EXPLICIT ERROR PROPAGATION FOR IMPROVING TIMESTEP SCHEDULING OF GENERATIVE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Discrete timestep schedules are among the last heuristic components of diffusion and flow-based samplers. Common choices, such as time-uniform and logSNR-uniform, rely on unrealistic simplifying assumptions: locally stationary score errors, slowly varying noise levels, and uniform error propagation along the sampling trajectory. We introduce *EP-Flow*, a principled framework for timestep scheduling based on explicit error propagation. By modeling the sampler as a discrete-time dynamical system, we derive a tractable surrogate functional decomposing total generation error into three components: (i) a persistent directional bias arising from score approximation error, (ii) a residual variance from step-local stochastic error, and (iii) a local truncation error. Each component is weighted by trajectory-dependent propagators quantifying how errors at each step accumulate at the final sample. This objective applies to any initial schedule: EP-Flow both improves heuristic spacings and provides a complementary signal for refining already-learned ones. Across four standard datasets—CIFAR-10, FFHQ, AFHQv2, and ImageNet-64—EP-Flow consistently improves few-step generation over the widely used Karras schedule. It generalizes across training recipes (EDM2), refines learned schedules in Stable Diffusion v1.5, and extends to large-scale text-to-image generation with FLUX.1-dev.

1 INTRODUCTION

Diffusion and flow-based generative models (Ho et al., 2020; Lipman et al.) have been dominant for high-quality image generation tasks, yet their sampling procedures remain computationally expensive. A central yet under-theorized component is the *timestep schedule*: the choice of discretization nodes along the continuous-time trajectory. Existing approaches range from static schedules, such as uniform-logSNR (Lu et al., 2022) and the ρ-schedule of Karras et al. (2022), to distribution-level optimization methods (Sabour et al., 2024) and solver-aware error-bound approaches (Xue et al., 2024; Gao et al., 2023). However, these approaches do not explicitly model how local errors propagate through the dynamics and affect the final generated sample. We argue that this propagation structure should be the core principle for allocating function evaluations (NFE).

At each step, two distinct error sources accumulate: *score approximation error*, arising from the imperfect learned velocity field, and *discretization error*, from the numerical integration scheme. Critically, neither acts in isolation: both are weighted according to the trajectory's sensitivity structure. The resulting terminal error therefore depends on how these local errors propagate through the subsequent dynamics.

Crucially, the strength of this error propagation is highly non-uniform across modern pretrained models. On FLUX.1-dev (Labs, 2024), for example, we observe that trajectory sensitivity varies non-monotonically across noise levels (as shown in Fig. 1). As a result, different timesteps contribute very differently to the final generation error. To the best of our knowledge, existing schedule heuristics do not account for how local errors are differentially amplified along the trajectory. Uniform $\Delta\lambda$ or $\Delta\sigma$ spacing is reasonable only under restrictive assumptions, recovered as a special case of our framework. Hence, modern pretrained models are far from this regime, leaving a systematic gap for a propagation-aware schedule.

Turning this propagation-aware principle into an optimal timestep schedule, however, is challenging because the exact objective is intractable. First, forcing terms exhibit non-trivial cross-covariance

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.

across timesteps, precluding additive per-step decomposition. Second, errors at step i perturb the Jacobian at $i+1$, coupling the recursion nonlinearly and precluding closed-form analysis. We propose EP-Flow to address these two challenges by exploiting their underlying structure: a first-order linearization decouples the recursion from upstream forcing, while the observed low-rank structure of score error reduces the cross-covariance term to a tractable surrogate. This yields an optimization over N step sizes that is solved once per pretrained model, with zero inference overhead. The framework is instantiated for Euler solver and extends naturally to DPM-Solver (Lu et al., 2022).

We evaluate EP-Flow across model settings from pixel-space diffusion to large-scale text-to-image generative models. On the EDM settings (CIFAR-10, FFHQ, AFHQv2, and ImageNet-64), it consistently improves low-NFE generation over the widely used Karras schedule. The method is agnostic to the training recipe (EDM2) (Karras et al., 2024), refines the already-learned schedule of Stable Diffusion v1.5, and extends to large-scale text-to-image generation on FLUX.1-dev. We highlight our technical contributions as follows:

(i) Error decomposition. We derive a tractable surrogate objective that decomposes sampling error into bias, variance, and discretization terms with trajectory-dependent weighting, with uniform spacing arising as a special case of our framework.

(ii) Schedule optimization. We formulate a low-dimensional step-size optimization that is solved offline per pretrained model with zero inference overhead, which applies to any initial schedule, both improving heuristic spacings and providing a complementary signal for learned ones.

2 Related Works

Recent work mainly addresses the adaptive timestep scheduling in two ways:

Offline schedule optimization. This line of work optimizes a fixed schedule offline by minimizing a surrogate for sampling error. Sabour et al. (2024) use a distribution-level KL bound between the true and discretized generative SDEs, while Xue et al. (2024) derive solver-weighted per-step error bounds. Both aggregate local errors across intervals without accounting for error propagation along subsequent dynamics.

Instance-adaptive scheduling. A parallel line adapts sampling to each input rather than fixing a schedule. INDIS (Yuan et al., 2026) trains a network that maps the initial noise $\boldsymbol{x}_T$ and condition $\boldsymbol{c}$ to a per-instance timestep allocation. FastFlow (Bajpai et al., 2026) uses a multi-armed bandit to decide how many steps to skip and extrapolate the skipped velocities. These methods determine the schedule per input at inference time.

To the best of our knowledge, ours is the first method to explicitly model error propagation while optimizing a single global schedule, with model-specific statistics collected once and reused across different step budgets.

3 Preliminaries

3.1 ODE-Based Sampling for Diffusion and Flow Models

Flow matching defines a generative process directly as an ODE (Lipman et al.); diffusion models admit an equivalent probability flow ODE with identical marginals (Song et al., 2021). Both reduce to:

$$\frac{d\boldsymbol{x}_t}{dt} = f_\theta(\boldsymbol{x}_t, t), \tag{1}$$

where f_θ is a learned vector field and t parameterizes the trajectory from noise to data. We consider a numerical integration over the steps $\{t_i\}_{i=0}^N$ with step sizes $h_i := t_i - t_{i-1}$. At each step i, the error comprises *score approximation error*[1]

$$\boldsymbol{\eta}(\boldsymbol{x}^*_{t_i}, t_i) := f^*(\boldsymbol{x}^*_{t_i}, t_i) - f_\theta(\boldsymbol{x}^*_{t_i}, t_i), \tag{2}$$

[1] $\boldsymbol{x}^*(t)$ denotes the exact ODE solution; $\tilde{\boldsymbol{x}}_{t_i}$ the numerical approximation.

and the *discretization error* $\boldsymbol{\delta}_i$ from finite-order quadrature, whose form depends on solver order. Anchoring $\boldsymbol{\eta}$ at the exact $\boldsymbol{x}^*_{t_i}$ allows it to be treated as an exogenous forcing term, which is essential for the analysis in Section 4.

3.2 Solver Instantiations: DPM-Solver and Euler

Diffusion probabilistic models and DPM-Solver. A diffusion probabilistic model (Ho et al., 2020) defines a forward process $q_{0t}(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\alpha_t \boldsymbol{x}_0, \sigma_t^2 \mathbf{I})$ (Kingma et al., 2021). Song et al. (2021) show that the forward process admits an equivalent reverse process from time T to 0:

$$d\boldsymbol{x}_t = \left[f(t)\boldsymbol{x}_t - g^2(t)\nabla_{\boldsymbol{x}} \log q_t(\boldsymbol{x}_t)\right] dt + g(t)\, d\bar{\boldsymbol{w}}_t, \quad \boldsymbol{x}_T \sim q_T(\boldsymbol{x}_T), \tag{3}$$

where $\bar{\boldsymbol{w}}_t$ is a standard Wiener process in reverse time, $f(t) = \dot{\alpha}_t/\alpha_t$, and $g^2(t) = \dot{\sigma}_t^2 - 2\sigma_t^2 \dot{\alpha}_t/\alpha_t$. Because equation 3 admits a deterministic counterpart with identical marginals, Song et al. (2021) derive an associated *probability flow ODE*, enabling sampling via standard deterministic solvers. Substituting the intractable score $\nabla_{\boldsymbol{x}} \log q_t(\boldsymbol{x}_t) = -\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)/\sigma_t$ with the learned noise prediction model instantiates f_θ in equation 1 as the *diffusion ODE*:

$$\frac{d\boldsymbol{x}_t}{dt} = f(t)\boldsymbol{x}_t + \frac{g^2(t)}{2\sigma_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t), \quad \boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \tilde{\sigma}^2 \mathbf{I}). \tag{4}$$

Letting $\lambda_t := \log(\alpha_t/\sigma_t)$ denote the half log-SNR, Lu et al. (2022) show that under the change of variable $t \mapsto \lambda$, equation 4 admits the exact solution

$$\boldsymbol{x}_t = \frac{\alpha_t}{\alpha_s}\boldsymbol{x}_s - \alpha_t \int_{\lambda_s}^{\lambda_t} e^{-\lambda} \hat{\boldsymbol{\epsilon}}_\theta(\hat{\boldsymbol{x}}_\lambda, \lambda)\, d\lambda, \tag{5}$$

where $\hat{\boldsymbol{\epsilon}}_\theta(\hat{\boldsymbol{x}}_\lambda, \lambda) := \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_{t(\lambda)}, t(\lambda))$. DPM-Solver approximates this integral by locally expanding $\hat{\boldsymbol{\epsilon}}_\theta$ in λ (Lu et al., 2022); for DPM-Solver-2, the intermediate evaluation at $\boldsymbol{u}_i$ provides a second-order correction whose deviation from the first-order approximation defines the local discretization error $\boldsymbol{\delta}_i$ used in Section 4 (full update in Appendix E).

Flow matching and Euler discretization. Flow matching (Lipman et al.; Liu et al., 2022) learns a velocity field $\boldsymbol{v}_\theta(\boldsymbol{x}_t, t)$ transporting noise $\boldsymbol{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ to data $\boldsymbol{x}_0 \sim q_{\text{data}}$, integrated from $t = 1$ to $t = 0$, with $f_\theta = \boldsymbol{v}_\theta$ in equation 1. Sampling uses the first-order Euler solver: denoting the schedule $1 = \sigma_0 > \sigma_1 > \cdots > \sigma_N = 0$, each step advances

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} + h_i\, \boldsymbol{v}_\theta(\boldsymbol{x}_{i-1}, \sigma_{i-1}), \qquad h_i := \sigma_i - \sigma_{i-1} < 0. \tag{6}$$

The gap between equation 6 and the exact integral $\boldsymbol{x}^*_i = \boldsymbol{x}^*_{i-1} + \int_{\sigma_{i-1}}^{\sigma_i} \boldsymbol{v}^*(\boldsymbol{x}^*_\tau, \tau)\, d\tau$ is the local error at step i, whose accumulation across steps is the target of the schedule optimization in Section 4.

4 Error Propagation Framework for Flow Matching

First we establish the theory for flow matching with Euler discretization as the primary instantiation, and extend it to DPM-Solver in Appendix E.

4.1 Error Recursion and Propagation

Let $\boldsymbol{e}_i := \boldsymbol{x}^*_i - \tilde{\boldsymbol{x}}_i$ denote the trajectory deviation at step i. Subtracting equation 6 from the exact solution and inserting $\pm \boldsymbol{v}^*(\boldsymbol{x}^*_{i-1}, \sigma_{i-1})$ and $\pm \boldsymbol{v}_\theta(\boldsymbol{x}^*_{i-1}, \sigma_{i-1})$ gives $\boldsymbol{e}_i = \boldsymbol{e}_{i-1} + \boldsymbol{\delta}_i + \boldsymbol{\eta}_i + h_i \bar{\boldsymbol{J}}_i \boldsymbol{e}_{i-1}$, where

$$\boldsymbol{\delta}_i := \int_{\sigma_{i-1}}^{\sigma_i} \boldsymbol{v}_\theta(\boldsymbol{x}^*_\tau, \tau)\, d\tau - h_i \boldsymbol{v}_\theta(\boldsymbol{x}^*_{i-1}, \sigma_{i-1}), \qquad \boldsymbol{\eta}_i := \int_{\sigma_{i-1}}^{\sigma_i} \boldsymbol{\eta}(\boldsymbol{x}^*_\tau, \tau)\, d\tau, \tag{7}$$

are the discretization error and accumulated score error, respectively, and the last term is the deviation propagated via the trajectory-averaged Jacobian $\bar{\boldsymbol{J}}_i := \int_0^1 \nabla_{\boldsymbol{x}} \boldsymbol{v}_\theta(\tilde{\boldsymbol{x}}_{i-1} + \tau \boldsymbol{e}_{i-1}, \sigma_{i-1})\, d\tau$. The one-step recursion and its unrolled form are

$$\boldsymbol{e}_i = \underbrace{(I + h_i \bar{\boldsymbol{J}}_i)}_{P^*_i}\, \boldsymbol{e}_{i-1} + \underbrace{\boldsymbol{\delta}_i + \boldsymbol{\eta}_i}_{\mathbf{f}_i} \quad \xrightarrow{\mathbf{e}_0 = \mathbf{0}} \quad \mathbf{e}_N = \sum_{j=1}^{N} \Gamma_j\, \mathbf{f}_j, \tag{8}$$

where P_i^* is the one-step propagator, $\mathbf{f}_i$ the local forcing, and $\Gamma_j := \prod_{k=j+1}^{N} P_k^*$ the accumulated propagator from step j to the final sample. Thus each local forcing $\mathbf{f}_j$ contributes to the final error only through its accumulated propagator Γ_j, so its effect depends on where along the trajectory step j occurs.

4.2 From Exact Propagation to Tractable Surrogate

We define the cost functional $\mathcal{C} := \mathbb{E}\|\mathbf{e}_N\|^2 = \mathbb{E}\left\|\sum_j \Gamma_j \mathbf{f}_j\right\|^2$, and decompose $\mathbf{f}_j = \mathbf{f}_j^{\eta} + \mathbf{f}_j^{\mathrm{disc}}$:

$$\mathbf{f}_j^{\eta} := \int_{\sigma_{j-1}}^{\sigma_j} \boldsymbol{\eta}(\mathbf{x}_\tau^*, \tau)\, d\tau, \qquad \mathbf{f}_j^{\mathrm{disc}} := \int_{\sigma_{j-1}}^{\sigma_j} \left[\boldsymbol{v}_\theta(\boldsymbol{x}_\tau^*, \tau) - \boldsymbol{v}_\theta(\boldsymbol{x}_{j-1}^*, \sigma_{j-1})\right] d\tau, \tag{9}$$

Intractability. Directly optimization is intractable due to trajectory-dependent coupling: each propagator Γ_j depends on states determined by all preceding forcings $\{\mathbf{f}_i\}_{i<k}$. This induces a statistical dependence between Γ_j and $\mathbf{f}_j$, preventing factorization of $\mathbb{E}[\Gamma_j \mathbf{f}_j]$ and ruling out an additive decomposition of $\mathcal{C}$ into independent per-step contributions. One might hope to sidestep the coupling by imposing simplifying assumptions on $\{\mathbf{f}_j\}$: an i.i.d. or white-noise assumption $\mathbb{E}[\mathbf{f}_j^\top \mathbf{f}_{j'}] = \sigma^2 \delta_{jj'}$ decouples cross-step terms but discards the systematic bias structure of $\boldsymbol{\eta}$. We therefore resolve the coupling directly rather than assume it away.

Resolving the Γ–$\mathbf{f}$ coupling. We therefore approximate the trajectory-averaged Jacobian $\widetilde{\boldsymbol{J}}_i$ by the Jacobian evaluated at the exact trajectory point:

$$\widetilde{\boldsymbol{J}}_i \approx J_i^* := \nabla_{\boldsymbol{x}} \boldsymbol{v}_\theta(\boldsymbol{x}_{i-1}^*, \sigma_{i-1}), \tag{10}$$

so that $P_i^* = I + h_i J_i^*$ depends only on the exact trajectory and is independent of the upstream forcing $\{\mathbf{f}_j\}_{j<k}$. This first-order linearization removes the causal entanglement in the recursion, discarding only $O(\xi_0^2)$ terms in the per-step error and perturbing the cost at $O(\bar{L}_H \xi_0^3)$ overall (App. B).

4.2.1 Temporal structure: the forcing kernel

By the exact identity equation 9, the forcing splits into a score-error (η) and a discretization (disc) part, which scale as first and second order in h_j, respectively. Both enter the cost through an *uncentered second-moment* kernel of the same form, which we denote by $\mathcal{K}_\bullet(\mu, \mu') := \mathbb{E}[\mathbf{f}_\bullet(\mu)^\top \mathbf{f}_\bullet(\mu')]$, where $\bullet \in \{\eta, \mathrm{disc}\}$.

These kernels enter the cost through $\mathcal{C} = \sum_{j,k} \Gamma_j \Gamma_k \mathcal{K}(\mu_j, \mu_k)$. Splitting each forcing into mean and fluctuation gives the exact

$$\mathcal{K}_\bullet(\mu, \mu') = \underbrace{\langle \bar{\mathbf{f}}_\bullet(\mu), \bar{\mathbf{f}}_\bullet(\mu') \rangle}_{\text{coherent (mean): non-decaying}} + \underbrace{\mathrm{Cov}(\mathbf{f}_\bullet(\mu), \mathbf{f}_\bullet(\mu'))}_{\text{fluctuation}}. \qquad \bar{\mathbf{f}}_j := \mathbb{E}[\mathbf{f}_j] \tag{11}$$

The identity equation 11 isolates a coherent contribution from the mean forcing. While this term is not generally separable, its potential persistence over long separations motivates the separable far-field structure posited in Assumption A1. One might instead center spatially by subtracting the mean, but this requires an explicit estimator and cannot remove zero-mean, non-decaying modes. More generally, we write $\mathcal{K}_\bullet(\mu, \mu') = \rho_\infty, \sigma_\bullet(\mu)\sigma_\bullet(\mu') + \mathcal{K}_\bullet^{\mathrm{res}}(\mu, \mu')$, where ρ_∞ is chosen to capture the far-field behavior of the normalized kernel.

Assumption A1 (far-field plateau). The normalized kernel $\rho_\bullet(\mu, \mu') := \mathcal{K}_\bullet(\mu, \mu')/[\sigma_\bullet(\mu)\sigma_\bullet(\mu')]$ converges, as $|\mu - \mu'| \to \infty$, to a limit ρ_∞ independent of (μ, μ'). This choice forces the residual to vanish at long range:

$$\lim_{|\mu - \mu'| \to \infty} \mathcal{K}_\bullet^{\mathrm{res}}(\mu, \mu') = 0. \tag{12}$$

This is directly falsifiable: it predicts a *flat*, pair-independent plateau in the normalized error-correlation curve, rather than decay to zero or pair-dependent drift. On FLUX.1-dev, $\rho_\eta(\mu, \mu')$ indeed plateaus at $\rho_\infty \approx 0.77$. By contrast, SD v1.5 (Rombach et al., 2022) serves as a stress test: its global limit does not exist, as the plateau drifts across noise levels. We therefore relax A1 to hold *locally* within distinct regimes, yielding a piecewise-stationary far-field structure C. This extension preserves the framework, demonstrating robustness beyond the strict global-plateau setting.

Bounding Error Components. Applying AM–GM yields the surrogate $\mathcal{C} \le 2(\Delta^{\eta} + \Delta^{\mathrm{disc}})$, where $\Delta^{(\cdot)} := \mathbb{E}\big\| \sum_j \Gamma_j \mathbf{f}_j^{(\cdot)} \big\|^2$, so each forcing can be bounded separately. Once the coherent ρ_∞ component is separated, the residual has a summable tail $C_\rho := \sum_{d\ge 1} \rho(d) < \infty$, which exactly sandwiches the cost within a factor $(1 \pm C_\rho)$ (Appendix C), tightening as $C_\rho \to 0$ ($C_\rho \approx 7 \times 10^{-4}$ on FLUX.1-dev). For discretization, a first-order Taylor expansion gives $\mathbf{f}_j^{\mathrm{disc}} \approx (h_j^2/2)\dot{\boldsymbol{v}}_\theta$; since $\dot{\boldsymbol{v}}_\theta$ exhibits *no* long-range plateau, it contributes only the banded term. Each component then reduces to

$$\Delta^{\eta} \lesssim \rho_\infty \Big(\sum_j \Gamma_j a_j \Big)^2 + \sum_j \Gamma_j^2 \tilde{V}_j, \qquad \Delta^{\mathrm{disc}} \lesssim \sum_j \Gamma_j^2 \tilde{D}_j, \tag{13}$$

with $a_j := \int_{\sigma_{j-1}}^{\sigma_j} \sigma_\eta \, d\tau$, $\tilde{V}_j := (1 + 2C_\rho) V_j^{\mathrm{res}}$, and $\tilde{D}_j := (1 + 2C_\rho^d) \frac{h_j^4}{4} D_j$. Summing gives $\mathcal{C} \lesssim \rho_\infty (\sum_j \Gamma_j a_j)^2 + \sum_j \Gamma_j^2 (\tilde{V}_j + \tilde{D}_j)$, minimized subject to $\sum_j |h_j| = \Lambda$.

4.2.2 Spatial structure: the propagator

Evaluating $\Gamma_j = \prod_{k>j} (I + h_k J_k^*)$ exactly is infeasible for two reasons: *statistically*, training provides only independent marginals $\mathbf{x}_\sigma \sim q_\sigma$, with no cross-$\sigma$ pairing to estimate trajectory-level joint statistics; *computationally*, the immense dimensionality ($d \sim 10^5$) makes forming or eigendecomposing each Jacobian $J_k^* \in \mathbb{R}^{d\times d}$ intractable. The former limits us to single-σ expectations, while the latter necessitates a low-dimensional summary. We therefore use the Jacobian's normalized trace as a tractable scalar measure of local propagation strength, which can be efficiently estimated using Hutchinson's estimator:

$$\bar{g}(\sigma) = \frac{1}{d} \mathbb{E}_{p_{\mathrm{train}}} [\mathrm{Tr}(\nabla_{\boldsymbol{x}} \boldsymbol{v}_\theta(\boldsymbol{x}_\sigma, \sigma))] \approx \frac{1}{nd} \sum_{k=1}^{n} \boldsymbol{z}_k^\top \frac{\boldsymbol{v}_\theta(\boldsymbol{x} + \delta \boldsymbol{z}_k, \sigma) - \boldsymbol{v}_\theta(\boldsymbol{x}, \sigma)}{\delta}, \quad \boldsymbol{z}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{14}$$

Distortion as a joint event. $\bar{g}(\sigma)$ has a natural probabilistic interpretation as the expected local response under an isotropic prior. Define the relative growth along a unit direction $\hat{\boldsymbol{\eta}}$ via the Rayleigh quotient $Q(\hat{\boldsymbol{\eta}}) = \hat{\boldsymbol{\eta}}^\top J_{\mathrm{sym}} \hat{\boldsymbol{\eta}}$, where $J_{\mathrm{sym}} = \frac{1}{2}(J + J^\top)$, Under isotropic averaging, $\mathbb{E}_{\hat{\boldsymbol{\eta}} \sim \mathrm{Unif}(S^{d-1})}[Q(\hat{\boldsymbol{\eta}})] = \frac{1}{d}\mathrm{Tr}(J_{\mathrm{sym}}) = \bar{g}(\sigma)$. The distortion admits the exact decomposition

$$Q(\hat{\boldsymbol{\eta}}) - \bar{g} = \sum_{i=1}^{d} (\lambda_i - \bar{g})(c_i - \tfrac{1}{d}), \qquad c_i := \cos^2(\hat{\boldsymbol{\eta}}, u_i), \tag{15}$$

where $(\lambda_i - \bar{g})$ and $(c_i - \frac{1}{d})$ capture spectral and directional anisotropy, respectively. This inner product reveals distortion as a joint event: neither factor alone is sufficient; distortion arises from their interaction.

A calibrated correction. The distortion bound in equation 15 motivates a correction beyond the isotropic baseline, to better align the scheduling objective with the actual generation quality. Because directional alignment is unknown, the offline signal $\kappa(\sigma)$ serves as a measure of the *potential* local distortion. Rather than explicitly modeling this unobservable alignment, we retain its σ-dependent variation through $\kappa(\sigma)$ and use a single calibration parameter α to control the overall correction strength:

$$g_\alpha(\sigma) = \bar{g}(\sigma) \cdot \frac{\kappa(\sigma)^\alpha}{\mathrm{mean}_\sigma[\kappa(\sigma)^\alpha]}, \qquad \kappa(\sigma) = \frac{\lambda_{\max}(J_{\mathrm{sym}})}{\bar{g}(\sigma)} \tag{16}$$

Thus, α calibrates the overall strength of the correction, while $\kappa(\sigma)$ determines its variation across noise levels. Appendix D details the construction of $\kappa(\sigma)$ and its connection to equation 15.

Theoretical α-scaling with NFE. The propagator mismatch between the scalar proxy equation 14 and the true propagator equation 10 accumulates multiplicatively along the trajectory. Taking logarithms yields the additive decomposition

$$\log \frac{\Gamma_j^{\mathrm{true}}}{\Gamma_j^{\mathrm{proxy}}} = \sum_{k=j+1}^{N} \log \frac{1 + h_k Q_k}{1 + h_k \bar{g}_k} \approx \sum_{k=j+1}^{N} h_k \bar{g}_k \, (\kappa_{\mathrm{dir},k} - 1), \qquad \kappa_{\mathrm{dir}}(\sigma) := \frac{Q(\hat{\eta}; \sigma)}{\bar{g}(\sigma)} \in [\kappa_{\min}, \kappa], \tag{17}$$

Algorithm 1 EP-Flow: Offline Schedule Optimization

Require: Pretrained model $\boldsymbol{v}_\theta$, endpoints $\sigma_0 > \sigma_N$, NFE N, held-out set $\mathcal{D}$, restarts R
Ensure: Optimized schedule $\{\sigma_i^\star\}_{i=0}^N$
// Phase 1: offline, once per model
Estimate $\hat{\sigma}_\eta(\sigma), \bar{g}(\sigma), \kappa(\sigma), \hat{D}(\sigma)$ on $\mathcal{D}$
Fit PCHIP interpolants; compute $\hat{\rho}_\infty, \hat{C}_\rho$
// Phase 2: once per NFE budget
Initialize $\mathbf{u}^{(0)}$ via warm-start; parameterize $\mathbf{s} = \mathrm{softmax}(\mathbf{u})$
for $r = 1, \ldots, R$ **do**
    Evaluate $\mathcal{C}(\mathbf{s}(\mathbf{u}))$; update $\mathbf{u}$ via Adam GD
end for
return schedule $\{\sigma_i^\star\}$ for $\mathbf{u}^\star = \arg\min_r \mathcal{C}_r$

a schedule-dependent quadrature of the total mismatch accumulated mismatch, valid to first order when $h_k \bar{g}_k \ll 1$. The correction equation 16 offsets this mismatch with per-α capacity $J = \int_{\sigma_N}^{\sigma_0} \bar{g}(\sigma) \log \kappa(\sigma)\, d\sigma$; balancing capacity against the residual left unresolved at finite step count gives $\alpha^\star J = R(\mathrm{NFE})$.

Since J is independent of NFE, the NFE dependence enters entirely through R. For the first-order scheme used throughout, $R = O(\bar{h})$ with $\bar{h} = (\sigma_0 - \sigma_N)/\mathrm{NFE}$, so that

$$\alpha^\star(\mathrm{NFE}) \approx \frac{C}{\mathrm{NFE}}, \qquad C = \bar{\tau}_w C', \tag{18}$$

where the functional form is derived and C factorizes into an NFE-independent alignment factor $\bar{\tau}_w \in [0, 1]$ and an $O(1)$ integrator constant C' (Appendix D).

4.3 Schedule Optimization

The stationarity conditions $\partial \mathcal{C}/\partial t_k = 0$ are coupled across nodes through the propagators $\{\Gamma_j\}$. In the degenerate limit $\Gamma_j \equiv I$, they reduce to an equidistribution principle—recovering uniform $\Delta\sigma$ for Euler and uniform $\Delta\lambda$ for DPM-Solver (Appendix F).

All required profiles are single-σ expectations estimated once per model without retraining, with statistic collection taking approximately 3 hours on a single A6000. The score-error variance $\sigma_\eta^2(\sigma) = \mathbb{E}_{p_{\mathrm{train}}}[\|\boldsymbol{\eta}(\boldsymbol{x}_\sigma, \sigma)\|^2]$ from the training-residual MSE, and the curvature D_j via finite differences in σ; and $\bar{g}$ equation 14 along with κ via lightweight Jacobian-power iterations. These profiles render $\mathcal{C}$ a scalar functional of N nodes. Requiring no network evaluations, the optimization runs in minutes per NFE budget (Algorithm 1) to yield a fixed schedule with zero inference overhead.

5 Experiments

Models and baselines. We evaluate EP-Flow across latent- and pixel-space settings: FLUX.1-dev (Labs, 2024) (Euler Solver in σ-space), Stable Diffusion v1.5 (Rombach et al., 2022) (DPM-Solver-2 in λ-space (Lu et al., 2022)), and EDM/EDM2 (Karras et al., 2022; 2024) (using `edm2-img512-s-dino` with FD-DINOv2). All methods share identical solver, endpoints, and NFE, differing only in the interior timestep allocation (Appendix G). Initializing from each baseline schedule, EP-Flow ensures reported gains reflect the error-propagation objective.

Datasets and evaluation For text-to-image generation, we evaluate fine-grained semantic alignment on the GenEval benchmark (Ghosh et al., 2023)(553 prompts across 4 seeds), and human preference alignment on HPSv2 (Wu et al., 2023) (4 styles $\times$ 800 prompts). For unconditional and class-conditional generation, we report FID and FD-DINOv2 on CIFAR-10 (Krizhevsky, 2009), FFHQ (Kazemi & Sullivan, 2014), AFHQ (Choi et al., 2020), and ImageNet (Russakovsky et al., 2015) over 50k generated samples (see Appendix J for qualitative samples).

Empirical probing of model statistics. Uniform-t and logSNR-uniform spacings, which our framework recovers as special cases, rely on structural assumptions that are visibly violated by modern

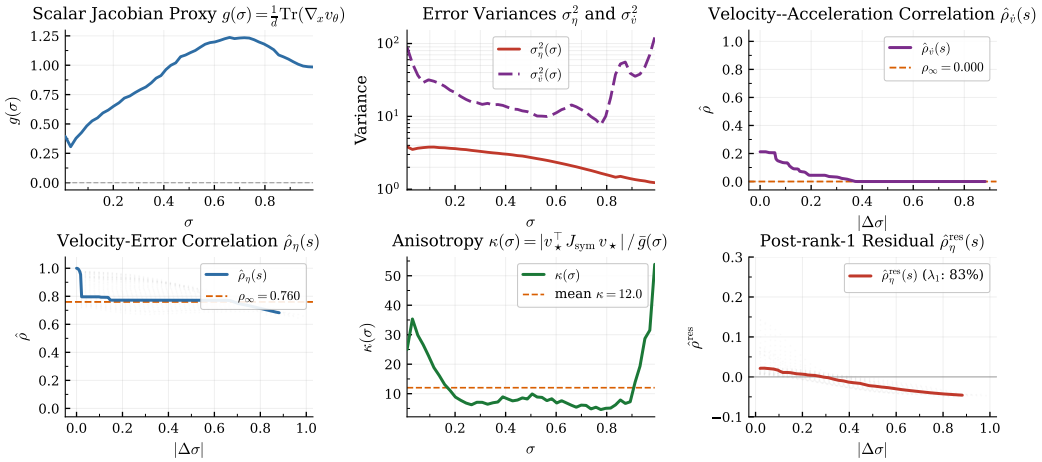


Figure 1: Empirical statistics of FLUX.1-dev (Labs, 2024) supporting the assumptions in §4: far-field structure and anisotropy, with the latter motivating the α-correction.

pretrained models (Fig. 1). *Temporally*, the velocity-error kernel on FLUX.1-dev (Labs, 2024) displays a stable far-field plateau with a decaying banded residual, whose finiteness $C_\rho < \infty$ is precisely what our error bounds require; tied to the decomposition in equation 11, this long-range behavior is not specific to latent models—we observe the same pattern in pixel-space models such as EDM and EDM2 (Appendix C). *Spatially*, the observed anisotropy ($\bar\kappa$) motivates refining the isotropic surrogate toward an objective that more faithfully captures anisotropic effects on generation quality.

5.1 Improving Heuristic Schedules via Error Propagation

Throughout this section we use the isotropic proxy ($\alpha = 0$). We demonstrate the efficacy of our framework in three steps: starting from the special case of our framework, expanding to strong hand-crafted heuristics, and concluding with a stress test under non-ideal conditions.

Recovering and surpassing exact special cases. Uniform-t and logSNR spacing are principled, widely adopted choices that serve as the exact degenerate limits of our objective under a stationary regime (App. F). Initializing from these special cases allows us to precisely isolate the impact of accounting for error propagation. This modeling yields consistent gains on GenEval and HPSv2 1, confirming that it captures real error dynamics not reflected by uniform allocation.

Generalization across models and recipes. Extending our analysis to strong hand-crafted heuristics, optimizing our objective starting from the Karras schedule on EDM (Karras et al., 2022) induces non-trivial shifts in the resulting schedule, outperforming the unoptimized baseline across diverse datasets and resolutions. Crucially, these gains carry over to EDM2 (Karras et al., 2024) despite a substantially different training recipe. This cross-setup consistency confirm that the gains stem from our error-aware objective rather than overfitting to a specific setting.

Robustness beyond stationarity. SD v1.5 (Rombach et al., 2022) does not exhibit the globally stationary structure assumed by our analysis, providing a natural stress test of the framework. In this setting, a piecewise-kernel formulation (Appendix C.2) localizes the estimate where stationarity breaks down, allowing the objective to improve upon the logSNR schedule when initialized from it. This result supports the robustness of our framework to deviations from the stationary setting.

5.2 Refining Competitive Schedules

Table 1 shows that when schedules leave significant room for optimization, the isotropic proxy ($\alpha = 0$) already recovers much of this headroom, yielding consistent gains over baseline. The scope extends to strong learned ones, such as the optimization-based CTR (Xue et al.,

Table 2: Refining Xue et al. (2024).

Method / NFE	4	6	8
GenEval ↑			
CTR (Xue et al., 2024)	0.274	0.345	0.365
/ Ours (α=0)	**0.329**	**0.377**	**0.401**
HPSv2 ↑			
CTR (Xue et al., 2024)	18.28	21.14	22.98
/ Ours (α=0)	**19.32**	**22.51**	**23.56**

EDM
FID ↓

Dataset	NFE	Karras	Ours
CIFAR-10 32×32	6	36.97	**28.02**
	10	15.57	**13.48**
	14	9.65	**8.76**
	18	7.03	6.56
ImageNet 64×64	6	41.09	**31.71**
	10	17.26	**15.24**
	14	10.24	**9.56**
	18	7.30	7.06
AFHQ 64×64	6	22.90	**17.21**
	10	10.13	**9.34**
	14	6.62	6.28
	18	5.09	4.90
FFHQ 64×64	6	43.64	**33.41**
	10	20.92	**17.77**
	14	13.26	**12.39**
	18	9.74	9.86

SD v1.5
GenEval Overall ↑

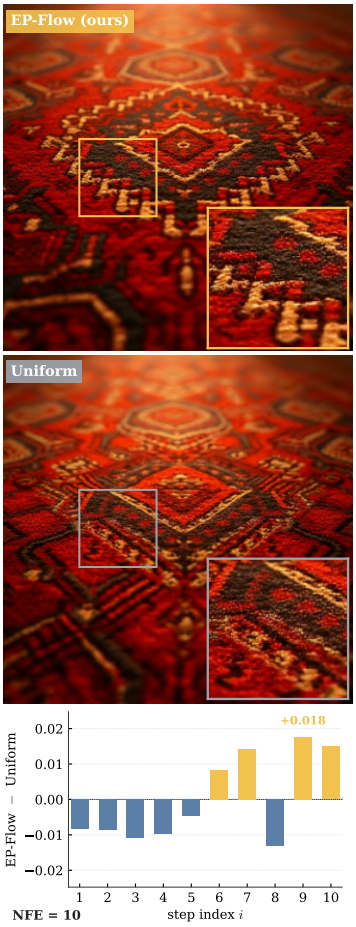
NFE	logSNR	Ours
4	0.261 19	**0.30681**
6	0.355 79	**0.36729**
8	0.378 93	**0.39004**
10	0.398 73	**0.40665**

EDM2
FD-DINOv2 ↓

NFE	Karras	Ours
5	614.35	**509.59**
10	166.76	**148.68**
15	111.86	**104.98**

FLUX-1.dev
GenEval Overall ↑

NFE	Uniform	Ours
4	0.332 45	**0.36799**
6	0.471 61	**0.49479**
8	0.549 59	**0.55880**
10	0.581 29	0.581 37



HPSv2 Human Preference Score ↑

	Baseline				Ours			
Model	**4**	**6**	**8**	**10**	**4**	**6**	**8**	**10**
SD v1.5	17.14	20.74	22.36	23.27	**17.99**	**20.97**	**23.20**	**23.86**
FLUX	20.48	25.56	28.38	29.27	**21.55**	**26.22**	28.59	29.53

Table 1: Comparison of timestep schedules across different generative models and sampling budgets. Lower is better for FID and FD-DINOv2, while higher is better for GenEval.

2024)[2] (Table 2) and the theoretically derived AYS (Sabour et al., 2024) (Table 4). The isotropic proxy suffices in general, but for schedules with limited remaining optimization headroom, such as the empirically validated cosine schedule for FLUX.1-dev (App. I), a finer signal is needed. In those settings, correcting for anisotropy (§4) reshapes the cost landscape and restores an optimization signal better aligned with generation quality. Below, we verify that the anisotropic correction behaves as the theory predicts across two regimes.

Low NFE — the correction follows the predicted scaling. At low NFE, large steps amplify anisotropic distortion, and our theory predicts an approximately $\propto 1/\text{NFE}$ scaling of the optimal correction strength in a near-optimal regime. Calibrating α on GenEval Overall yields values broadly consistent with this trend (Tab. 16), while applying the same α to PickScore without re-tuning preserves baseline perceptual quality (Tab. 3). Tab. 17 further shows that the calibrated α is statistically indistinguishable from the empirical PickScore optimum.

High NFE — diminishing returns. As NFE increases, shrinking headroom yields diminishing returns for schedule optimization. On FLUX.1-dev (NFE=20), optimizing from the cosine schedule matches the baseline on GenEval (0.6517 for ours at α=0 vs. 0.6507 for cosine) with a negligible PickScore difference ($+0.0014$, 95% CI $[-0.0135, +0.0164]$). Thus, as optimization

Table 3: Best α_{GenEval} transfers to PickScore without re-tuning (*: $p < 0.05$; n.s.: not significant).

NFE	Pred α^*	PickScore Δ vs. cosine
4	0.50	$+0.0316$ ([+0.0184, +0.0452]) *
6	0.33	-0.0284 ([-0.0646, +0.0075]) n.s.
8	0.25	$+0.0074$ ([-0.0098, +0.0252]) n.s.

[2]We refer to it by its mechanism, constrained trust region (CTR).

Table 4: Comparison with Align Your Steps (AYS) on Stable Diffusion v1.5. GenEval Overall and HPSv2 are reported at different NFE budgets. Higher is better for all metrics.

	NFE = 10		NFE = 15		NFE = 20	
Method	GenEval ↑	HPSv2 ↑	GenEval ↑	HPSv2 ↑	GenEval ↑	HPSv2 ↑
AYS	0.40829	24.00	0.41890	24.54	0.42975	24.72
Ours	**0.41477**	**24.22**	**0.42423**	**24.61**	0.42776	24.73

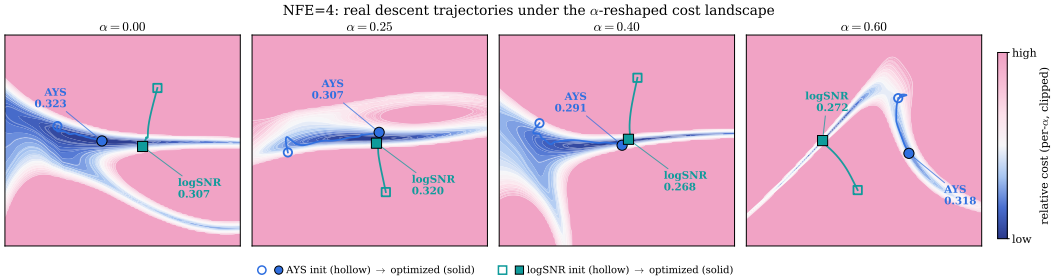


Figure 2: **Optimization landscape at** NFE=4 **under varying** α**.** 2D PCA projections of descent trajectories initialized from AYS (blue) and logSNR (orange). Hollow and solid markers denote initialization and convergence; annotations report GenEval-Overall.

headroom vanishes, further refinement produces negligible performance changes without degrading perceptual quality.

The anisotropic correction reflects the same shrinking headroom: the scaling implies $\alpha^\star \to 0$ as NFE grows. Probing this regime, we find that varying α does not yield a uniformly better schedule, but instead trades off competing metrics: α=0 performs best on Texture, while α=0.75 performs best on Numeracy (full per-category breakdown in Tab. 14). Optimization thus navigates a family of quality profiles rather than a single optimum.

5.3 Understanding Optimization Behavior

Preceding sections demonstrate that our method improves heuristic schedules and refines existing ones. We further investigate the underlying optimization behavior.

Analysis of learned schedules. We next inspect the learned schedules to understand how optimization modifies the initial discretization. At NFE=6, starting from the Karras schedule with step sizes $[49.82, 20.46, 7.21, 2.05, 0.42, 0.05]$, optimization yields $[47.20, 16.65, 14.26, 1.65, 0.22, 0.03]$. This shows that cost-driven optimization induces non-trivial step-size redistribution that substantially alters generation quality.

Visualization of optimization landscape. Figure 2 visualizes our optimization objective and descent trajectories at NFE=4, using 2D PCA projections of the three-dimensional schedule space, with AYS and logSNR as initializations. Two patterns emerge. *(i) The objective is strongly non-convex.* Different initializations converge to distinct basins with varying downstream quality, highlighting the role of initialization. *(ii)* α *reshapes the landscape.* At α=0.40, both initializations converge to a shared basin, whereas at α=0.60, they settle in well-separated basins with different quality. These observations motivate the multi-restart procedure in Algorithm 1.

6 Conclusion

In this paper, we introduce a new framework for timestep schedule optimization in diffusion and flow-based models. We believe that EP-Flow shifts from a myopic, per-interval viewpoint to a global perspective that explicitly accounts for error propagation. Empirically, we observe that it consistently improves upon heuristic schedules and can further refine learned schedules across EDM/EDM2, SD v1.5, and FLUX.1-dev. Limitations and further discussions are detailed in Appendix K.

AI USE STATEMENT

In this work, we did not use generative AI tools for tasks requiring disclosure under the ICLR AI Policy for Authors (generating synthetic datasets; developing theoretical models or conceptual frameworks; formulating mathematical claims assisting in the writing of proofs; proposing or refining hypotheses; designing or research methodology or experiments; implementing methods; assisting with translation; cleaning and reformatting datasets; supporting qualitative or thematic data analysis; and interpreting results); these tasks are not applicable to this work. Additionally, we used generative AI tools for two tasks under the recommended-disclosure category: (i) polishing the language of the manuscript to improve readability, and (ii) assisting with the styling of scientific figures. We have reviewed all AI-assisted work: all language edits were checked by the authors against the intended technical meaning, and all AI-assisted figures were verified against the underlying experimental data before inclusion. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

We confirm that we have read and adhere to the conference Code of Ethics. To the best of our knowledge, this work is original and all relevant prior work has been properly cited. We propose timestep scheduling methods for pretrained diffusion and flow-matching models to improve sampling efficiency. Our methods operate exclusively at inference time, without modifying model weights or introducing new training datasets. Consequently, any potential biases or visual artifacts are inherited directly from the underlying pretrained models. All evaluations are conducted on standard, publicly available research benchmarks without collecting personal or sensitive data. While reducing computational barriers to generative modeling can broaden access, it may also lower the resource requirements for misuse. We strongly encourage responsible deployment and domain-specific safety evaluations prior to real-world application.

REPRODUCIBILITY STATEMENT

To ensure full reproducibility, Appendix G provides pseudocode for our optimization algorithm alongside complete theoretical proofs, detailed experimental configurations, and implementation details for all baselines. As our framework optimizes sampling schedules without altering model parameters, we report the exact optimized schedules in Appendix H and will release our codebase, allowing for direct and efficient reproduction of reported results.

REFERENCES

Divya Jyoti Bajpai, Dhruv Bhardwaj, Soumya Roy, Tejas Duseja, Harsh Agarwal, Aashay Sandansing, and Manjesh Kumar Hanawal. Fastflow: Accelerating the generative flow matching models with bandit inference. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 131476–131494, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/d58e4a96eebe794c6ab8a96893013dee-Paper-Conference.pdf.

Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

Yansong Gao, Zhihong Pan, Xin Zhou, Le Kang, and Pratik Chaudhari. Fast diffusion probabilistic model sampling through the lens of backward error analysis. *arXiv preprint arXiv:2304.11446*, 2023.

Dhruba Ghosh, Hannaneh Hajishirzi, and Ludwig Schmidt. Geneval: An object-focused framework for evaluating text-to-image alignment. *Advances in Neural Information Processing Systems*, 36: 52132–52152, 2023.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022.

Tero Karras, Miika Aittala, Jaakko Lehtinen, Janne Hellsten, Timo Aila, and Samuli Laine. Analyzing and improving the training dynamics of diffusion models. In *Proc. CVPR*, 2024.

Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021.

Alex Krizhevsky. Learning multiple layers of features from tiny images. 2009. URL `https://api.semanticscholar.org/CorpusID:18268744`.

Black Forest Labs. Flux. `https://github.com/black-forest-labs/flux`, 2024.

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

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

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. *Advances in neural information processing systems*, 35:5775–5787, 2022.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, June 2022.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.

Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your steps: Optimizing sampling schedules in diffusion models. *arXiv preprint arXiv:2404.14507*, 2024.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *arXiv preprint arXiv:2306.09341*, 2023.

Shuchen Xue, Zhaoqiang Liu, Fei Chen, Shifeng Zhang, Tianyang Hu, Enze Xie, and Zhenguo Li. Accelerating diffusion sampling with optimized time steps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8292–8301, 2024.

Liangyu Yuan, Ruoyu Wang, Tong Zhao, Dingwen Fu, Mingkun Lei, Beier Zhu, and Chi Zhang. Few-step diffusion sampling through instance-aware discretizations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 35882–35892, June 2026.

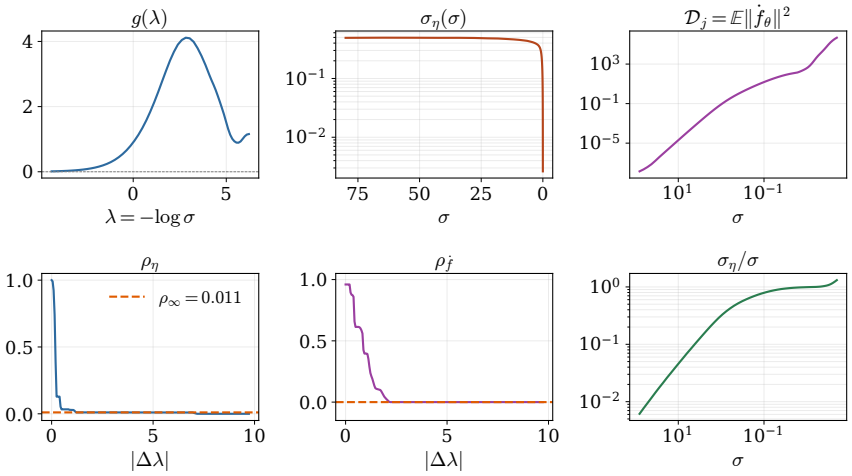


Figure 3: Offline statistics collected from EDM2 (Karras et al., 2024)

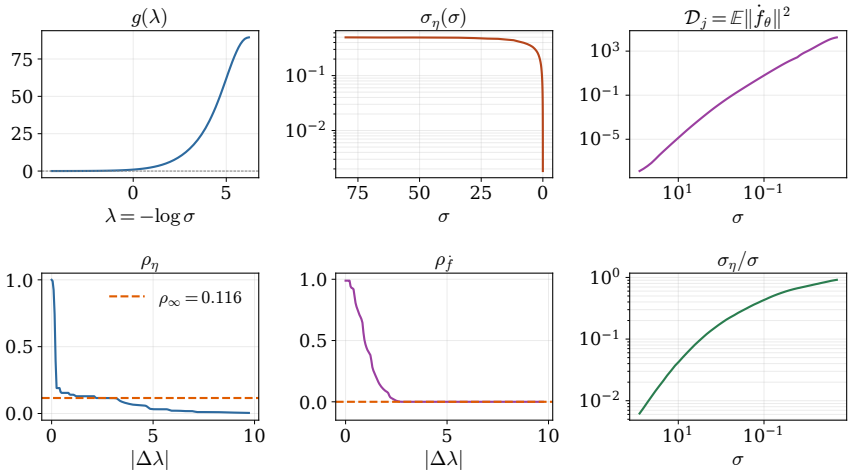


Figure 4: Offline statistics collected from EDM-cifar10 32×32 (Karras et al., 2022)

Summary of Appendix

01 **Solver Instantiation** Full derivation of the terminal-error objective and its decomposition into systematic bias, residual variance, and discretization error under DPM-Solver-2 setting.

02 **Propagation structure** Cross-step covariance, banded characterization, and the resulting local per-step objective.

03 **Anisotropy correction** Directional variation of the error field and the corresponding schedule correction.

04 **Degenerate limits** Recovery of logSNR-uniform spacing for semilinear diffusion and uniform time spacing for flow matching.

A Offline Statistics

To examine whether the far-field plateau is model-specific, we additionally collect offline statistics for EDM/EDM2, as shown in Figs. 4 and 3. Notably, the measured C_ρ is close to 1 for both EDM and EDM2, which is substantially larger than the $C_\rho \approx 7 \times 10^{-4}$ observed for FLUX.1-dev. The improvements demonstrated in Table. 1 further highlight that our approach is agnostic to the exact magnitude of C_ρ.

B Leading Order Analysis

We first write the feasible set

$$\Delta_T := \Big\{\mathbf{h} \in \mathbb{R}^N : h_i \geq 0,\ \textstyle\sum_{i=1}^N h_i = T\Big\}, \qquad \mathbf{1}^\perp := \{\mathbf{v} : \mathbf{1}^\top \mathbf{v} = 0\}, \qquad P := I - \tfrac{1}{N}\mathbf{1}\mathbf{1}^\top,$$

so that all feasible variations $\delta\mathbf{h}$ lie in the tangent space $\mathbf{1}^\perp$ and P is the orthogonal projector onto it. Let e_N be the terminal sampling error and $\xi = \eta + \text{disc}$ the total local forcing. The accumulated propagator gain is $\Gamma_j := \prod_{k>j}(1+\epsilon_k)$. Under the decomposition $\mathcal{K}_\bullet(\mu,\mu') = \rho_\infty\, \sigma_\bullet(\mu)\sigma_\bullet(\mu') + \mathcal{K}_\bullet^{\text{res}}(\mu,\mu')$, the leading-order cost $\mathcal{C}(\mathbf{h}) := \mathbb{E}\|e_N^{(1)}\|^2$ splits into a scalar *coherent floor* plus a *residual* part:

$$\boxed{\mathcal{C}(\mathbf{h}) \;=\; \underbrace{\rho_\infty\, \Phi(\mathbf{h})^2}_{\text{coherent}} \;+\; \underbrace{\mathcal{C}_{\text{res}}(\mathbf{h})}_{\text{residual+disc}}}, \qquad \Phi(\mathbf{h}) = \boldsymbol{\Gamma}^\top \boldsymbol{\sigma} = \sum_{j=1}^N \Gamma_j A_j, \tag{19}$$

with $A_j = \int_{t_{j-1}}^{t_j} \sigma_\eta(\mu)\, d\mu$ and $\mathcal{C}_{\text{res}}(\mathbf{h}) = (1+2C_\rho)\sum_j \Gamma_j^2 \big(V_j^{\text{res}} + D_j^{\text{disc}}\big)$, where $D_j^{\text{disc}} = \frac{1}{4}\,\mathbb{E}\|\dot f_\theta\|^2\, h_j^4$.

Assumptions.

(H1) Bounded path Hessian. There exists a constant $\bar L_H < \infty$ such that, within an $\mathcal{O}(\xi_0)$-tube of the exact trajectory $x^\star(\lambda)$,

$$\left\| \int_0^1 (1-s)\, \nabla_x^2 v_\theta\big(x^\star(\lambda) + s e(\lambda), \lambda\big)\, \mathrm{d}s \right\| \leq \bar L_H, \qquad \forall\, \lambda \in [0,T].$$

(H2) Bounded second-moment forcing. Let $\xi = \eta + \text{disc}$ denote the total forcing, consisting of the vector-field approximation and discretization errors. Assume

$$\xi_0 := \sup_\lambda \big(\mathbb{E}\|\xi(\lambda)\|^2\big)^{1/2} < \infty.$$

Theorem 1 (Leading-Order Sufficiency)**.** *Let* $h^\star \in \arg\min_{\Delta_T} \mathcal{C}$ *and* $h^{\star\star} \in \arg\min_{\Delta_T} \mathbb{E}\|e_N\|^2$. *Under* (H1)–(H2)*:*

(i) $\mathbb{E}\|e_N\|^2 = \mathcal{C}(\mathbf{h}) + \mathcal{O}(\bar L_H \xi_0^3)$ *uniformly on* Δ_T*;*

(ii) $\mathbb{E}\|e_N(h^\star)\|^2 - \mathbb{E}\|e_N(h^{\star\star})\|^2 = \mathcal{O}(\bar L_H \xi_0^3)$.

Leading-order Approximation

The proof establishes two properties:

(i) **leading-order accuracy:** $\mathcal{C} = \text{MSE} + \mathcal{O}(\bar L_H \xi_0^3)$;

(ii) **suboptimality:** scheduling with $h^\star$ attains the true MSE optimum up to $\mathcal{O}(\bar L_H \xi_0^3)$.

Proof. Linearization via a first-order remainder.

Denote the exact-point Jacobian $\mathbf{J}_i^* := \nabla_x v_\theta(x_{i-1}^*, t_{i-1})$. Applying first-order Taylor with integral remainder to $\bar{\mathbf{J}}_{t_{i-1}} = \int_0^1 \nabla_x v_\theta\big(\tilde x_{i-1} + \tau(x_{i-1}^* - \tilde x_{i-1}), t_{i-1}\big)\, d\tau$ about x_{i-1}^* gives

$$\bar{\mathbf{J}}_{t_{i-1}} = \mathbf{J}_i^* + \mathbf{R}_i, \qquad \mathbf{R}_i := \int_0^1 (1-s)\, \nabla_x^2 v_\theta\big(x_{i-1}^* - s\, e_{i-1}, t_{i-1}\big)[e_{i-1}, \cdot]\, ds.$$

Under (H1), $\|\mathbf{R}_i\| \leq \bar L_H \|e_{i-1}\| = \mathcal{O}(\xi_0)$

Two-term error split. The one-step recursion $e_i = (I + h_i \bar{\mathbf{J}}_{t_{i-1}})e_{i-1} + f_i$ separates into an exact-trajectory propagator and a remainder,

$$e_i = P_i^{(0)} e_{i-1} + \delta P_i\, e_{i-1} + f_i, \qquad P_i^{(0)} := I + h_i \mathbf{J}_i^*, \quad \delta P_i := h_i \mathbf{R}_i, \quad \|\delta P_i\| \leq h_i \bar L_H \|e_{i-1}\|.$$

Writing $e = e^{(1)} + \tilde{e}$, the linearized error solves $e_i^{(1)} = P_i^{(0)} e_{i-1}^{(1)} + f_i$, i.e.

$$e_N^{(1)} = \sum_{j=1}^{N} \Gamma_j f_j, \qquad \|e_N^{(1)}\| = \mathcal{O}(\xi_0),$$

and the remainder collects the discarded term,

$$\tilde{e}_N = \sum_{i=1}^{N} \Pi_{i\to N}^{(0)} \, \delta P_i \, e_{i-1}, \qquad \|\tilde{e}_N\| \le \sum_i \Pi_{i\to N}^{(0)} \, h_i \bar{L}_H \|e_{i-1}\|^2 = \mathcal{O}(\bar{L}_H \xi_0^2),$$

where the per-step factors h_i sum to $\sum_i h_i = T = \mathcal{O}(1)$, so no additional $\bar{h}$ survives at the accumulated level.

Property (i). By Cauchy–Schwarz,

$$\delta\mathcal{C}(\mathbf{h}) := \mathbb{E}\|e_N\|^2 - \mathcal{C}(\mathbf{h}) = 2\,\mathbb{E}\big[e_N^{(1)\top} \tilde{e}_N\big] + \mathbb{E}\|\tilde{e}_N\|^2 = \mathcal{O}(\xi_0 \cdot \bar{L}_H \xi_0^2) + \mathcal{O}(\bar{L}_H^2 \xi_0^4) = \mathcal{O}(\bar{L}_H \xi_0^3),$$

uniformly on Δ_T.

Property (ii). With $M_\delta := \sup_{\Delta_T} |\delta\mathcal{C}| = \mathcal{O}(\bar{L}_H \xi_0^3)$ from (i),

$$\mathbb{E}\|e_N(h^\star)\|^2 - \mathbb{E}\|e_N(h^{\star\star})\|^2 = \underbrace{\big[\mathcal{C}(h^\star) - \mathcal{C}(h^{\star\star})\big]}_{\le 0 \ (h^\star \text{ min } \mathcal{C})} + \underbrace{\big[\delta\mathcal{C}(h^\star) - \delta\mathcal{C}(h^{\star\star})\big]}_{\le 2M_\delta} \ \le \ 2M_\delta, \tag{20}$$

Together, (i)–(iii) show that the surrogate $\mathcal{C}$ differs from the true MSE by $\mathcal{O}(\xi_0^3)$, while $\mathcal{C} = \mathcal{O}(\xi_0^2)$ at leading order. Thus, the surrogate captures the true MSE to leading order, with the neglected terms one order higher in ξ_0. □

C Temporal Structure

Setup. Fix $\bullet \in \{\eta, \mathrm{disc}\}$ and abbreviate $\mathcal{K}_{jk} := \mathcal{K}_\bullet(\mu_j, \mu_k)$, $\sigma_j^2 := \mathcal{K}_{jj} = \mathbb{E}\|\mathbf{f}_j\|^2$. The exact cost functional equation 59 splits into diagonal and off-diagonal parts:

$$\mathcal{C} = \sum_{j,k} \Gamma_j \Gamma_k \, \mathcal{K}_{jk} = \underbrace{\sum_j \Gamma_j^2 \sigma_j^2}_{\mathcal{C}^{\mathrm{diag}}} + \underbrace{\sum_{j\neq k} \Gamma_j \Gamma_k \, \mathcal{K}_{jk}}_{\mathcal{C}^{\mathrm{off}}}.$$

Introduce the weights $w_j := \Gamma_j \sigma_j$ and the normalized kernel $r_{jk} := \mathcal{K}_{jk}/(\sigma_j \sigma_k)$, i.e. $\rho_\bullet(\mu_j, \mu_k)$ in Section 4. By Cauchy–Schwarz, $|r_{jk}| \le 1$ and $r_{jj} = 1$. In matrix form, $\mathcal{C} = \mathbf{w}^\top R \, \mathbf{w}$ and $\mathcal{C}^{\mathrm{diag}} = \|\mathbf{w}\|^2$.

Proposition 1 (Exact sandwich bound)**.** *Let $\beta := \max_j \sum_{k\neq j} |r_{jk}|$. Then, without any structural assumption on the kernel,*

$$(1-\beta)\,\mathcal{C}^{\mathrm{diag}} \ \le \ \mathcal{C} \ \le \ (1+\beta)\,\mathcal{C}^{\mathrm{diag}}. \tag{21}$$

Proof. For $j \neq k$, AM–GM gives $|\Gamma_j \Gamma_k \mathcal{K}_{jk}| = |w_j||w_k||r_{jk}| \le \frac{1}{2}(w_j^2 + w_k^2)\,|r_{jk}|$. Summing over $j \neq k$ and using the symmetry $|r_{jk}| = |r_{kj}|$,

$$|\mathcal{C}^{\mathrm{off}}| \le \sum_j w_j^2 \sum_{k\neq j} |r_{jk}| \le \beta\, \mathcal{C}^{\mathrm{diag}}.$$

The claim follows from $\mathcal{C}^{\mathrm{diag}} - |\mathcal{C}^{\mathrm{off}}| \le \mathcal{C} \le \mathcal{C}^{\mathrm{diag}} + |\mathcal{C}^{\mathrm{off}}|$. □

Remarks. (i) The bound equation 21 holds for every kernel; the only data-dependent quantity is β, which we measure directly. The diagonal term is a $(1 \pm \beta)$-accurate proxy for $\mathcal{C}$.

(ii) A convenient sufficient condition for small β is banded decay, $|r_{jk}| \le \rho(|j-k|)$ with ρ non-increasing, which gives $\beta \le 2\sum_{d\ge 1} \rho(d)$. Note that this condition applies to the plateau-subtracted residual kernel r_{jk}^{res} introduced below (see Fig. 1).

Table 5: Eigen-spectrum of K_η directly computed on FLUX.1-dev ($N = 10$).

Matrix	$\lambda_1/\sum_i \lambda_i$	λ_2/λ_1	**Top-3**	**PR**
K_η (Toeplitz)	0.792	0.050	0.858	1.58
K_η (empirical)	0.865	0.052	0.936	1.33
Identity (uncorrelated)	0.100	1.000	0.300	10.0
Pure exp ($\rho_\infty = 0$)	0.815	0.333	0.815	9.85

Relation to the far-field plateau (A1). Under A1, $r_{jk} \to \rho_\infty$ in the long-range limit, yielding $\beta \approx (N-1)\rho_\infty$, which renders the bound trivial for the full kernel. We therefore remove the plateau at the level of the continuous kernel, where A1 is stated. Each per-step entry is an integral over the step intervals I_j (of length h_j):

$$\mathcal{K}_{jk} = \int_{I_j}\int_{I_k} \mathcal{K}_\bullet(\mu,\mu')\,d\mu\,d\mu'.$$

Inserting the exact split $\mathcal{K}_\bullet = \rho_\infty\,\sigma_\bullet(\mu)\sigma_\bullet(\mu') + \mathcal{K}_\bullet^{\mathrm{res}}$ gives

$$\mathcal{K}_{jk} = \rho_\infty\,a_j a_k + \mathcal{K}_{jk}^{\mathrm{res}}, \qquad a_j := \int_{I_j}\sigma_\bullet(\mu)\,d\mu, \qquad \mathcal{K}_{jk}^{\mathrm{res}} := \int_{I_j}\int_{I_k}\mathcal{K}_\bullet^{\mathrm{res}}(\mu,\mu')\,d\mu\,d\mu',$$

and hence, exactly,

$$\mathcal{C} = \rho_\infty\Big(\sum_j \Gamma_j a_j\Big)^2 + \mathcal{C}^{\mathrm{res}}, \qquad \mathcal{C}^{\mathrm{res}} := \sum_{j,k}\Gamma_j\Gamma_k\,\mathcal{K}_{jk}^{\mathrm{res}}.$$

Note that $\mathcal{K}_{jj} = \rho_\infty a_j^2 + \mathcal{K}_{jj}^{\mathrm{res}}$. Write $V_j^{\mathrm{res}} := \mathcal{K}_{jj}^{\mathrm{res}}$ and $r_{jk}^{\mathrm{res}} := \mathcal{K}_{jk}^{\mathrm{res}}/\sqrt{V_j^{\mathrm{res}}V_k^{\mathrm{res}}}$. Proposition 1 applied to $\mathcal{K}^{\mathrm{res}}$ gives

$$(1-\beta^{\mathrm{res}})\sum_j \Gamma_j^2 V_j^{\mathrm{res}} \;\le\; \mathcal{C}^{\mathrm{res}} \;\le\; (1+\beta^{\mathrm{res}})\sum_j \Gamma_j^2 V_j^{\mathrm{res}}, \qquad \beta^{\mathrm{res}} := \max_j \sum_{k\neq j}|r_{jk}^{\mathrm{res}}|.$$

By A1, $\mathcal{K}_\bullet^{\mathrm{res}}$ vanishes at long range, so β^{res} is controlled by the near-field residual alone.

C.1 Empirical Eigen-Spectrum of the Forcing Kernel

Does the forcing kernel K_η exhibit a single dominant spectral mode corresponding to the far-field plateau in 1?

We address this by diagonalizing the forcing kernel K_η for FLUX.1-dev at $N = 10$. Table 5 evaluates K_η by comparision with three reference structures: an uncorrelated identity kernel, a plateau-free exponential ($\rho_\infty = 0$), and a raw empirical estimate (for evaluating the Toeplitz approximation).

The results show that K_η is **sharply rank-one**: the leading eigenvalue accounts for 79% of the trace with $\mathrm{PR} = 1.58$, consistent with the dominant principal-component structure of the far-field plateau observed in Fig. 1.

C.2 Piecewise kernel

Stable Diffusion v1.5 does not exhibit a global far-field plateau, motivating us to characterize its piecewise non-stationary structure and assess the robustness of our framework beyond the global-plateau regime (as shown in Fig. 5).

The empirical correlation matrix $\rho(\lambda_i, \lambda_j)$ exhibits coarse block-diagonal structure across noise levels (Fig. 6, left). Fitting an optimal K-segment model to $\rho_\infty(\lambda)$ reveals that a coarse two-bin split into low- and high-noise regimes captures the majority of variance, with further splits yielding diminishing returns (Fig. 6, right).

D Spatial Structure

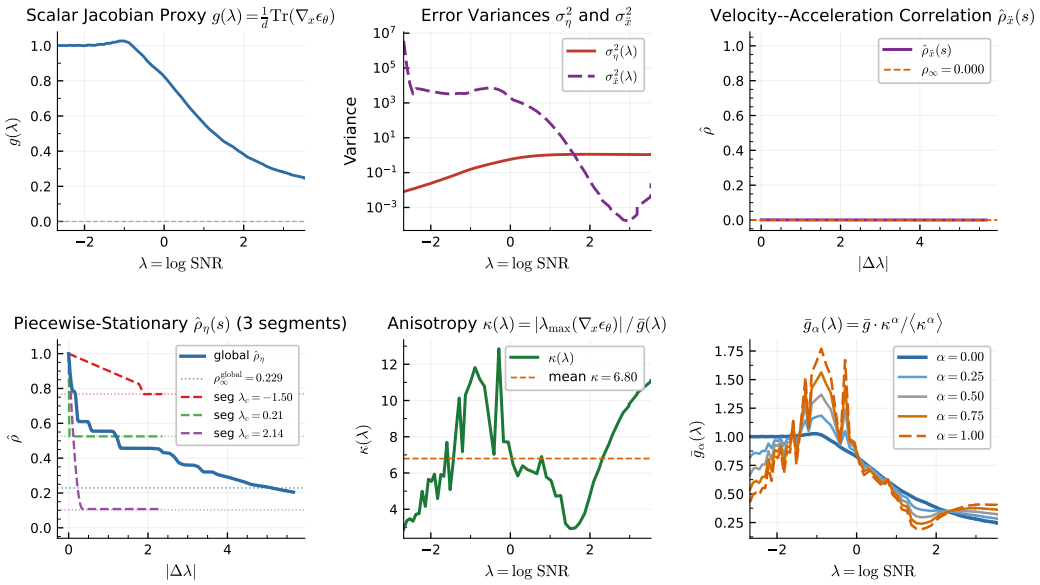


Figure 5: Offline statistics collected from Stable Diffusion v1.5 (Rombach et al., 2022)

SD v1.5, CFG=7.5

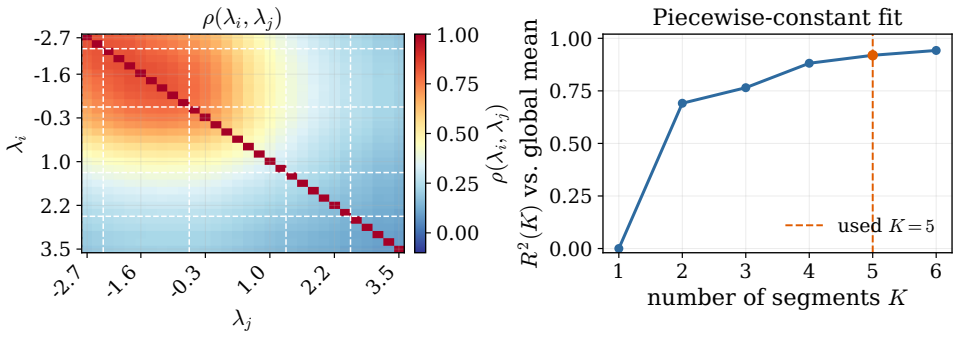


Figure 6: Empirical correlation matrix of SD v1.5

Takeaway. A large κ is necessary but insufficient for distortion; distortion requires both large κ and error alignment with high-eigenvalue directions.

D.1 DISTORTION

Core recurrence and directional amplification. Consider the error recurrence

$$e_i = P_i^* e_{i-1} + f_i. \tag{22}$$

the propagated error satisfies

$$\|P_i^* e_{i-1}\|^2 = e_{i-1}^\top (P_i^*)^\top P_i^* e_{i-1} = \|e_{i-1}\|^2 \left(1 + 2h_i Q_i(\hat{\eta}) + O(h_i^2)\right), \tag{23}$$

where

$$P_i^* = I + h_i J_i^*, \qquad J_i^* := J_{t_{i-1}}^*, \tag{24}$$

and

$$J_{\mathrm{sym},i}^* := \frac{J_i^* + J_i^{*\top}}{2}. \tag{25}$$

Here

$$\hat{\eta} := \frac{e_{i-1}}{\|e_{i-1}\|}, \qquad \|\hat{\eta}\| = 1, \tag{26}$$

is the direction of the *already accumulated error*, and

$$Q_i(\hat{\eta}) := \hat{\eta}^\top J_{\mathrm{sym},i}^* \hat{\eta} \tag{27}$$

is its Rayleigh quotient. Hence,

$$\lambda_{\min}(J_{\mathrm{sym},i}^*) \le Q_i(\hat{\eta}) \le \lambda_{\max}(J_{\mathrm{sym},i}^*). \tag{28}$$

Thus, $Q_i(\hat{\eta})$ characterizes the local first-order relative change of the accumulated error along its current direction, with its value determined by the alignment of $\hat{\eta}$ with the eigendirections of $J_{\mathrm{sym},i}^*$.

From Isotropic Averages to Directional Response. Any square matrix admits the unique decomposition

$$J = J_{\mathrm{sym}} + J_{\mathrm{skew}}, \qquad J_{\mathrm{sym}} = \frac{J + J^\top}{2}, \qquad J_{\mathrm{skew}} = \frac{J - J^\top}{2}. \tag{29}$$

Since J_{skew} is skew-symmetric, $\hat{\eta}^\top J_{\mathrm{skew}} \hat{\eta} = 0$, and thus

$$Q(\hat{\eta}) = \hat{\eta}^\top J_{\mathrm{sym}} \hat{\eta}. \tag{30}$$

Because J_{sym} is real symmetric, it admits the eigendecomposition

$$J_{\mathrm{sym}} = U \Lambda U^\top = \sum_{i=1}^{d} \lambda_i u_i u_i^\top, \tag{31}$$

giving

$$Q(\hat{\eta}) = \sum_{i=1}^{d} \lambda_i (\hat{\eta}^\top u_i)^2 = \sum_{i=1}^{d} \lambda_i c_i, \qquad c_i := (\hat{\eta}^\top u_i)^2, \quad c_i \ge 0, \quad \sum_i c_i = 1. \tag{32}$$

Hence, the Rayleigh quotient is a direction-dependent weighted average of the eigenvalues.

Isotropic Baseline and Directional Deviation. The isotropic baseline is given by the average Rayleigh quotient over all directions,

$$\bar{g} := \frac{1}{d} \operatorname{tr}(J_{\mathrm{sym}}) = \frac{1}{d} \sum_{i=1}^{d} \lambda_i, \tag{33}$$

where $\{\lambda_i\}_{i=1}^d$ are the eigenvalues of J_{sym}. Importantly, $\bar{g}$ also admits a natural probabilistic interpretation: if the error direction is uniformly distributed on the unit sphere,

$$\hat{\eta} \sim \mathrm{Unif}(S^{d-1}), \tag{34}$$

then

$$\mathbb{E}_{\hat{\eta}}\left[\hat{\eta}^\top J_{\mathrm{sym}}\hat{\eta}\right] = \frac{1}{d}\operatorname{tr}(J_{\mathrm{sym}}) = \bar{g}. \tag{35}$$

Thus, $\bar{g}$ is precisely the expected directional response under an isotropic error-direction assumption, while $Q_{\mathrm{real}} = \hat{\eta}^\top J_{\mathrm{sym}}\hat{\eta}$ measures the response along the actual accumulated error direction.

The resulting directional deviation is

$$\Delta := Q_{\mathrm{real}} - \bar{g} = \sum_{i=1}^{d}(\lambda_i - \bar{g})\left(c_i - \frac{1}{d}\right), \qquad c_i = (\hat{\eta}^\top u_i)^2. \tag{36}$$

Hence, spectral anisotropy alone does not cause the isotropic baseline to suffer from substantial distortion. A large κ must also align with the error in high-eigenvalue directions for this anisotropy to lead to actual distortion.

D.2 Theoretical α-scaling

Setup and notation. Write $\mathbb{E}_\sigma[\phi] := \frac{1}{\sigma_0 - \sigma_N}\int_{\sigma_N}^{\sigma_0}\phi(\sigma)\,d\sigma$ for the uniform average over the noise range. All σ-dependent spectral quantities $(\bar{g}, \lambda_{\min\,/\,\max}, Q, \kappa)$ are those of $J^*_{\mathrm{sym}}(\sigma)$ from the Distortion analysis.

Derivation of the α-scaling

We determine how the optimal correction strength $\alpha^\star$ scales with NFE through four steps:

(i) **bracket the directional factor**: bound the unobservable directional factor using a computable spectral proxy;

(ii) **compute the correction capacity**: derive the *exact* first-order response of propagation to an α-reweighting;

(iii) **compress the remaining uncertainty**: reduce all unobservability to a single NFE-independent scalar;

(iv) **match correction to residual**: match the correction capacity to the finite-NFE residual to obtain the leading-order scaling.

Although derived for the first-order Euler solver, this analysis extends to p-th order schemes, yielding $\alpha^\star \propto \mathrm{NFE}^{-p}$ due to their $O(\bar{h}^p)$ residual. We establish the scaling analytically and confirm it empirically (on Euler Solver, Tab. 16.

Step (i): a computable bracket for the unobservable directional factor

The distortion admits the exact inner-product form $\Delta = Q - \bar{g} = \langle \lambda - \bar{g}\mathbf{1},\, c - \frac{1}{d}\mathbf{1}\rangle$ with Popoviciu's inequality,

$$\frac{|\Delta|}{|\bar{g}|} \;\le\; \frac{\sqrt{d}\,\sigma_\lambda\,\sigma_c}{|\bar{g}|}, \qquad \sigma_\lambda \le \tfrac{1}{2}(\lambda_{\max} - \lambda_{\min}), \tag{37}$$

with $\sigma_\lambda^2 = \frac{1}{d}\sum_i(\lambda_i - \bar{g})^2$ and $\sigma_c^2 = \sum_i(c_i - \frac{1}{d})^2$. Define the true (unobservable) and proxy (computable) measure of the *potential* local distortion

$$\kappa_{\mathrm{dir}}(\sigma) := \frac{Q(\hat{\eta};\sigma)}{\bar{g}(\sigma)}, \qquad \kappa(\sigma) := \frac{\lambda_{\max}(\sigma)}{\bar{g}(\sigma)}; \tag{38}$$

the Rayleigh bounds give the *exact* bracket $\kappa_{\min} \le \kappa_{\mathrm{dir}} \le \kappa$. Only κ is available, with the unobserved gap $\kappa - \kappa_{\mathrm{dir}}$ absorbed into one scalar in Step (iii).

Step (ii): α-reweighting and its exact first-order capacity.

Reweight the propagator across timesteps through the proxy κ, controlled by a single exponent α:

$$g_\alpha(\sigma) := \bar{g}(\sigma)\,\frac{\kappa(\sigma)^\alpha}{\mathbb{E}_\sigma[\kappa^\alpha]}, \qquad g_0 = \bar{g}. \tag{39}$$

The normalization $\mathbb{E}_\sigma[\kappa^\alpha]$ gives the exact, *mean-zero* first-order perturbation:

$$\partial_\alpha \log g_\alpha\big|_{\alpha=0} = \log\kappa - \mathbb{E}_\sigma[\log\kappa], \qquad \mathbb{E}_\sigma\big[\partial_\alpha \log g_\alpha|_{\alpha=0}\big] = 0, \tag{40}$$

With the cumulative propagator $\Gamma_\alpha = \prod_k (1 + h_k g_\alpha(\sigma_k))$ and $\log(1+x) = x + O(x^2)$ in the small-step regime, the correction capacity is

$$J := \partial_\alpha \log \Gamma_\alpha\big|_0 = \int_{\sigma_N}^{\sigma_0} \bar{g}\,\big(\log\kappa - \mathbb{E}_\sigma[\log\kappa]\big)\,d\sigma = (\sigma_0 - \sigma_N)\,\mathrm{Cov}_\sigma\big(\bar{g}, \log\kappa\big). \tag{41}$$

The covariance form makes the leverage condition explicit: α has first-order effect *iff* the high-amplification steps (large $\bar{g}$) coincide with the anisotropic ones (large $\log\kappa$). If $\bar{g} \perp \log\kappa$ across σ, then $J = 0$ and no reweighting can help to first order. *Empirically, we verify that this covariance is indeed nonzero.*

Step (iii): unobservability absorbed into single scalar.

In the continuum limit, linearizing the cumulative true-vs-proxy propagation ratio for small h_k yields:

$$\begin{aligned} \log\frac{\Gamma^{\mathrm{true}}}{\Gamma^{\mathrm{proxy}}} = \sum_k \log\frac{1 + h_k\bar{g}_k\kappa_{\mathrm{dir},k}}{1 + h_k\bar{g}_k} &\overset{\log(1+x)=x+O(x^2)}{=} \sum_k h_k\bar{g}_k(\kappa_{\mathrm{dir},k} - 1) + O(h^2) \\ &\xrightarrow[\bar{h}\to 0]{} I_\tau := \int_{\sigma_N}^{\sigma_0} \bar{g}\,(\kappa_{\mathrm{dir}} - 1)\,d\sigma. \end{aligned} \tag{42}$$

Introduce the coefficient $\tau(\sigma) := \dfrac{\kappa_{\mathrm{dir}}(\sigma) - 1}{\kappa(\sigma) - 1}$, the weight $w := \bar{g}(\kappa - 1) \geq 0$, and $I := \int w\,d\sigma$.

$$I_\tau = \int w(\sigma)\tau(\sigma)\,d\sigma = \bar{\tau}_w\,I, \qquad \bar{\tau}_w := \frac{\int w\tau\,d\sigma}{\int w\,d\sigma}. \tag{43}$$

All residual unobservability is now the *single* weighted average $\bar{\tau}_w$, which depends only on the model's spectral geometry and error-direction distribution. By construction, $\bar{\tau}_w \in [0, 1]$.

Step (iv): finite-NFE residual and the $1/\mathrm{NFE}$ scaling.

The optimal α is determined by the residual mismatch R and the correction capacity J, giving $\alpha^\star = R/J$. We next characterize their NFE-dependence. The capacity J from equation 41 is a property of the continuous flow and is NFE-independent, whereas the residual R arises from finite-step discretization and therefore inherits the global truncation order of the first-order numerical scheme. We thus model

$$R(\mathrm{NFE}) = C'\,\bar{h}\,I_\tau + O(\bar{h}^2), \qquad C' = O(1), \tag{44}$$

so $R \to 0$ as $\mathrm{NFE} \to \infty$: with infinitely many steps the continuous flow is recovered and no correction is needed. Substituting $I_\tau = \bar{\tau}_w I$ and $\bar{h} = (\sigma_0 - \sigma_N)/\mathrm{NFE}$,

$$\boxed{\alpha^\star(\mathrm{NFE}) \;\approx\; \underbrace{C'\,\bar{\tau}_w}_{\substack{\text{model constant} \\ \in [0, C']}} \cdot \underbrace{\frac{I}{J}\,(\sigma_0 - \sigma_N)}_{\text{offline-computable}} \cdot \frac{1}{\mathrm{NFE}}} \tag{45}$$

Only $\bar{h}$ carries NFE, so the prediction is that $\alpha^\star\cdot\mathrm{NFE}$ is approximately constant—a leading-order law for the near-optimal regime. The prefactor splits into an offline spectral ratio I/J, computable from $\{\bar{g}(\sigma), \kappa(\sigma)\}$ alone, and the single bounded constant $C'\bar{\tau}_w$ absorbing the unobservable direction information. Back-solving from the observed optimum gives $C'\bar{\tau}_w \approx 0.469 \in [0, C']$, confirming $\alpha^\star$ is captured by a model-level constant within the admissible range.

E Derivation of DPM-Solver-2 Error Propagation

E.1 Setup and Notation

We analyze the error propagation of DPM-Solver-2 under a fixed step schedule (Lu et al., 2022). Let $\boldsymbol{x}^*_{t_i}$ denote the exact solution and $\tilde{\boldsymbol{x}}_{t_i}$ the numerical solution. Define the error at step i:

$$e_i := \boldsymbol{x}^*_{t_i} - \tilde{\boldsymbol{x}}_{t_i}$$

We treat the neural network's prediction error as a perturbation evaluated at exact trajectory points:

$$\eta(\boldsymbol{x}^*_t, t) := \epsilon_\theta(\boldsymbol{x}^*_t, t) - \epsilon^*(\boldsymbol{x}^*_t, t)$$

E.2 Deviation of Error Propagation

The exact solution satisfies:

$$\boldsymbol{x}^*_{t_i} = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}\boldsymbol{x}^*_{t_{i-1}} - \alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}\epsilon^*(\boldsymbol{x}^*_\lambda, \lambda)\, d\lambda$$

The algorithm calculation:

$$\tilde{\boldsymbol{x}}_{t_i} = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}\tilde{\boldsymbol{x}}_{t_{i-1}} - \sigma_{t_i}(e^{h_i}-1)\epsilon_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1}) - \frac{\sigma_{t_i}}{2r_1}(e^{h_i}-1)(\epsilon_\theta(\boldsymbol{u}_i, \boldsymbol{s}_i) - \epsilon_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1}))$$

Simultaneously add and subtract $\pm\,\sigma_{t_i}(e^{h_i}-1)\left[\left(1-\frac{1}{2r_1}\right)\epsilon^*(\boldsymbol{x}^*_{t_{i-1}}, t_{i-1}) + \frac{1}{2r_1}\epsilon^*(\boldsymbol{x}^*_{s_i}, s_i)\right]$ yields three terms:

Term (I) — Linear propagation :

$$\frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}e_{i-1}$$

Term (II) — Total score error:

$$\sigma_{t_i}(e^{h_i}-1)\left[\left(1-\frac{1}{2r_1}\right)\underbrace{(\epsilon_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1}) - \epsilon^*(\boldsymbol{x}^*_{t_{i-1}}, t_{i-1}))}_{\text{network error at previous step}} + \frac{1}{2r_1}\underbrace{(\epsilon_\theta(\boldsymbol{u}_i, s_i) - \epsilon^*(\boldsymbol{x}^*_{s_i}, s_i))}_{\text{network error at intermediate step}}\right]$$

Term (III) — Numerical discretization error:

$$\delta_{disc} = \sigma_{t_i}(e^{h_i}-1)\left[\left(1-\frac{1}{2r_1}\right)\epsilon^*(\boldsymbol{x}^*_{t_{i-1}}, t_{i-1}) + \frac{1}{2r_1}\epsilon^*(\boldsymbol{x}^*_{s_i}, s_i)\right] - \alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}\epsilon^*(\boldsymbol{x}^*_\lambda, \lambda)\, d\lambda$$

Define the score approximation errors at the two evaluation points:

$$\Delta_i^{(0)} = \epsilon_\theta(\tilde{x}_{t_{i-1}}, t_{i-1}) - \epsilon^*(x^*_{t_{i-1}}, t_{i-1}), \quad \Delta_i^{(1)} = \epsilon_\theta(u_i, s_i) - \epsilon^*(x^*_{s_i}, s_i)$$

Then **Term (II)** can be written compactly as $\sigma_{t_i}(e^{h_i}-1)\left[\left(1-\frac{1}{2r_1}\right)\Delta_i^{(0)} + \frac{1}{2r_1}\Delta_i^{(1)}\right]$. Add-and-Substract yields:

$$\Delta_i^{(0)} = \underbrace{\epsilon_\theta(\tilde{x}_{t_{i-1}}, t_{i-1}) - \epsilon_\theta(x^*_{t_{i-1}}, t_{i-1})}_{\text{(A) State error — Lipschitz in } e_{i-1}} + \underbrace{\epsilon_\theta(x^*_{t_{i-1}}, t_{i-1}) - \epsilon^*(x^*_{t_{i-1}}, t_{i-1})}_{\eta(x^*_{t_{i-1}}, t_{i-1}) \text{ — forcing at exact point}}$$

$$\Delta_i^{(1)} = \underbrace{\epsilon_\theta(\boldsymbol{u}_i, s_i) - \epsilon_\theta(\boldsymbol{x}^*_{s_i}, s_i)}_{\text{(C) State error — Lipschitz in } e_{i-1}} + \underbrace{\epsilon_\theta(\boldsymbol{x}^*_{s_i}, s_i) - \epsilon^*(\boldsymbol{x}^*_{s_i}, s_i)}_{\eta(x^*_{s_i}, s_i) \text{ — forcing at exact point}}$$

E.3 Intermediate-Point Deviation

The DPM-Solver-2 predictor is evaluated at the intermediate point

$$\boldsymbol{u}_i = \frac{\alpha_{s_i}}{\alpha_{t_{i-1}}}\tilde{\boldsymbol{x}}_{t_{i-1}} - \sigma_{s_i}(e^{r_1 h_i}-1)\,\epsilon_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1}).$$

Subtracting the exact solution $\boldsymbol{x}^*_{s_i}$ and performing the same add-and-subtract decomposition gives

$$\boldsymbol{u}_i - \boldsymbol{x}^*_{s_i} = -\frac{\alpha_{s_i}}{\alpha_{t_{i-1}}}e_{i-1} - \alpha_{s_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{s_i}} e^{-\lambda}\eta(\boldsymbol{x}^*_\lambda, \lambda)\, d\lambda - \delta^{\text{half}}_{\text{disc}} + \sigma_{s_i}(e^{r_1 h_i}-1)\,\mathbf{A}, \tag{46}$$

where $\delta^{\text{half}}_{\text{disc}}$ is the sub-step quadrature error on $[\lambda_{t_{i-1}}, \lambda_{s_i}]$:

$$\delta^{\text{half}}_{\text{disc}} := \sigma_{s_i}(e^{r_1 h_i}-1)\,\epsilon_\theta(\boldsymbol{x}^*_{t_{i-1}}, t_{i-1}) - \alpha_{s_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{s_i}} e^{-\lambda}\epsilon_\theta(\boldsymbol{x}^*_\lambda, \lambda)\, d\lambda,$$

and the mean-value Jacobian $\mathbf{A}$ captures the state-error contribution at the previous step:

$$\mathbf{A} := \int_0^1 \nabla_{\boldsymbol{x}}\epsilon_\theta(\boldsymbol{x}^*_{t_{i-1}} - \tau\, e_{i-1},\, t_{i-1})\, d\tau \cdot e_{i-1}.$$

E.4 Collecting Terms: Full Error Recursion

Substituting equation 46 into $\Delta_i^{(1)}$ via the mean-value theorem,

$$\Delta_i^{(1)} = \underbrace{\widetilde{J}_i(\boldsymbol{u}_i - \boldsymbol{x}^*_{s_i})}_{\text{(C)}} + \eta(\boldsymbol{x}^*_{s_i}, s_i), \qquad \widetilde{J}_i := \int_0^1 \nabla_{\boldsymbol{x}} \epsilon_\theta(\boldsymbol{x}^*_{s_i} + \tau(\boldsymbol{u}_i - \boldsymbol{x}^*_{s_i}),\, s_i)\, d\tau, \tag{47}$$

Assembling Terms (I)–(III), we group the $\mathcal{O}(e_{i-1})$ contributions and the remaining forcing separately.

Propagation coefficient. The terms linear in e_{i-1} collect as

$$C_{e_{i-1}} := \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}} - \frac{\widetilde{J}_i}{2r_1} \cdot \frac{\alpha_{s_i}}{\alpha_{t_{i-1}}}\, \sigma_{t_i}(e^{h_i} - 1). \tag{48}$$

The $\mathbf{A}$ terms (which are also linear in e_{i-1} through the mean-value Jacobian) contribute an additional piece:

$$\sigma_{t_i}(e^{h_i} - 1)\left[-\left(1 - \frac{1}{2r_1}\right) + \frac{\sigma_{s_i}(e^{r_1 h_i} - 1)}{2r_1}\widetilde{J}_i\right]\mathbf{A}.$$

Forcing terms. All terms independent of e_{i-1} collapse, after using the identity $\sigma_{t_i}(e^{h_i} - 1) = \alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda} d\lambda$, to the full-half structure:

$$\begin{aligned} f_i = {} & \underbrace{\alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}\eta(\boldsymbol{x}^*_\lambda, \lambda)\, d\lambda}_{\text{full-step } \eta \text{ integral}} - \frac{\widetilde{J}_i}{2r_1}\, \sigma_{t_i}(e^{h_i} - 1)\, \underbrace{\alpha_{s_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{s_i}} e^{-\lambda}\eta(\boldsymbol{x}^*_\lambda, \lambda)\, d\lambda}_{\text{half-step } \eta \text{ correction}} \\ & + \delta^\theta_{\text{disc}} - \frac{\widetilde{J}_i}{2r_1}\, \sigma_{t_i}(e^{h_i} - 1)\, \delta^{\text{half}}_{\text{disc}}, \end{aligned} \tag{49}$$

where the full-step discretization error is

$$\delta^\theta_{\text{disc}} := \alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}\left[\left(1 - \frac{1}{2r_1}\right)\epsilon_\theta(\boldsymbol{x}^*_{t_{i-1}}, t_{i-1}) + \frac{1}{2r_1}\epsilon_\theta(\boldsymbol{x}^*_{s_i}, s_i) - \epsilon_\theta(\boldsymbol{x}^*_\lambda, \lambda)\right] d\lambda.$$

Summary. The exact (to first order in e_{i-1}) step-wise recursion is therefore

$$\boxed{e_i = C_{e_{i-1}}\, e_{i-1} + \sigma_{t_i}(e^{h_i} - 1)\left[-(1 - \frac{1}{2r_1}) + \frac{\sigma_{s_i}(e^{r_1 h_i} - 1)}{2r_1}\widetilde{J}_i\right]\mathbf{A} + f_i,} \tag{50}$$

with $C_{e_{i-1}}$ given by equation 48 and f_i by equation 49.

E.5 Full-Step / Half-Step Correction Structure

Substituting equation 46 into $\Delta_i^{(1)}$ and assembling all three terms, the step error admits the following decomposition:

$$\begin{aligned} e_i = {} & \underbrace{\alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}\eta(\boldsymbol{x}^*_\lambda, \lambda)\, d\lambda}_{\text{full-step } \eta \text{ integral}} - \frac{\widetilde{J}_i}{2r_1}\, \sigma_{t_i}(e^{h_i} - 1)\, \underbrace{\alpha_{s_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{s_i}} e^{-\lambda}\eta(\boldsymbol{x}^*_\lambda, \lambda)\, d\lambda}_{\text{half-step } \eta \text{ correction}} \\ & + \underbrace{\delta^\theta_{\text{disc}}}_{\text{full-step disc. error}} - \frac{\widetilde{J}_i}{2r_1}\, \sigma_{t_i}(e^{h_i} - 1)\, \underbrace{\delta^{\text{half}}_{\text{disc}}}_{\text{half-step disc. correction}} \\ & + \underbrace{C_{e_{i-1}}\, e_{i-1} + \sigma_{t_i}(e^{h_i} - 1)\left[-(1 - \frac{1}{2r_1}) + \frac{\sigma_{s_i}(e^{r_1 h_i} - 1)}{2r_1}\widetilde{J}(\lambda_{s_i})\right]\mathbf{A}}_{\text{propagated previous-step error}}, \end{aligned} \tag{51}$$

where $C_{e_{i-1}} := \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}} - \frac{\widetilde{J}_i}{2r_1} \cdot \frac{\alpha_{s_i}}{\alpha_{t_{i-1}}} \sigma_{t_i}(e^{h_i} - 1), \quad \mathbf{A} := \int_0^1 \nabla_{\boldsymbol{x}} \epsilon_\theta(\boldsymbol{x}^*_{t_{i-1}} - \tau\, e_{i-1},\, t_{i-1})\, d\tau \cdot e_{i-1}.$

The structure in equation 51 reveals the corrector mechanism of DPM-Solver-2: the half-step evaluation at s_i enters only as a *correction* to the full-step integral, weighted by the intermediate Jacobian $\widetilde{J}_i/(2r_1)$. Both discretization terms $\delta^\theta_{\text{disc}}$ and $\delta^{\text{half}}_{\text{disc}}$ appear in the same paired fashion.

E.6 First-Order Linearization of the Propagator

To obtain a tractable recursion, we expand each mean-value Jacobian around the *exact* trajectory Jacobian. Define

$$J^*(\lambda) := \nabla_{\boldsymbol{x}} \epsilon_\theta(\boldsymbol{x}^*_\lambda, \lambda),$$

and write the mean-value form as

$$\widetilde{J}(\lambda)\, v = J^*(\lambda)\, v + \delta J(\lambda;\, v)\, v, \qquad \delta J(\lambda;\, v) := \int_0^1 (1-s)\, \nabla^2_{\boldsymbol{x}} \epsilon_\theta(\boldsymbol{x}^*_\lambda + sv,\, \lambda)\, [v,\, \cdot\,]\, ds,$$

where contracting one copy of v into the Hessian makes $\delta J(\lambda;\, v)$ a $d \times d$ matrix, of the same type as $J^*(\lambda)$. Assuming only that $\|\nabla^2_{\boldsymbol{x}} \epsilon_\theta(\boldsymbol{x}, \lambda)\|$ is bounded along the interpolation path $\{\boldsymbol{x}^*_\lambda + s\, v : s \in [0,1]\}$—a weaker condition than global Lipschitz continuity of $\nabla_{\boldsymbol{x}} \epsilon_\theta$—we obtain

$$\|\delta J(\lambda;\, v)\| \;\le\; \tfrac{1}{2} \Big(\sup_{s \in [0,1]} \big\| \nabla^2_{\boldsymbol{x}} \epsilon_\theta(\boldsymbol{x}^*_\lambda + sv, \lambda) \big\| \Big)\, \|v\| \;=\; \mathcal{O}(\|v\|),$$

and hence $\delta J(\lambda;\, v)\, v = \mathcal{O}(\|v\|^2)$. In particular $\delta J(\lambda_{t_{i-1}};\, \cdot\,) = \mathcal{O}(e_{i-1})$ and, since $u_i - \boldsymbol{x}^*_{s_i} = \mathcal{O}(e_{i-1})$, also $\delta J(\lambda_{s_i};\, \cdot\,) = \mathcal{O}(e_{i-1})$, so each Hessian correction contributes at second order, $\delta J\, v = \mathcal{O}(e^2_{i-1})$. Collecting the linear parts (driven by J^*) into a single propagation operator P_i, and grouping the remaining terms into the forcing f_i, yields

$$e_i = P_i\, e_{i-1} + f_i + \mathcal{O}(e^2_{i-1}).$$

where the propagator is

$$P_i = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}} \Big(1 + a_i\, J^*_{t_{i-1}} + b_i\, J^*_{s_i} + c_i\, J^*_{t_{i-1}} J^*_{s_i} \Big), \tag{52}$$

with coefficients

$$a_i = -\left(1 - \frac{1}{2r_1}\right) \alpha_{t_{i-1}}\, I^{\text{full}}_i, \quad b_i = -\frac{\alpha_{s_i}}{2r_1}\, I^{\text{full}}_i, \quad c_i = \frac{\alpha_{s_i} \alpha_{t_{i-1}}}{2r_1}\, I^{\text{half}}_i\, I^{\text{full}}_i,$$

$$I^{\text{full}}_i := \int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}\, d\lambda, \qquad I^{\text{half}}_i := \int_{\lambda_{t_{i-1}}}^{\lambda_{s_i}} e^{-\lambda}\, d\lambda.$$

The higher-order corrections to P_i due to δJ are:

$$\delta P^{(1)}_i = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}} \Big[(a_i + c_i J^*_{s_i})\, \delta J_{t_{i-1}}\, e_{i-1} + (b_i + c_i J^*_{t_{i-1}})\, \delta J_{s_i}\, e_{s_i} \Big] = \mathcal{O}(e^2_{i-1}),$$

and are absorbed into the $\mathcal{O}(e^2_{i-1})$ remainder throughout.

E.7 Series Expansion of the Propagator Product

Define the per-step Jacobian correction

$$\varepsilon_k := a_k J^*_{t_{k-1}} + b_k J^*_{s_k} + c_k J^*_{t_{k-1}} J^*_{s_k},$$

so that $P^{(0)}_k = \frac{\alpha_{t_k}}{\alpha_{t_{k-1}}}(1 + \varepsilon_k)$ and the propagator product telescopes to

$$\Pi^*_{j \to N} = \frac{\alpha_{t_N}}{\alpha_{t_j}} \prod_{k=j+1}^{N} (1 + \varepsilon_k).$$

Expanding the product order by order in ε:

$$\prod_{k=j+1}^{N}(1+\varepsilon_k) = \underbrace{1}_{m=0} + \underbrace{\sum_{k>j}\varepsilon_k}_{m=1} + \underbrace{\sum_{j<k_1<k_2}\varepsilon_{k_1}\varepsilon_{k_2}}_{m=2} + \cdots \tag{53}$$

Substituting into the unrolled error $e_N = \sum_{j=1}^{N}\Pi^*_{j\to N} f_j$ gives

$$e_N = \alpha_{t_N}\sum_{j=1}^{N}\frac{f_j}{\alpha_{t_j}}\left[1+\sum_{k>j}\varepsilon_k + \sum_{j<k_1<k_2}\varepsilon_{k_1}\varepsilon_{k_2}+\cdots\right]. \tag{54}$$

Truncating at second order (i.e. retaining $m \le 2$) and writing $\tilde{f}_j := f_j/\alpha_{t_j}$, $F^<_k := \sum_{j=1}^{k-1}\tilde{f}_j$:

$$e_N^{[\le 2]} = \alpha_{t_N}\left[\underbrace{\sum_j \tilde{f}_j}_{e_N^{[0]}} + \underbrace{\sum_k \varepsilon_k F^<_k}_{e_N^{[1]}} + \underbrace{\sum_{k_1<k_2}\varepsilon_{k_1}\varepsilon_{k_2}F^<_{k_1}}_{e_N^{[2]}}\right].$$

E.8 Cost Functional: General Form

With the unrolled recursion $e_N = \sum_{j=1}^{N}\Gamma_j f_j$, the mean squared error at the final step is

$$\mathcal{C} := \mathbb{E}\|e_N\|^2 = \mathbb{E}\left\|\sum_{j=1}^{N}\Gamma_j f_j\right\|^2 = \sum_{j=1}^{N}\sum_{k=1}^{N}\Gamma_j\Gamma_k\,\mathbb{E}[f_j^\top f_k]. \tag{55}$$

Decomposing the forcing cross-covariance. Each forcing term decomposes as $f_j = f_j^\eta + f_j^{\mathrm{disc}}$, giving

$$\mathbb{E}[f_j^\top f_k] = \underbrace{\mathbb{E}[(f_j^\eta)^\top f_k^\eta]}_{\text{network error cross-cov.}} + \underbrace{\mathbb{E}[(f_j^{\mathrm{disc}})^\top f_k^{\mathrm{disc}}]}_{\text{disc. error cross-cov.}} + \underbrace{\mathbb{E}[(f_j^\eta)^\top f_k^{\mathrm{disc}}] + \mathbb{E}[(f_j^{\mathrm{disc}})^\top f_k^\eta]}_{\text{cross terms}},$$

so the cost functional splits accordingly:

$$\mathcal{C} = \mathcal{C}^\eta + \mathcal{C}^{\mathrm{disc}} + \mathcal{C}^{\mathrm{cross}}. \tag{56}$$

Network error component. Recalling the weight kernel from equation 49,

$$f_j^\eta = \alpha_{t_j}\int_{\lambda_{t_{j-1}}}^{\lambda_{t_j}} e^{-\lambda}\eta(\boldsymbol{x}^*_\lambda,\lambda)\,d\lambda \;-\; \frac{\widetilde{J}_{s_j}}{2r_1}\,\sigma_{t_j}(e^{h_j}-1)\,\alpha_{s_j}\int_{\lambda_{t_{j-1}}}^{\lambda_{s_j}} e^{-\lambda}\eta(\boldsymbol{x}^*_\lambda,\lambda)\,d\lambda,$$

which may be written compactly as

$$f_j^\eta = \int_{\lambda_{t_{j-1}}}^{\lambda_{t_j}} w_j(\mu)\,\eta(\boldsymbol{x}^*_\mu,\mu)\,d\mu, \qquad w_j(\mu) = e^{-\mu}\begin{cases}\alpha_{t_j} - m_j\,\alpha_{s_j} & \mu\in[\lambda_{t_{j-1}},\lambda_{s_j}]\\ \alpha_{t_j} & \mu\in(\lambda_{s_j},\lambda_{t_j}]\end{cases}$$

with $m_j = \frac{J^*_{t_{j-1}}}{2r_1}\,\alpha_{t_j}\,I_j^{\mathrm{full}}$. The network error cross-covariance term is therefore

$$\mathcal{C}^\eta = \sum_{j,k}\Gamma_j\Gamma_k\int_{\lambda_{t_{j-1}}}^{\lambda_{t_j}}\int_{\lambda_{t_{k-1}}}^{\lambda_{t_k}} w_j(\mu)\,w_k(\nu)\,\underbrace{\mathbb{E}[\eta(\boldsymbol{x}^*_\mu,\mu)^\top\eta(\boldsymbol{x}^*_\nu,\nu)]}_{=:\,\mathcal{K}_\eta(\mu,\nu)}\,d\mu\,d\nu. \tag{57}$$

Discretization error component. By the Peano kernel theorem, the local truncation error satisfies $f_j^{\text{disc}} = \mathcal{O}(h_j^3)$ for DPM-Solver-2, one order higher than the network forcing $f_j^\eta = \mathcal{O}(h_j)$. Its cross-covariance contribution is

$$\mathcal{C}^{\text{disc}} = \sum_{j,k} \Gamma_j \Gamma_k \, \mathbb{E}[(f_j^{\text{disc}})^\top f_k^{\text{disc}}]. \tag{58}$$

Cross term: exact form and Cauchy–Schwarz control. One might assume $\mathbb{E}[\eta] = 0$. However, Eq. equation 11 and the plateau in Fig. 1 show that this assumption generally does not hold, except for models such as EDM (Karras et al., 2022). Writing $e_N^\eta = \sum_j \Gamma_j f_j^\eta$ and $e_N^{\text{disc}} = \sum_j \Gamma_j f_j^{\text{disc}}$, the cross contribution is controlled by Cauchy–Schwarz,

$$\left|\mathcal{C}^{\text{cross}}\right| = 2\left|\mathbb{E}[(e_N^\eta)^\top e_N^{\text{disc}}]\right| \;\le\; 2\sqrt{\mathcal{C}^\eta \, \mathcal{C}^{\text{disc}}}.$$

As $f^{\text{disc}} = \mathcal{O}(h^3)$ is one order higher than $f^\eta = \mathcal{O}(h)$, both $\mathcal{C}^{\text{disc}}$ and $\mathcal{C}^{\text{cross}}$ are higher-order in h relative to $\mathcal{C}^\eta$, which is therefore the leading objective.

Combining equation 56 with the Cauchy–Schwarz control of the cross term, the leading objective reduce to

$$\boxed{\mathcal{C} \approx \sum_{j=1}^{N}\sum_{k=1}^{N} \Gamma_j \Gamma_k \iint w_j(\mu)\, \mathcal{K}_\eta(\mu,\nu)\, w_k(\nu)\, d\mu\, d\nu \;+\; \sum_{j=1}^{N}\sum_{k=1}^{N} \Gamma_j \Gamma_k \, \mathbb{E}[(f_j^{\text{disc}})^\top f_k^{\text{disc}}].} \tag{59}$$

This is the full cost functional, containing both diagonal and off-diagonal terms. The off-diagonal terms $j \neq k$ couple forcing across different steps through the full covariance $\mathcal{K}_\eta(\mu,\nu)$. Appendix C shows this coupling admits an exact two-sided (sandwich) bound: equation 59 is bracketed between banded quadratic forms, so restricting to a diagonal neighbourhood incurs no approximation and yields the per-step decomposition used in equation 13.

F Special Cases

`logSNR-uniform`	*(degenerate case of semilinear diffusion ODE, log-SNR coordinate λ)*
`uniform`	*(degenerate case of flow matching / Euler, coordinate t)*

logSNR-uniform

Recall the reduced cost from the banded analysis,

$$\mathcal{C} \;=\; \alpha_{t_N}^2 \sum_{j=1}^{N} \Gamma_j^2 \, V_j^{\text{res}} \;+\; \rho_\infty \Big(\textstyle\sum_j \Gamma_j a_j \Big)^2, \qquad \Gamma_j = \textstyle\prod_{k>j}(1+\bar{\varepsilon}_k),$$

whose first block carries the per-step residual variance $V_j^{\text{res}} = \mathbb{E}\|\tilde{f}_j^{\text{res}}\|^2$ of the normalized forcing $\tilde{f}_j = \int_{\lambda_{j-1}}^{\lambda_j} e^{-\mu}\xi(\mu)\, d\mu$, and whose second block is the leading (coherent) mode of the systematic bias, with $a_j = \int_{\lambda_{j-1}}^{\lambda_j} e^{-\mu}\sigma_\eta(\mu)\, d\mu$. The cross-step terms have already been bracketed two-sidedly by a node-independent $1 + \mathcal{O}(C_\rho)$ factor, so the schedule sees only the diagonal energy above. The residual variance is the doubly-weighted kernel integral

$$V_j^{\text{res}} = \Psi(\lambda_{j-1}, \lambda_j), \qquad \Psi(a,b) := \int_a^b \int_a^b e^{-(\mu+\mu')}\, \mathcal{K}_\xi^{\text{res}}(\mu,\mu')\, d\mu\, d\mu',$$

with the stationary residual kernel in separable form $\mathcal{K}_\xi^{\text{res}}(\mu,\mu') = \sigma_\xi(\mu)\sigma_\xi(\mu')\rho(|\mu-\mu'|)$.

Degeneration to logSNR-uniform

The schedule optimization degenerates to *uniform log-SNR spacing* under two idealizations:

(i) **no propagator correction**: $J^* = 0$ (so $\bar{\varepsilon}_k = 0$, $\Gamma_j \equiv 1$); and

(ii) **small steps**: $\Delta\lambda_i \ll \ell$, over which $e^{-\mu}$ and σ_ξ are locally constant (Step 3).

Then the minimizer of $\mathcal{C}$ subject to $\sum_j \Delta\lambda_j = \Lambda$ is

$$\boxed{\Delta\lambda_j = \frac{\Lambda}{N} \quad \forall j}$$

independently of the decay profile of ρ and of the global shape of σ_ξ. Both conditions are exactly what schedule optimization must break: propagator corrections are present ($J^* \neq 0$), and the interesting regime is the step-limited one ($\Delta\lambda_i \not\ll \ell$, few NFEs). logSNR-uniform is thus the vanishing-curvature, small-step corner our method departs from.

Step 0: the systematic bias is an additive constant. Under (i), $\Gamma_j \equiv 1$, reducing the bias block to $\rho_\infty S^2$ with $S = \sum_j a_j = \int_0^\Lambda e^{-\mu}\sigma_\eta(\mu)\,d\mu$ fixed by the endpoints. This makes the bias a schedule-independent *additive constant* that shifts the cost without altering its $\arg\min$. Thus, optimal nodes depend solely on the residual block, and equidistribution holds for any ρ_∞ without requiring a zero-mean ($\rho_\infty = 0$) assumption. (The bias only affects the optimum when $J^* \neq 0$ renders Γ_j node-dependent and breaks S, which is the regime our method addresses.)

Step 1: Stationarity at node λ_i. The node λ_i is the right endpoint of step i and the left endpoint of step $i+1$. From

$$\frac{\partial\Psi(a,b)}{\partial b} = 2\int_a^b e^{-(\mu+b)}\mathcal{K}_\xi^{\mathrm{res}}(\mu,b)\,d\mu, \qquad \frac{\partial\Psi(a,b)}{\partial a} = -2\int_a^b e^{-(\mu+a)}\mathcal{K}_\xi^{\mathrm{res}}(\mu,a)\,d\mu,$$

the condition $\partial V_i^{\mathrm{res}}/\partial\lambda_i + \partial V_{i+1}^{\mathrm{res}}/\partial\lambda_i = 0$ becomes, after factoring $2e^{-\lambda_i}$,

$$\int_{\lambda_{i-1}}^{\lambda_i} e^{-\mu}\,\mathcal{K}_\xi^{\mathrm{res}}(\mu,\lambda_i)\,d\mu = \int_{\lambda_i}^{\lambda_{i+1}} e^{-\mu}\,\mathcal{K}_\xi^{\mathrm{res}}(\mu,\lambda_i)\,d\mu.$$

Step 2: Separable kernel. Substituting the separable form and cancelling $\sigma_\xi(\lambda_i)$, then setting $s = |\mu - \lambda_i|$,

$$\int_0^{\Delta\lambda_i} e^{-(\lambda_i - s)}\sigma_\xi(\lambda_i - s)\rho(s)\,ds = \int_0^{\Delta\lambda_{i+1}} e^{-(\lambda_i + s)}\sigma_\xi(\lambda_i + s)\rho(s)\,ds.$$

This is exact—no decay assumption on ρ has been used yet.

Step 3: Small-step limit. For $\Delta\lambda_i \ll \ell$, $e^{\pm s} \approx 1$ and $\sigma_\xi(\lambda_i \pm s) \approx \sigma_\xi(\lambda_i)$; the common factor $e^{-\lambda_i}\sigma_\xi(\lambda_i)$ cancels, leaving the correlation mass

$$R(\Delta\lambda_i) = R(\Delta\lambda_{i+1}), \qquad R(\Delta\lambda) := \int_0^{\Delta\lambda}\rho(s)\,ds.$$

Since $\rho > 0$ makes R strictly increasing, $\Delta\lambda_i = \Delta\lambda_{i+1}$ for all i, i.e. $\Delta\lambda_j = \Lambda/N$. Both the shape of ρ and the profile of σ_ξ drop out in the cancellation, so uniform log-SNR spacing follows from the separable structure alone—not from white noise. □

Why the operating point is not logSNR-uniform. Each idealization above fails in practice, and our method targets exactly these gaps: (i) propagator corrections are present ($J^* \neq 0$), making Γ_j node-dependent so that the systematic bias re-enters through $\partial_{\lambda_i}\Gamma_j$ and pulls the optimum off uniform spacing; (ii)–(iii) the residual variance $\sigma_\xi(\lambda)$ is strongly non-stationary and steps are not uniformly small against ℓ, so the cancellation in Step 3 is only approximate. logSNR-uniform is thus the $J^* \to 0$, stationary-error limit; the schedule we optimize departs from it precisely where these assumptions break.

UNIFORM

Tthe integrating factor of flow matching $\dot{\boldsymbol{x}} = v_\theta(\boldsymbol{x}, t)$ degenerates to the identity ($e^{-\mu} \to 1$) and the working coordinate collapses from λ back to t. The entire logSNR-uniform derivation carries over verbatim with $e^{-\mu} \to 1$, $\lambda \to t$, $\Lambda \to T$, giving

$$\boxed{\Delta t_j = \frac{T}{N} \quad \forall j} \qquad \textit{(uniform).}$$

logSNR-uniform and uniform are thus the same equidistribution principle read in two coordinates: the semilinear diffusion sampler equidistributes in λ, and flow matching sampler equidistributes in t.

G ALGORITHUMS

G.1 BASELINE

Uniform (Flux.1-dev). Nodes are spaced uniformly in flow-matching time t ($t_i = t_{\max} - \frac{i}{N}(t_{\max} - t_{\min})$), equivalent to uniform spacing in σ.

logSNR (Stable Diffusion v1.5). Nodes are spaced uniformly in DPM-Solver's log-SNR variable $\lambda_t = \log(\alpha_t/\sigma_t)$ ($\lambda_i = \lambda_{\max} - \frac{i}{N}(\lambda_{\max} - \lambda_{\min})$) (Lu et al., 2022).

Karras (EDM/EDM2). The ρ-warped schedule of Karras et al. (2022),

$$\sigma_i = \left(\sigma_{\max}^{1/\rho} + \tfrac{i}{N}\left(\sigma_{\min}^{1/\rho} - \sigma_{\max}^{1/\rho}\right)\right)^\rho, \qquad i = 0, \dots, N, \tag{60}$$

with $\rho = 7$ recommended.

AYS and CTR. For Stable Diffusion v1.5, we compare with AYS (Sabour et al., 2024) and Xue et al. (2024) (CTR). For budgets > 10 steps, we apply AYS's recommended log-linear interpolation (Listing 1).

Matched-NFE comparison. All comparisons use *matched* NFE. EP-Flow is initialized from the corresponding baseline schedule for each NFE budget and optimized at that budget. The resulting schedules and their initializations are listed in Appendix H (Tables 8–11 and 6–7).

Listing 1: Log-linear interpolation for AYS schedules.

```
def loglinear_interp(t_steps, num_steps):
    xs = np.linspace(0, 1, len(t_steps))
    ys = np.log(t_steps[::-1])

    new_xs = np.linspace(0, 1, num_steps)
    new_ys = np.interp(new_xs, xs, ys)

    return np.exp(new_ys)[::-1].copy()
```

G.2 ALGORITHUM

We instantiate the `EP-flow` framework for two model families whose solver structure differs.

- **FLUX.1-dev** (flow matching, Euler): the propagator reduces to $P_k^{(0)} = 1 + h_k g(t_{k-1})$ since $b_k = c_k = 0$ (no half-step corrector, equation 52).
- **Stable Diffusion v1.5** (VP-SDE, DPM-Solver-2): the half-step corrector restores $b_k, c_k \neq 0$. Moreover, the far-field covariance is not globally stationary, so we adopt the *piecewise-kernel* formulation of Appendix C.

The corresponding scheduling procedures are given in Algorithm 2 for FLUX.1-dev and Algorithm 3 for Stable Diffusion v1.5.

Algorithm 2 EP-FLOW timestep scheduler — FLUX.1-dev (flow matching, Euler)

Require: offline statistics $(\rho, \sigma_\eta^2, \sigma_{\dot{v}}^2, g)$; endpoints $(\sigma_{\max}, \sigma_{\min})$; steps N; iterations M
Ensure: strictly decreasing schedule $\{\sigma_j\}_{j=0}^{N}$

Offline preparation
load interpolators $g(\sigma),\ \sigma_\eta^2(\sigma),\ \sigma_{\dot{v}}^2(\sigma)$
$\rho_\infty \leftarrow \textsc{ExtractPlateau}(\rho)$; $\phi^{\text{res}} \leftarrow \textsc{ResidualKernel}(\rho, \rho_\infty)$
$C_\rho^\eta \leftarrow \int \rho_\eta^{\text{res}}(s)\, ds$; $C_\rho^d \leftarrow \int \rho_{\dot{v}}(s)\, ds$
$\boldsymbol{\theta} \leftarrow \textsc{Init}(N, \sigma_{\max}, \sigma_{\min})$

for $m = 1, \dots, M$ **do**
Project parameters → schedule
$\mathbf{h} \leftarrow \mathrm{softmax}(\boldsymbol{\theta})\,(\sigma_{\max} - \sigma_{\min})$; $\sigma_j \leftarrow \sigma_{\max} - \sum_{i \le j} h_i$ ▷ $\sigma_N = \sigma_{\min}$
Per-step quantities $(j = 1, \dots, N)$
$\epsilon_j \leftarrow h_j\, g(\sigma_{j-1})$; $\Gamma_j \leftarrow \prod_{k>j}(1 + \epsilon_k)$
$a_j \leftarrow \int_{\sigma_j}^{\sigma_{j-1}} \sigma_\eta(\tau)\, d\tau$; $V_j^{\text{res}} \leftarrow \sigma_\eta^2(\bar{\sigma}_j)\, \phi^{\text{res}}(h_j, h_j)$; $D_j \leftarrow \frac{1}{4} h_j^4\, \sigma_{\dot{v}}^2(\sigma_{j-1})$
Cost

$$C = \underbrace{\rho_\infty \Big(\textstyle\sum_j \Gamma_j a_j \Big)^2}_{\text{rank-1 bias}} + \underbrace{(1 + 2C_\rho^\eta) \textstyle\sum_j \Gamma_j^2 V_j^{\text{res}}}_{\text{residual variance}} + \underbrace{(1 + 2C_\rho^d) \textstyle\sum_j \Gamma_j^2 D_j}_{\text{discretization}}$$

Update
$\widehat{\nabla}_{\boldsymbol{\theta}} C \leftarrow \textsc{FiniteDiff}(C, \delta)$; $\boldsymbol{\theta} \leftarrow \textsc{Adam}(\boldsymbol{\theta}, \widehat{\nabla}_{\boldsymbol{\theta}} C)$; $\mathrm{lr} \leftarrow \gamma \cdot \mathrm{lr}$
if $\Delta C / C < \tau$ **then break**
end for
return $\{\sigma_j\}_{j=0}^{N}$

H Optimized Schedules

Tables 6–11 list the optimized schedules used in our experiments, comparing each baseline against EP-Flow initialized from it: Flux.1-dev under uniform initialization (Table 8) and Stable Diffusion v1.5 under the AYS, logSNR, and CTR initializations (Tables 9–11).

I Additional Benchmark Results

This section presents additional benchmark results that complement the evidence for§5.

Table 12 shows that the cosine schedule is an empirically strong baseline for FLUX.1-dev.

Tables 13, 15, and 14 show that, as the optimization signal fades in the high-NFE regime, α mainly serves as a knob controlling the trade-off across competing evaluation axes.

Table 16 shows a trend consistent with the predicted α-scaling. Table 17 provides the evidence that predicted α also transfers effectively to unseen metrics like PickScore without re-tuning, matching the performance of the metric's own empirically tuned optimum.

J Qualitative Samples

We provide qualitative comparisons across different model families, sampling budgets, and timestep scheduling methods. The examples include EDM/EDM2, FLUX.1-dev, and Stable Diffusion v1.5, with comparisons against Karras, AYS, and Uniform schedules, as shown in Figs. 7 and 8–11. These samples provide a visual reference for the differences between timestep schedules and complement the quantitative evaluation presented in Table 1.

Algorithm 3 EP-FLOW timestep scheduler — SD v1.5 (VP-SDE, DPM-Solver-2)

Require: segment statistics $\{(\rho^{(k)}, \rho_\infty^{(k)}, s^{(k)})\}_{k=1}^K$; steps N; iterations M; corrector ratio r_1
Ensure: strictly increasing schedule $\{\lambda_j\}_{j=0}^N$

Offline preparation
build segment kernels $\phi^{(k)}$ from $\rho^{(k)}$; plateaus $\rho_\infty^{(k)}$; banded scaling $b^{(k)} \leftarrow 1 + 2C_\rho^{(k)}$
initialize $\boldsymbol{\theta} \in \mathbb{R}^N$

for $m = 1, \dots, M$ **do**
Project parameters → schedule
$\mathbf{h} \leftarrow \mathrm{softmax}(\boldsymbol{\theta})\,(\lambda_{\max} - \lambda_{\min})$; $\lambda_j \leftarrow \lambda_{\min} + \sum_{i \le j} h_i$
Segment assignment $(j = 1, \dots, N)$
$k(j) \leftarrow \textsc{Locate}(\lambda_j)$; $\mathcal{S}_k \leftarrow \{\, j : k(j) = k \,\}$; $(\phi_j, b_j) \leftarrow (\phi^{(k(j))}, b^{(k(j))})$
Per-step quantities $(j = 1, \dots, N)$: $\Gamma_j,\ \tilde{A}_j,\ V_j^{\mathrm{res}},\ D_j$
Cost

$$C = \alpha_{t_N}^2 \Big[\underbrace{\sum_{k=1}^K \rho_\infty^{(k)} \Big(\sum_{j \in \mathcal{S}_k} \Gamma_j\, \tilde{A}_j \Big)^2}_{\text{piecewise rank-1 bias}} + \underbrace{\sum_j b_j\, \Gamma_j^2\, (V_j^{\mathrm{res}} + D_j)}_{\text{residual variance \& discretization}} \Big]$$

Update
$\widehat{\nabla}_{\boldsymbol{\theta}} C \leftarrow \textsc{FiniteDiff}(C, \delta)$; $\boldsymbol{\theta} \leftarrow \textsc{Adam}(\boldsymbol{\theta}, \widehat{\nabla}_{\boldsymbol{\theta}} C)$; $\mathrm{lr} \leftarrow \gamma \cdot \mathrm{lr}$
if $\|\widehat{\nabla}_{\boldsymbol{\theta}} C\| < \tau$ **then break**
end for
return $\{\lambda_j\}_{j=0}^N$

K Limitations and Future work

Limitations. We optimize the derived cost functional with finite-difference numerical gradients and a simple multi-restart strategy; the finite-difference estimates can become noisy as the optimization headroom shrinks, while analytical gradients may improve optimization efficiency and precision. Current use of standard Adam GD leaves room for more advanced optimization methods to better exploit the resulting structured, multi-basin landscape (Fig. 2).

Future Work. Beyond schedule optimization, the propagator–forcing decomposition may also provide a principled basis for designing advanced numerical solvers that explicitly account for error propagation.

Table 6: Learned σ schedules on EDM (Karras et al., 2022)

Dataset	NFE	Learned σ schedule
AFHQv2 64×64	6	80.0000, 23.3520, 5.1098, 1.2365, 0.1970, 0.0356, 0.0020
	10	80.0000, 44.8172, 23.6135, 11.4173, 4.7921, 1.7216, 0.6623, 0.2245, 0.0654, 0.0167, 0.0020
	14	80.0000, 53.1713, 34.3337, 21.3966, 12.7337, 7.1004, 3.5801, 1.7331, 0.8869, 0.4274, 0.1920, 0.0791, 0.0303, 0.0098, 0.0020
	18	80.0000, 58.3553, 41.8592, 29.4481, 20.2419, 13.5200, 8.6990, 5.3157, 3.0290, 1.7185, 1.0227, 0.5910, 0.3278, 0.1740, 0.0877, 0.0424, 0.0200, 0.0071, 0.0020
FFHQ 64×64	6	80.0000, 16.5791, 5.7300, 1.5527, 0.2180, 0.0328, 0.0020
	10	80.0000, 44.5699, 19.0317, 6.3062, 3.3311, 1.9810, 0.6020, 0.2007, 0.0606, 0.0205, 0.0020
	14	80.0000, 53.2892, 34.5341, 21.6534, 13.0266, 7.4129, 3.8829, 1.7796, 0.8740, 0.4203, 0.1876, 0.0762, 0.0282, 0.0097, 0.0020
	18	80.0000, 45.2166, 36.4942, 22.8066, 16.8126, 12.4005, 6.1517, 4.7682, 1.9827, 1.3255, 0.6986, 0.3712, 0.2496, 0.1615, 0.0711, 0.0365, 0.0160, 0.0057, 0.0020
ImageNet 64×64	6	80.0000, 29.6496, 8.9708, 1.7033, 0.3331, 0.0453, 0.0020
	10	80.0000, 44.9141, 23.7682, 11.6038, 4.9892, 1.6944, 0.6614, 0.2241, 0.0652, 0.0166, 0.0020
	14	80.0000, 53.2414, 34.4525, 21.5484, 12.9056, 7.2817, 3.7475, 1.6914, 0.8795, 0.4240, 0.1903, 0.0781, 0.0296, 0.0097, 0.0020
	18	80.0000, 58.4210, 41.9744, 29.5999, 20.4203, 13.7164, 8.9055, 5.5213, 3.1998, 1.7057, 1.0131, 0.5862, 0.3251, 0.1721, 0.0862, 0.0412, 0.0190, 0.0070, 0.0020
CIFAR-10 32×32	6	80.0000, 15.0198, 2.6808, 1.1802, 0.1249, 0.0191, 0.0020
	10	80.0000, 44.9167, 23.7724, 11.6088, 4.9955, 1.7017, 0.6610, 0.2230, 0.0644, 0.0160, 0.0020
	14	80.0000, 53.2456, 34.4597, 21.5577, 12.9165, 7.2941, 3.7610, 1.6973, 0.8747, 0.4208, 0.1880, 0.0764, 0.0283, 0.0095, 0.0020
	18	80.0000, 58.4278, 41.9865, 29.6162, 20.4397, 13.7386, 8.9306, 5.5495, 3.2332, 1.7395, 1.0192, 0.5890, 0.3252, 0.1710, 0.0847, 0.0397, 0.0177, 0.0070, 0.0020

Table 7: Learned schedules for EDM2 (Karras et al., 2024)

Schedule	NFE	σ schedule
5	Karras	80.0000, 24.4083, 5.8389, 0.9654, 0.0851, 0.0020
	EP-Flow	80.0000, 20.7148, 3.9740, 0.4514, 0.0873, 0.0020
10	Karras	80.0000, 45.3137, 24.4083, 12.3816, 5.8389, 2.5152, 0.9654, 0.3183, 0.0851, 0.0167, 0.0020
	EP-Flow	80.0000, 50.3272, 28.7040, 12.3086, 10.2556, 2.6545, 0.7812, 0.3018, 0.1411, 0.0691, 0.0020
15	Karras	80.0000, 55.0508, 37.0918, 24.4083, 15.6401, 9.7232, 5.8389, 3.3686, 1.8543, 0.9654, 0.4700, 0.2107, 0.0851, 0.0300, 0.0088, 0.0020
	EP-Flow	80.0000, 54.6593, 36.4185, 23.5360, 14.6304, 8.6216, 4.6823, 2.2460, 1.2750, 0.6851, 0.3521, 0.1754, 0.0852, 0.0303, 0.0089, 0.0020

Table 8: Optimized sampling schedules on **Flux.1-dev**. For each NFE we report the uniform baseline and EP-Flow initialized from it (steps in t, $0.99 \to 0.01$).

NFE	Method	Schedule
4	Uniform	0.99, 0.745, 0.5, 0.255, 0.01
	EP-Flow	0.99, 0.777, 0.52, 0.246, 0.01
6	Uniform	0.99, 0.827, 0.663, 0.5, 0.337, 0.173, 0.01
	EP-Flow	0.99, 0.84, 0.701, 0.552, 0.374, 0.204, 0.01
8	Uniform	0.99, 0.868, 0.745, 0.623, 0.5, 0.378, 0.255, 0.133, 0.01
	EP-Flow	0.99, 0.872, 0.754, 0.636, 0.506, 0.371, 0.243, 0.118, 0.01
10	Uniform	0.99, 0.892, 0.794, 0.696, 0.598, 0.5, 0.402, 0.304, 0.206, 0.108, 0.01
	EP-Flow	0.99, 0.9, 0.811, 0.724, 0.635, 0.542, 0.436, 0.324, 0.239, 0.123, 0.01

Table 9: Optimized schedules on **Stable Diffusion v1.5**, AYS family. Baseline: AYS (Sabour et al., 2024); EP-Flow initialized from AYS (discrete timesteps, $999 \to 0$).

NFE	Method	Schedule
10	AYS	999, 850, 736, 645, 545, 455, 343, 233, 125, 24, 0
	EP-Flow	999, 917, 853, 779, 696, 602, 505, 397, 247, 85, 0
15	AYS	999, 903, 814, 736, 677, 613, 545, 486, 418, 343, 268, 192, 125, 44, 8, 0
	EP-Flow	999, 946, 904, 858, 806, 748, 684, 615, 546, 472, 390, 280, 167, 61, 3, 0
20	AYS	999, 928, 850, 795, 736, 692, 645, 596, 545, 501, 455, 399, 343, 286, 233, 174, 125, 58, 24, 4, 0
	EP-Flow	999, 946, 911, 874, 832, 787, 743, 696, 643, 591, 541, 488, 435, 372, 282, 188, 130, 87, 50, 3, 0

Table 10: Optimized schedules on **Stable Diffusion v1.5**, logSNR family (discrete timesteps).

NFE	Method	Schedule
4	logSNR	999, 681, 233, 20, 0
	EP-Flow	999, 628, 359, 30, 0
6	logSNR	999, 803, 538, 233, 52, 7, 0
	EP-Flow	999, 836, 671, 516, 349, 3, 0
8	logSNR	999, 858, 681, 461, 233, 81, 20, 4, 0
	EP-Flow	999, 870, 728, 581, 440, 252, 53, 3, 0
10	logSNR	999, 989, 969, 932, 865, 753, 597, 425, 265, 125, 1
	EP-Flow	999, 906, 814, 711, 601, 497, 390, 244, 74, 17, 1

Table 11: Optimized schedules on **Stable Diffusion v1.5**, CTR family. Baseline: CTR (Xue et al., 2024).

NFE	Method	Schedule
4	CTR	999, 576, 313, 98, 0
	EP-Flow	999, 660, 412, 90, 0
6	CTR	999, 660, 433, 304, 174, 60, 0
	EP-Flow	999, 812, 635, 471, 273, 55, 0
8	CTR	999, 850, 729, 679, 518, 346, 187, 58, 0
	EP-Flow	999, 904, 799, 680, 556, 427, 247, 55, 0

Table 12: **Human Preference Score (HPSv2) results.** Comparison across four visual categories. Mean score ± standard deviation. **Bold** = best

Method	ANIME	CONCEPT	PAINTINGS	PHOTOS	AVERAGE
FLUX Default	32.13 ± 0.27	30.17 ± 0.34	30.62 ± 0.18	28.67 ± 0.44	30.45
logSNR-uniform	32.38 ± 0.32	30.03 ± 0.32	30.64 ± 0.24	28.73 ± 0.46	30.44
Cosine	**32.89 ± 0.31**	**30.75** ± 0.30	**31.22** ± 0.20	**29.19** ± 0.43	**31.01**
Beta	32.42 ± 0.32	30.22 ± 0.35	30.73 ± 0.21	28.64 ± 0.46	30.50
Uniform-σ	31.13 ± 0.33	28.95 ± 0.32	29.61 ± 0.21	27.38 ± 0.48	29.27

Table 13: Ablation of α scaling across multi-dimensional benchmarks in the medium-budget regime (NFE=10). Optimal values for each category are highlighted in **bold**.

Method	Color	Shape	Texture	2D Spat.	3D Spat.	Non-spat.	Numeracy
Cosine	0.7829	0.4831	0.6719	0.2268	0.3627	0.3105	0.6091
Ours ($\alpha = 0.00$)	0.7820	0.4928	**0.6906**	0.2336	0.3685	0.3111	0.6234
Ours ($\alpha = 0.25$)	0.7801	0.4964	0.6842	**0.2354**	0.3620	0.3108	0.6188
Ours ($\alpha = 0.50$)	**0.7855**	**0.4973**	0.6860	0.2152	**0.3700**	0.3109	0.6171
Ours ($\alpha = 0.75$)	0.7744	0.4853	0.6870	0.2333	0.3638	0.3111	**0.6300**
Ours ($\alpha = 1.00$)	0.7788	0.4825	0.6739	0.2337	0.3629	0.3108	0.6238

Table 14: **Quantitative comparison on T2ICompBench.** We report the performance across multiple dimensions including local attributes and global compositions. **Bold** = best, underline = second best.

Method	COLOR	SHAPE	TEXTURE	2D-SPATIAL	3D-SPATIAL	NUMERACY	NON-SPATIAL	COMPLEX
	B-VQA	B-VQA	B-VQA	UniDet	UniDet	UniDet	CLIP	3-in-1
logSNR-uniform	0.7774	0.4789	0.6252	0.2640	0.3871	0.5988	0.3078	0.4407
Cosine	**0.7827**	0.4769	0.6254	0.2576	0.3866	0.6021	0.3075	0.4449
Beta	0.7729	0.4861	0.6118	**0.2657**	0.3808	0.5944	0.3072	0.4312
FLUX Default	0.7531	**0.4922**	0.6164	0.2628	**0.3941**	0.6049	**0.3083**	0.4345
EP-Flow (Ours)	0.7767	0.4798	**0.6294**	0.2629	0.3909	**0.6134**	0.3080	**0.4463**

Table 15: **GenEval compositional accuracy across categories.** We report the performance of EP-Flow compared to standard diffusion schedules. **Bold** = best, underline = second best.

Method	2-OBJ	1-OBJ	COUNT	COLOR	POS.	CLRS	OVERALL
logSNR-uniform	**77.53**	99.38	71.25	45.25	18.75	76.86	0.64835
Cosine	74.75	**99.69**	**72.50**	45.75	18.75	78.72	0.65026
Beta	72.22	**99.69**	**72.50**	43.25	19.00	77.39	0.64609
FLUX Default	73.99	99.38	68.12	41.50	**19.50**	**79.26**	0.63624
EP-Flow (Ours)	77.02	99.38	70.62	**46.50**	19.25	77.66	**0.65072**

Table 16: Performance comparison across different NFE and α values.

NFE	Cosine	$\alpha = 0$	$\alpha = 0.15$	$\alpha = 0.25$	$\alpha = 0.33$	$\alpha = 0.40$	$\alpha = 0.50$	$\alpha = 0.55$
4	0.5372	0.5053	0.5382	0.5356	0.5351	0.5349	**0.5414**	0.5312
6	0.6268	0.6118	0.6215	0.6257	**0.6279**	0.6259	0.6212	-
8	0.6478	0.6418	0.6479	**0.6489**	0.6445	0.6418	0.6443	-

Table 17: **Transfer to PickScore** (without re-tuning, $n = 553 \times 4$ seeds). Gaps are defined as predicted $-$ comparator (positive values favor our prediction); brackets denote 95% bootstrap CIs. The predicted α matches the metric's own optimum (argmax) across all NFE levels and significantly outperforms the worst α (argmin) at NFE $= 4$ and 6, demonstrating that our $1/\text{NFE}$ scaling successfully lands at the high-performing end.

NFE	Predicted α	vs. argmax-α (best): Δ [95% CI], p	vs. argmin-α (worst): Δ [95% CI], p
4	0.50	-0.0039 $[-0.0117, +0.0038]$, n.s.	$+0.0267$ $[+0.0134, +0.0394]$, $p = 2.131 \times 10^{-6}$
6	0.33	-0.0092 $[-0.0229, +0.0039]$, n.s.	$+0.0180$ $[+0.0084, +0.0273]$, $p = 1.178 \times 10^{-6}$
8	0.25	-0.0036 $[-0.0122, +0.0047]$, n.s.	$+0.0031$ $[-0.0067, +0.0133]$, n.s.



Figure 7: Qualitative comparison of samples generated with Karras and EP-Flow timestep schedules. Results are shown for EDM2 at NFE=5 (left) and EDM on AFHQv2 at NFE=10 (right).

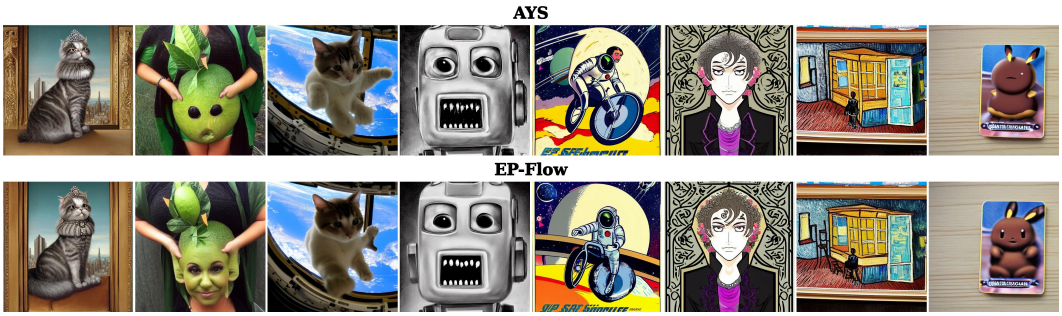


Figure 8: HPSv2 samples generated with AYS and EP-Flow timestep schedules at NFE=20.



Figure 9: HPSv2 samples generated with Uniform and EP-Flow timestep schedules at NFE=6.

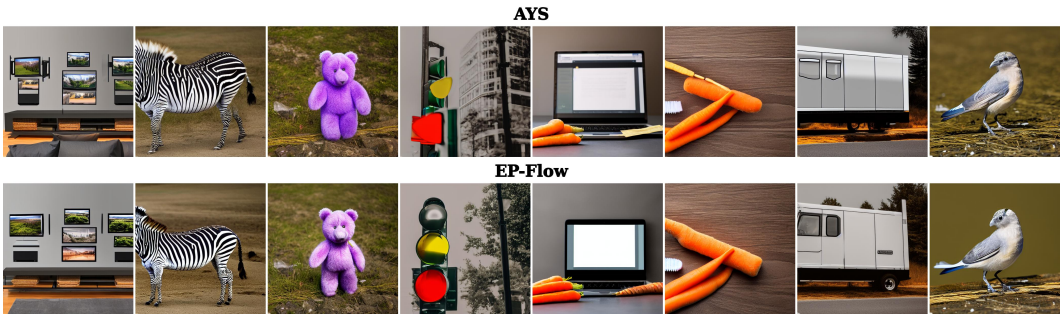


Figure 10: GenEval samples generated with AYS and EP-flow timestep schedules at NFE=10.

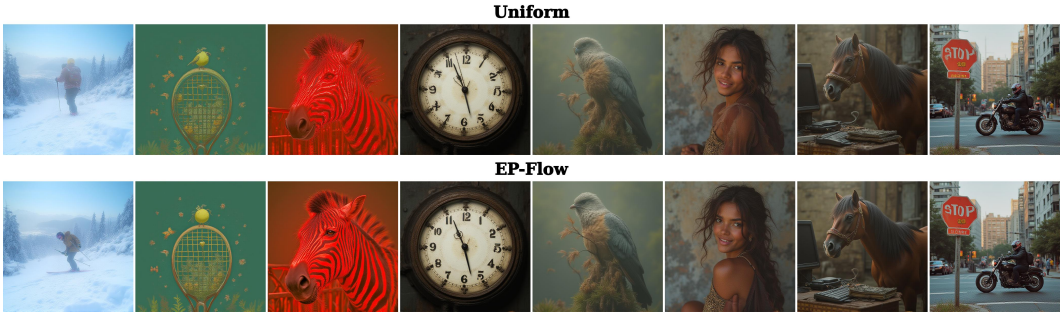


Figure 11: GenEval samples generated with Uniform and EP-flow timestep schedules at NFE=6.