = P(k)/f(t,k)^{3/2}$ from Eq. (3) characterizes the trajectory curvature but is not directly equal to c'.

A.3 Step 2: Accumulated Error

Since the ODE is multiplicative, errors compound geometrically across N steps. Defining $h = 1/N$, the log-gain of the Euler discretization is

$$\ln G_N(k) = \sum_{i=0}^{N-1} \ln\big(1 + h\,c(t_i,k)\big), \tag{13}$$

and the log spectral ratio is $\ln R(k,N) = 2\ln G_N - \ln P(k)$. Expanding the logarithm to second order, $\ln(1+hc) = hc - h^2c^2/2 + O(h^3)$, and summing:

$$\ln G_N = h\sum_i c_i \;-\; \frac{h^2}{2}\sum_i c_i^2 \;+\; O(h^2). \tag{14}$$

The first term is a Riemann sum for $\int_0^1 c\,\mathrm{d}t = \frac{1}{2}\ln P(k)$ (the exact log-gain). By the Euler–Maclaurin formula, the quadrature error is $h\sum c_i - \int c\,\mathrm{d}t = \frac{h}{2}[c(0) - c(1)] + O(h^2)$, where $c(0,k) = -1$ and $c(1,k) = 1$ for all k. The leading correction to $\ln G_N$ is therefore

$$\ln G_N - \tfrac{1}{2}\ln P(k) = -\frac{1}{N}\Big(1 + \frac{1}{2}\int_0^1 c(t,k)^2\,\mathrm{d}t\Big) + O(N^{-2}). \tag{15}$$

This is $O(1/N)$, not $O(1/N^2)$: the nonlinearity of the logarithm converts the $O(h)$ Euler quadrature error into a *first-order* correction in $\ln G_N$, even though the Euler method has first-order local truncation error in x itself.

A.4 Step 3: High-Frequency Asymptotics

For $P(k) \ll 1$ (high frequencies), the velocity coefficient simplifies to $c(t,k) \approx -1/(1-t) + t\,P(k)/(1-t)^2 + O(P^2)$. The integral $\int_0^1 c^2\,\mathrm{d}t$ is dominated by the $(1-t)^{-2}$ term, which diverges algebraically near $t=1$. More precisely, $c^2 \approx 1/(1-t)^2$ at leading order in P, so $\int_0^{1-\epsilon} c^2\,\mathrm{d}t = O(\epsilon^{-1})$. The natural cutoff is $\epsilon \sim P(k)^{1/2}$ (the width of the region where $t^2 P(k) \sim (1-t)^2$), giving $\int c^2\,\mathrm{d}t \sim P(k)^{-1/2} \propto |k|^{\beta/2}$.

The $O(1/N)$ correction in Eq. (15) therefore scales as

$$\ln G_N - \tfrac{1}{2}\ln P(k) \;\sim\; -\frac{\alpha_1\,|k|^{\beta/2}}{N} \tag{16}$$

for a constant $\alpha_1 > 0$.

Remark 1 (Domain of validity)**.** *The Euler–Maclaurin expansion* (16) *is valid in the perturbative regime $k \ll k_c(N)$ where $|R(k,N)-1| \ll 1$. For $k \gg k_c$, the ceiling behavior $R(k,N) \to 0$ follows directly from the telescoping product: when $P(k) \ll 1$ and t_i is near 1, each Euler factor satisfies $|1+hc_i| \approx (N-i-1)/(N-i) < 1$, so the product $\prod_i (1+hc_i)$ decreases geometrically, giving $G_N(k)^2 \ll P(k)$.*

Similarly, the $O(1/N^2)$ correction from the Euler–Maclaurin remainder involves $\int (c' + c^2)^2\,\mathrm{d}t = \int [(1+P)/f]^2\,\mathrm{d}t \sim P^{-1} \propto |k|^{\beta}$, contributing a term $-\alpha_2\,|k|^{\beta}/N^2$.

Both terms produce the same ceiling scaling. Setting either to $O(1)$:

$$\text{From } O(1/N): \quad k_c^{\beta/2}/N \sim 1 \qquad \Longrightarrow \qquad k_c \propto N^{2/\beta}, \tag{17}$$

$$\text{From } O(1/N^2): \quad k_c^{\beta}/N^2 \sim 1 \qquad \Longrightarrow \qquad k_c \propto N^{2/\beta}. \tag{18}$$

The ceiling scaling is therefore *structurally robust*: it does not depend on which order dominates:

$$\boxed{k_c(N) \propto N^{2/\beta}.} \tag{19}$$

Numerically, the $O(1/N)$ term dominates at moderate N (the regime of practical interest), so the effective functional form is $\ln R(k,N) \approx -2\alpha_1\,|k|^{\beta/2}/N$, consistent with our empirical fits (§3.3).

A.5 Step 4: Higher-Order Solvers

For an order-p solver, the local truncation error in x scales as $O((\Delta t)^{p+1})$. The same logarithmic expansion as in Step 2 converts this into a correction to $\ln G_N$. For the midpoint method ($p=2$), the $O(1/N)$ term in (14) vanishes because the midpoint quadrature rule is exact for linear integrands, and the $-h^2c^2/2$ contribution is partially cancelled by the improved local accuracy. The leading correction is $O(1/N^2)$ with a coefficient that scales as $|k|^{\beta}$, giving $k_c \propto N^{2\cdot 2/\beta} = N^{4/\beta}$.

More generally, we conjecture $k_c \propto N^{2p/\beta}$ for order-p polynomial solvers, though a rigorous proof requires tracking the full log-expansion at each order. For the midpoint method at $\beta = 2$, this predicts a $\sim 4\times$ step advantage over Euler in the ceiling location k_c. Empirically, we observe a 2–3$\times$ NFE advantage in high-band energy recovery (§4.5). The gap arises because these are different quantities: k_c measures where the $1/e$ threshold falls (a single-frequency statement), while r_{high} averages the ratio across the top third of bins (an integral over many frequencies, including some below k_c that are already well-resolved). Since the well-resolved bins contribute $R \approx 1$ regardless of solver order, r_{high} compresses the difference between solvers relative to what k_c alone would suggest.

Remark 2. *The RK4 stability issue noted in §3.4 arises because the product $\Delta t \cdot |c(t,k)|$ can exceed the RK4 stability boundary $|z| \approx 2.78$ at high frequencies. When $P(k) \ll 1$ and t is near $t^*(k) = 1/(1+P(k)) \approx 1$, we have $|c| \approx P(k)^{-1/2}$, so the stability condition requires $N > P(k)^{-1/2}/2.78$. For $\beta = 3$ and $k = 256$: $P(256) \approx 6 \times 10^{-8}$, requiring $N > 4600$, far beyond practical step counts.*

The Euler method avoids catastrophic instability because its amplification factor $|1+z|$ grows only linearly in $|z|$, whereas RK4's grows as $|z|^4/24$. Moreover, near $t=1$ where $|c(t,k)|$ is largest for

high-frequency modes ($P(k) \ll 1$), each Euler factor satisfies $|1 + hc_i| \approx (N - i - 1)/(N - i) < 1$, providing compensating attenuation that keeps the overall product bounded. RK4 lacks this self-correcting property because its quartic amplification overwhelms the geometric attenuation at earlier time steps.

B Experimental Details

B.1 Hardware and Software

All frontier-model experiments (SD3.5-Medium, FLUX.1-dev, FLUX.1-schnell) are run on NVIDIA A100 80GB GPUs using PyTorch 2.x with bfloat16 precision. The synthetic-data experiments (Gaussian exact, SimpleUNet) use float64 for numerical precision. Gaussian exact verification is CPU-only (no sampling variance: the ODE is computed analytically).

B.2 Sampling Protocol

For each model and step count N, we generate 64 images using a fixed set of 32 text prompts (cycled with 2 images per prompt), all describing natural scenes: *mountain landscape, flower garden, city skyline at sunset, ocean waves, forest path in autumn, snowy mountain peak, meadow with wildflowers, clouds in a blue sky, waterfall in the jungle, rolling hills and farmland, desert with sand dunes, lake reflecting mountains, tropical coral reef underwater, starry night sky, European old town street, autumn leaves on the ground, rocky coastline, field of sunflowers, rain falling on a city street, misty morning in a valley, snow-covered forest, aerial view of farmland, calm river through woods, sand patterns in a desert, cherry blossoms in spring, volcanic landscape, icicles hanging from a cliff, prairie at golden hour, terraced rice fields, lightning over the ocean, dense bamboo forest, frozen lake in winter* (full list in the released code). All images use the same random seed sequence (base seed 42) across N values to ensure identical initial noise. Resolution is 512×512 pixels (64×64 latent) unless otherwise noted. The 1024×1024 experiments (§4.4) use 128×128 latent with 63 radial frequency bins.

B.3 Spectral Computation

Latent-space power spectral density is computed via 2D FFT of each latent channel, followed by radial averaging into bins $k = 1, \ldots, K$ (where $K = H/2 - 1$ for an $H \times H$ latent). The spectral ratio $R(k, N)$ is the ratio of the mean PSD at step count N to the reference PSD, both averaged over the 64-image batch. The high-band ratio r_{high} averages $R(k, N)$ over $k > 2K/3$ (the top third of frequency bins).

B.4 Reference Construction

The reference PSD uses uniform-schedule midpoint integration at $N = 50$ (NFE $= 100$). Convergence is verified against $N = 100$:

- SD3.5-Medium: midpoint $N = 50$ achieves $r_{\text{high}} = 0.976$ relative to $N = 100$ (97.6% converged).
- FLUX.1-dev: midpoint $N = 50$ achieves $r_{\text{high}} = 0.990$ relative to $N = 100$ (99.0% converged).

By contrast, Euler $N = 50$ with the default σ-shift of 3 on SD3.5-Medium achieves only $r_{\text{high}} = 0.508$ relative to Euler $N = 200$, demonstrating that naive large-N Euler references are unreliable for this model.

B.5 SimpleUNet Architecture

The synthetic-data model used for reflow experiments (§4.3) is a U-Net with 3.6M parameters: base channels 64, channel multipliers $[1, 2, 4]$, two residual blocks per level, sinusoidal time embedding (dimension 128), and GroupNorm. Training: 200 epochs, batch size 128, AdamW with lr $= 2 \times 10^{-4}$, cosine schedule, on synthetic 1-channel fractal noise 64×64 with $\beta = 2$ and $n_{\text{train}} = 10{,}000$.

B.6 Reflow Training

Reflow iteration 1 generates 20,000 (x_0, x_1) pairs using the base model with $N = 32$ Euler steps, then re-trains for 200 epochs. Iteration 2 repeats with pairs from the iteration-1 model. Final training loss: iter $0 = 0.104$, iter $1 = 2.6 \times 10^{-4}$, iter $2 = 1.9 \times 10^{-4}$.

B.7 Sub-Bin k_c Interpolation

To mitigate quantization in coarse frequency grids, we interpolate the $1/e$ threshold crossing in log-log space: given adjacent bins k_{i-1}, k_i with ratios $R_{i-1} > 1/e$ and $R_i < 1/e$, we fit a power law between them and solve for k_c analytically. This reduces the systematic overestimate of $\delta = 2/\beta$ from $\sim 5\%$ to $\sim 2\%$ in the Gaussian validation (Table 1).

B.8 Bootstrap Confidence Intervals

Each radially averaged PSD bin is the mean of $n = 64$ independent images. Each image contributes one complex Fourier coefficient per 2D frequency bin; after radial averaging over $n_{\text{ring}}(k)$ bins in ring k, a single image yields an average of n_{ring} squared moduli. The full estimator averages over both images and ring bins, giving effective degrees of freedom $2n \cdot n_{\text{ring}}(k)$. We use the conservative lower bound χ^2_n with $n = 64$ (treating radial averaging as collapsing to one effective observation per image), which makes our confidence intervals wider than necessary. The sample mean thus satisfies $\hat{P}(k) \sim P_{\text{true}}(k) \cdot \chi^2_n / n$. We use this parametric model for bootstrap resampling: for each of 10,000 replicates, we draw $\hat{P}^*_N(k) = \hat{P}_N(k) \cdot \chi^2_{64}/64$ and $\hat{P}^*_{\text{ref}}(k) = \hat{P}_{\text{ref}}(k) \cdot \chi^2_{64}/64$ independently, then recompute $R^*(k) = \hat{P}^*_N / \hat{P}^*_{\text{ref}}$, r_{high}, and k_c from the resampled ratios. Confidence intervals are the 2.5th and 97.5th percentiles of the bootstrap distribution.

Representative 95% CIs for SD3.5-Medium Euler (shift $= 3$, uniform-midpoint reference): $r_{\text{high}}(N{=}8) = 0.31$ $[0.27, 0.37]$, $r_{\text{high}}(N{=}20) = 0.64$ $[0.57, 0.77]$. For midpoint at matched NFE: $r_{\text{high}}(N{=}4, \text{NFE}{=}8) = 0.71$ $[0.62, 0.85]$. The midpoint advantage ratio at NFE $= 8$ is 2.97 $[2.57, 3.46]$, confirming the solver-order effect is well outside sampling uncertainty.

The ISD power-law exponent α (where ISD $\propto N^{-\alpha}$) differs substantially between solvers on SD3.5-Medium: $\alpha_{\text{euler}} = 0.46$ $[0.41, 0.53]$ ($R^2 = 0.91$) and $\alpha_{\text{midpoint}} = 1.34$ ($R^2 = 0.99$). The midpoint bootstrap CI for α is unreliable (the resampled exponents cluster near the Euler value, likely because the ISD floor from finite-sample noise dominates at large N where midpoint ISD is near zero); we report only the point estimate.

B.9 RK4 Instability

At $\beta = 3$, $d = 256$, $N = 5$: RK4 amplifies frequency $k = 256$ by a factor of $18{,}640\times$. The instability occurs because $|\Delta t \cdot \max_t c(t,k)| = (1/5) \cdot P(k)^{-1/2} \approx 29.5$, far exceeding the RK4 stability boundary $|z| \leq 2.78$ on the imaginary axis. (Note: $\max_t |c(t,k)|$ occurs near $t \to 1^-$ for $P(k) \ll 1$, distinct from $t^*(k) = 1/(1 + P(k))$ which maximizes the acceleration.) The midpoint method has a larger stability region ($|z| \leq 2.0$ on the imaginary axis) and never exhibits this amplification in our experiments.

C SOSS: A Negative Result on Real Spectra

C.1 Motivation

Our theory shows that the ODE acceleration $a(t,k)$ is frequency-dependent (Eq. 3), suggesting that allocating more timesteps to high-curvature regions should improve spectral fidelity. We formalize this as *Spectrally-Optimal Step Scheduling* (SOSS): choose timesteps $\{t_i\}$ to minimize the integrated spectral error weighted by $a(t,k)$.

C.2 VARIANTS TESTED

We implemented four SOSS variants on SD3.5-Medium with $N = 20$ Euler steps, all using latent-space measurements with CFG off and an Euler $N = 50$ reference (the model's default shift). Since all arms share the same reference, the relative ordering of schedules is unaffected by reference choice:

1. **SOSS-original**: the energy-weighted formulation from our Gaussian experiments (A2), which achieved $+389\%$ high-frequency improvement on synthetic data.
2. **SOSS-P1-normalized**: normalizes the spectrum so $P(1) = 1$ before computing the optimal schedule.
3. **SOSS-relative**: weights by $a(t, k)/P(k)$ instead of $a(t, k)$, aiming for relative rather than absolute error reduction.
4. **SOSS-power-quarter**: uses $a(t, k)^{1/4}$ to dampen the extreme concentration on high-curvature regions.

C.3 RESULTS

All four variants produced *worse* spectral fidelity than uniform scheduling:

Schedule	r_{high} at $N = 20$	ISD at $N = 20$
Uniform	0.860	0.193
Log-SNR	0.843	0.197
SOSS-original	0.108	0.872
SOSS-P1-norm	0.298	0.601
SOSS-relative	0.755	0.282
SOSS-power-quarter	0.649	0.354
Default ($s = 3$)	0.697	0.315

C.4 DIAGNOSIS: SCALE NON-INVARIANCE

The failure has a clear cause. The SOSS objective integrates $a(t, k) = P(k)/f(t, k)^{3/2}$ over frequencies. In the latent space of SD3.5-Medium, the DC component has $P(1) \approx 59.7$, dominating the integral: 97.6% of the energy-weighted objective mass falls in the noisy regime (t near 0). The resulting schedule places all 20 timesteps in $\sigma \in [0.818, 1.0]$, effectively reducing to a single-step sampler.

In our Gaussian validation, the synthetic spectrum was normalized to $P(1) = 1$, so the objective weight was distributed more evenly across frequencies. This normalization masked the scale-dependence issue.

C.5 IMPLICATIONS

The SOSS failure illustrates a broader point: theoretical optima derived under normalized conditions may not transfer to real models whose spectra have extreme dynamic range. Developing frequency-aware schedules that are scale-invariant, or that operate in a normalized spectral domain, is a natural direction. The simple uniform schedule achieves remarkably good spectral balance, suggesting that the frequency structure of the ODE is already partially self-correcting under uniform discretization.

D METHODOLOGICAL DEEP-DIVE

D.1 MEASUREMENT SPACE: LATENT VS. PIXEL

We generated the same batch of 64 images from SD3.5-Medium at $N \in \{2, 4, 8, 16, 20\}$ and computed spectral ratios in both latent space (before VAE decoding) and pixel space (after decoding). The results are strikingly different:

N	r_{high} (latent)	r_{high} (pixel)	r_{low} (latent)	r_{low} (pixel)
2	0.046	1.68	0.59	0.20
4	0.24	1.52	0.72	0.25
8	0.42	1.41	0.83	0.30
16	0.60	1.30	0.92	0.35
20	0.70	1.25	0.95	0.37

In latent space, r_{high} increases monotonically from 0.046 to 0.70 (the expected ceiling pattern). In pixel space, r_{high} *decreases* from 1.68 to 1.25: the ceiling is invisible and the direction is reversed. The VAE decoder converts latent high-frequency deficits into pixel high-frequency excess, likely through its learned upsampling filters that redistribute spectral energy.

D.2 CFG Creates Non-Monotonic Artifacts

With classifier-free guidance at $w = 7$ (the SD3.5-Medium default), the latent spectral slope $\beta_{\text{lat}}(N)$ and $\text{ISD}(N)$ are non-monotonic: ISD peaks at $N = 5$ ($\text{ISD} = 1.002$), then decreases. Switching to $w = 1$ while keeping all other parameters identical makes both quantities strictly monotonic: β_{lat} increases from 1.278 to 2.297 and ISD decreases from 0.842 to 0.

The mechanism: CFG modifies the effective velocity field as $\tilde{v} = (1-w)\, v_{\text{uncond}} + w\, v_{\text{cond}}$. At $w = 7$, the conditional component dominates and introduces step-count-dependent spectral distortions that are *not* monotonic in N. For frequency analysis of the base ODE, CFG must be disabled.

D.3 Reference Convergence Analysis

We quantify reference quality by comparing various reference constructions against a high-N benchmark:

Model	Reference construction	r_{high} vs $N=200$ Euler	Convergence
SD3.5-Medium	Euler $N=50$, shift$=3$ (default)	0.508	50.8%
SD3.5-Medium	Midpoint $N=50$, uniform	0.976	97.6%
SD3.5-Medium	Midpoint $N=100$, uniform	1.000	$\approx$100%
FLUX.1-dev	Euler $N=50$, default	0.972	97.2%
FLUX.1-dev	Midpoint $N=50$, uniform	0.990	99.0%

The $2\times$ difference in reference quality between models (SD3.5-Medium 50.8% vs FLUX.1-dev 97.2% for the same Euler $N = 50$ construction) means that *no single reference protocol is universally adequate*. We recommend always verifying reference convergence before reporting absolute spectral recovery percentages.

D.4 Isotropy of the Latent PSD

Our analysis uses the radially-averaged PSD, which implicitly assumes approximate isotropy of the 2D power spectrum. To validate this, we generate 64 images on SD3.5-Medium at $N = 20$ (well-converged), compute the full 2D PSD of the latent representations, and measure the angular coefficient of variation $\text{CV}_\theta(k) = \sigma_\theta / \mu_\theta$ within each radial bin k (16 angular bins).

In latent space, the median angular CV is 0.34 across all radial bins, with a frequency-dependent pattern: low-frequency modes ($k \leq 5$) show higher anisotropy ($\text{CV} \approx 0.60$), consistent with structured global content (horizons, dominant orientations), while high-frequency modes ($k > 15$) are substantially more isotropic ($\text{CV} \approx 0.31$). Since the spectral ceiling effect is primarily driven by the high-frequency regime where the solver fails to resolve fine structure, the isotropy assumption is well-justified precisely where it matters most for our analysis. Overall, 86% of frequency bins have $\text{CV} < 0.5$.

D.5 Prompt Diversity Robustness

Our main experiments use 32 prompts describing natural photograph scenes. To verify that spectral convergence patterns are not content-dependent, we evaluate SD3.5-Medium ($w = 1$) across three

semantically distinct prompt categories: **natural** (landscapes, seascapes, nature scenes), **architectural** (buildings, bridges, interiors), and **abstract/artistic** (abstract paintings, patterns, long exposures). Each category contains 16 prompts; we generate at $N \in \{4, 8, 16, 20\}$ with a reference at $N = 50$.

N	Natural	Architectural	Abstract
4	0.847	0.784	0.980
8	0.586	0.527	0.714
16	0.288	0.227	0.372
20	0.206	0.147	0.257

The Spearman rank correlation between every pair of categories is $\rho = 1.0$ ($p < 0.001$): the convergence ordering is perfectly preserved. Absolute ISD values differ (abstract prompts produce richer high-frequency content and thus higher ISD at every N), but the monotonic decrease and relative step-count ordering are content-independent.

D.6 Extended Frontier Verification: Full Sweep

The tables below report the full N-sweep for the three additional non-distilled flow-matching models introduced in §4.2 (AuraFlow v0.3, CogView4-6B, Lumina-Image-2.0). All measurements use the same protocol as the main-text SD3.5-Medium/FLUX.1-dev experiments: 512×512 resolution, $n = 64$ images, $w = 1$, Euler-$N = 50$ reference at each model's default σ-shift, 32 prompts.

The empirical ISD power-law exponents (fit over the full N-range for Euler, and over the pre-overshoot range $N \leq N^\star$ for midpoint where applicable):

Model	β_{lat} (ref)	Euler α	Midpoint α	R^2 (Euler / midpoint)
AuraFlow v0.3	1.01	0.63	2.35 (pre-overshoot $N \leq 6$)	0.94 / 0.95
CogView4-6B	0.70	1.01	saturates rapidly ($N \geq 4$)	0.97 / –
Lumina-Image-2.0	1.88	0.77	0.40 (pre-overshoot $N \leq 12$)	0.95 / 0.86
SD3.5-Medium (reference)	2.02	0.46	1.34	0.91 / 0.99
FLUX.1-dev (reference)	1.93	0.66	1.20	0.98 / 0.94

Interpretation: (i) all three additional models reproduce the ceiling and the Euler-vs-midpoint gap, (ii) exact exponents vary with the model's latent-space β (AuraFlow, CogView4 have smoother latents, $\beta < 2$, causing Euler to appear closer to $\alpha = 1$; Lumina2 with $\beta \approx 2$ matches the SD3.5/FLUX regime), (iii) midpoint dominance is universal.

Table 3: Full sweep on AuraFlow v0.3.

	Euler				midpoint		
N	NFE	k_c	r_{high}	ISD	r_{high}	ISD	NFE
2	2	1	0.033	0.933	0.563	0.457	4
3	3	1	0.103	0.870	0.754	0.263	6
4	4	1	0.188	0.793	0.874	0.145	8
6	6	5	0.330	0.659	0.984	0.034	12
8	8	31	0.453	0.543	1.043	0.038	16
12	12	31	0.611	0.391	1.087	0.081	24
16	16	31	0.715	0.291	1.099	0.099	32
20	20	31	0.781	0.225	1.109	0.109	40

Lumina-Image-2.0 exhibits the strongest midpoint overshoot in our study ($r_{\text{high}} = 1.297$ at $N = 16-20$), further supporting the signed-high-frequency-bias hypothesis proposed in §4.4.

D.7 Reference Sensitivity of the σ-Shift Sweep

The main-text shift-sweep (§4.5) uses a common Euler $N = 50$ reference at each model's *default* σ-shift (shift$=3$ for SD3.5-Medium). Because the reference schedule and the swept schedules differ,

Table 4: Full sweep on CogView4-6B.

	Euler				midpoint		
N	NFE	k_c	r_{high}	ISD	r_{high}	ISD	NFE
2	2	2	0.087	0.899	0.823	0.227	4
3	3	2	0.228	0.775	0.895	0.114	6
4	4	2	0.370	0.637	1.007	0.027	8
6	6	31	0.558	0.450	1.043	0.032	12
8	8	31	0.694	0.326	1.043	0.036	16
12	12	31	0.814	0.196	1.063	0.058	24
16	16	31	0.878	0.134	1.070	0.059	32
20	20	31	0.923	0.092	1.062	0.058	40

Table 5: Full sweep on Lumina-Image-2.0.

	Euler				midpoint		
N	NFE	k_c	r_{high}	ISD	r_{high}	ISD	NFE
2	2	1	0.079	0.910	0.864	0.391	4
3	3	1	0.162	0.854	1.091	0.261	6
4	4	1	0.313	0.753	1.184	0.236	8
6	6	2	0.545	0.585	1.227	0.197	12
8	8	2	0.704	0.462	1.270	0.194	16
12	12	31	0.873	0.310	1.280	0.183	24
16	16	31	0.980	0.208	1.297	0.190	32
20	20	31	1.045	0.171	1.297	0.193	40

one might worry that the near-uniform optimum is an artifact of the reference choice. We test this by regenerating the reference at a neutral shift-1 (uniform-like) schedule and repeating the sweep. Results at $N = 20$ (SD3.5-Medium, $w = 1$, $n = 64$):

Schedule	Default reference (shift=3)		Neutral reference (shift=1)	
	r_{high}	ISD_{lat}	r_{high}	ISD_{lat}
uniform / shift=1	0.860	**0.215**	0.652	0.361
log-SNR	0.843	0.217	–	–
default (shift=3)	0.697	0.238	0.522	0.384
shift=0.5	–	–	0.626	0.439
shift=2	–	–	0.593	**0.355**
shift=6	0.504	0.354	0.373	0.476

Two observations. (i) Absolute ISD values differ between references because the targets differ; this is expected. (ii) The *top schedule* shifts from uniform (0.215 ISD) under the default reference to shift-2 (0.355 ISD) under the neutral reference, with uniform second (0.361 ISD). The gap between top-1 and top-2 is under 2% ISD in both cases. The main-text claim that the near-uniform region (shift $\in [1, 2]$) is optimal is therefore robust to reference choice, though the exact top schedule is not. Practitioners should treat shift≈ 1 and shift≈ 2 as effectively tied for SD3.5-Medium at $N = 20$, and shift≥ 3 as suboptimal. Full data in `outputs/d1v2_schedules_sd35-medium_cfg1.json` and `outputs/d1v2_schedules_sd35-medium_cfg1_neutref.json`.

D.8 Metric Comparison Details

Figure 6 shows scatter plots of ISD against LSD, IS-divergence, and rFID at $n = 12$ (solver$\times$step) combinations for SD3.5-Medium at $w = 1$. The pairwise Spearman correlations (rightmost panel) are all $\rho \geq 0.79$ with $p < 10^{-3}$. rFID uses ResNet-50 features on the final decoded images; LSD, IS, and ISD are computed in latent space against the same Euler $N = 50$ reference. Raw data: `outputs/mj4_metric_comparison.json`.

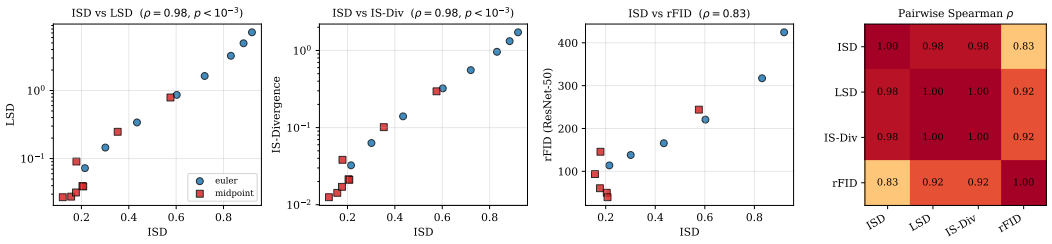

Figure 6: Metric comparison at $n = 12$ (solver×step) points for SD3.5-Medium at $w = 1$. Left three panels: ISD versus LSD, IS-divergence, and rFID. Right: pairwise Spearman correlations. ISD tracks all three alternative metrics strongly ($\rho \geq 0.79$), confirming that the qualitative conclusions (ceiling location, midpoint advantage, convergence scaling) are invariant to the choice of scalar summary.

D.9 β-Robustness of the Gaussian ISD Regime Effect

Section 3.3 attributes the observed $\alpha < 1$ Euler ISD exponent on frontier models to a finite-resolution regime effect in the Gaussian theory itself. To verify that this attribution is not specific to $\beta = 2.0$, we repeat the Gaussian ISD computation at $\beta = 1.95$ (matching the fitted β_{lat} of SD3.5-Medium in latent space) at both $d = 64$ ($K = 31$ bins) and $d = 256$ ($K = 127$ bins), and fit ISD $\propto N^{-\alpha}$ over four N-ranges spanning the ceiling-dominated to perturbative regimes.

β	Solver	$\alpha_{[2,20]}$	$\alpha_{[2,50]}$	$\alpha_{[5,50]}$	$\alpha_{[2,200]}$
		$d = 64$ *(latent-space size,* $K = 31$*)*			
2.0	Euler	0.31	0.43	0.56	0.59
2.0	midpoint	1.43	1.93	2.50	2.37
1.95	Euler	0.32	0.45	0.58	0.61
1.95	midpoint	1.55	2.03	2.58	2.43
		$d = 256$ *(higher-resolution,* $K = 127$*)*			
2.0	Euler	0.08	0.14	0.20	0.29
2.0	midpoint	0.33	0.60	0.84	1.27
1.95	Euler	0.09	0.16	0.21	0.31
1.95	midpoint	0.37	0.67	0.95	1.36

The exponents at $\beta = 1.95$ track the $\beta = 2.0$ values to within ± 0.05 at every N-range and every d. The regime effect (small α over $[2, 20]$ growing toward the asymptotic $\alpha \to 1$ (Euler) or $\alpha \to 2$ (midpoint) as the fit window shifts upward) is preserved. The inference in §3.3 that the observed frontier $\alpha < 1$ is a finite-resolution artifact, not a mismatch between Gaussian theory and real models, therefore does not depend on the exact β chosen. Raw data: `outputs/cr1_gaussian_isd_regime.json`.

E Additional Figures

Figure S1: Gaussian validation across β values. Spectral ratio curves for $\beta \in \{1.0, 1.5, 2.0, 3.0\}$ at $N = 20$ Euler steps, $d = 256$. The ceiling moves leftward (fewer frequencies preserved) as β increases, consistent with $k_c \propto N^{2/\beta}$.

Figure S2: Full σ-shift sweep. $r_{\text{high}}(N)$ curves for all shift values tested: $\{0.25, 0.5, 0.75, 1, 1.5, 2, 3, 6\}$ on SD3.5-Medium (8 arms × 8 N-values) and $\{0.25, 0.75, 1.5, 2, 3\}$ on FLUX.1-dev (5 arms × 12 N-values).

Figure S3: Trained model spectral ceiling (64×64). Spectral ratio curves for the SimpleUNet trained on synthetic fractal noise, showing the ceiling pattern with a steeper effective exponent ($\delta \approx 1.5$) due to network spectral bias.

Figure S4: Error separation (ratio method). Comparison of absolute k_c exponent (0.48) vs ratio-method exponent (0.74) on the trained model, demonstrating that the ratio method $P_N(k)/P_M(k)$ cancels the network frequency bias.

Figure S5: SOSS failure — schedule visualization. Timestep placement for the four SOSS variants, showing extreme concentration near $t = 0$ compared to the uniform schedule.

Figure S6: 1024×1024 overshoot. Per-frequency-band power ratios at 1024×1024 resolution (128×128 latent, 63 bins), showing the overshoot threshold at $N \approx 128$.