DISTRIBUTION-GUIDED CLASSIFIER-FREE GUIDANCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Classifier-free guidance (CFG) is central to conditional generation with diffusion models, but using a constant guidance strength overlooks how the diffusion rate, signal content, and score reliability vary throughout denoising. We introduce Distribution-Guided CFG (DG-CFG), a training-free schedule that reallocates a fixed guidance budget across timesteps to improve the balance among generation quality, diversity, and stability. A path-integral analysis of deterministic CFG identifies the temporal weight that governs how guidance accumulates along sampling trajectories. Combining this principle with signal-content weighting and error-motivated low-noise attenuation yields a middle-focused schedule that emphasizes informative and reliable stages of denoising. Controlled experiments with analytic scores support the path-integral predictions and illustrate the effect of temporal reweighting. Across different backbones and sampling methods, DG-CFG provides a favorable diversity–fidelity trade-off and improves generation quality under strong guidance while controlling saturation. Component ablations and sampling-budget evaluations further support the schedule design and robustness.

1 INTRODUCTION

Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021b) have emerged as a leading paradigm in generative modeling, achieving state-of-the-art results across image synthesis (Dhariwal & Nichol, 2021; Rombach et al., 2022), video generation (Ho et al., 2022), audio synthesis (Kong et al., 2021), and molecular conformation generation (Xu et al., 2022). Their operating principle is conceptually elegant: a forward stochastic differential equation (SDE) progressively corrupts data with noise, and a learned reverse process—parameterized by the score function—removes this noise to reconstruct realistic samples. This framework, unified by Song et al. (2021b), admits two equivalent reverse-time descriptions: a stochastic SDE and a deterministic probability flow ODE that preserves the same marginal distributions at every timestep.

For conditional generation, *classifier-free guidance* (CFG) (Ho & Salimans, 2021) has become the de facto standard. Unlike classifier-based guidance (Dhariwal & Nichol, 2021), which requires an auxiliary classifier, CFG blends the score functions of a conditional model p_{data} and a reference model q_{data} (typically the unconditional model) through a guidance strength parameter ω:

$$s_t^{\omega}(x) = \omega \nabla_x \log p_t(x) + (1-\omega)\nabla_x \log q_t(x). \tag{1}$$

When $\omega > 1$, the conditional score is extrapolated beyond the reference score, often improving visual quality and prompt alignment but reducing sample diversity and causing supersaturation. Because it requires neither auxiliary networks nor additional training, this simple extrapolation underpins modern text-to-image systems (Rombach et al., 2022; Saharia et al., 2022; Ramesh et al., 2022; Balaji et al., 2022) and has also been adopted in video (Ho et al., 2022), audio (Liu et al., 2023), and 3D synthesis (Poole et al., 2023).

Time-dependent guidance is already an established approach to adapting CFG across denoising stages. Interval CFG (Kynkäänniemi et al., 2024) uses a selected rectangular time window, while β–CFG (Malarz et al., 2025) adopts an empirical beta-shaped schedule without explicitly tying it to the diffusion rate $\beta(t)$. C^{2}FG (Gao et al., 2026) instead controls instantaneous conditional–unconditional score discrepancies. The remaining question is how these local guidance effects accumulate along the sampling trajectory and alter the generated distribution. This motivates a continuous-time path-integral formulation for the probability flow ODE (PF-ODE), linking temporal guidance allocation to cumulative distributional transport.

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.

In this work, we introduce Distribution-Guided CFG (DG-CFG), a time-dependent guidance schedule designed around this temporal allocation problem. A path-integral analysis of deterministic probability flow ODE sampling shows that the guidance contribution is weighted by the diffusion rate, which directly motivates balancing the factor $\beta(t)(\omega(t) - 1)$. DG-CFG combines this time-balancing principle with the remaining signal content and the amplification of score errors near the low-noise endpoint. The resulting schedule concentrates guidance at intermediate noise levels while preserving the integrated guidance budget of constant CFG. Our main contributions are:

1. **Method.** We propose DG-CFG, a training-free schedule that reallocates guidance toward intermediate denoising stages. Its schedule $\omega(t) - 1 \propto (1 - \bar{\alpha}_t)\sqrt{\bar{\alpha}_t}/\beta(t)$ balances the explicit diffusion-rate coefficient, weights retained signal content, and attenuates low-noise guidance motivated by network errors.
2. **Design principle.** We characterize the distributions induced by constant and time-dependent CFG through path-integral representations. The analysis reveals how conditional–reference score discrepancies accumulate along sampling trajectories and identifies the temporal weighting that motivates DG-CFG (Theorems 3.2 and 3.3).
3. **Evaluation.** Across SD1.5, SD2.1, and SDXL, DG-CFG improves generation quality while controlling saturation under strong guidance. Analytic experiments, component ablations, and NFE evaluations support the distributional interpretation, schedule design, and robustness across sampling budgets.

2 Preliminaries

2.1 Diffusion Models

Diffusion models generate data by learning to reverse a gradual noising process. In the forward direction, data are transported toward a simple prior—typically a standard Gaussian—according to a prescribed SDE

$$dx = f(x, t)\, dt + g(t)\, dw, \tag{2}$$

over $[0, T]$, where w is a standard Wiener process and p_t denotes the marginal density at time t. The associated reverse-time SDE (Anderson, 1982) is parameterized by the *score function* $\nabla_x \log p_t(x)$:

$$dx = \big(f(x, t) - g^2(t)\, \nabla_x \log p_t(x)\big)\, dt + g(t)\, d\bar{w}, \tag{3}$$

where $\bar{w}$ is a reverse-time Wiener process. Song et al. (2021b) showed that the same marginal densities are also realized by the *deterministic probability flow ODE*

$$dx = \Big(f(x, t) - \frac{1}{2} g^2(t)\, \nabla_x \log p_t(x)\Big)\, dt, \tag{4}$$

which preserves the same marginal distributions at every t.

2.2 Classifier-Free Guidance

Classifier-free guidance (CFG) (Ho & Salimans, 2021) steers sampling without a separately trained classifier. Given a target distribution p_{data} and a reference distribution q_{data} (typically the unconditional distribution), CFG replaces the score in the reverse process with the affine combination

$$s_t^\omega(x) = \omega \nabla_x \log p_t(x) + (1 - \omega) \nabla_x \log q_t(x), \tag{5}$$

where $\omega \in \mathbb{R}$ is the guidance strength and p_t, q_t denote the forward-noised marginals of $p_{\text{data}}, q_{\text{data}}$. Substituting s_t^ω into the probability flow ODE (4) yields the *guided deterministic sampler*

$$dx = \Big(f(x, t) - \frac{1}{2} g^2(t)\, s_t^\omega(x)\Big)\, dt. \tag{6}$$

Under CFG extrapolation, the reverse SDE and PF-ODE generally induce different marginal distributions, even with exact target and reference scores (Bradley & Nakkiran, 2024). We focus on the deterministic ODE formulation in (6). When $\omega = 1$, it reduces to ordinary sampling from the target distribution. When $\omega > 1$, the target score is extrapolated away from the reference score, increasing conditional fidelity but often reducing diversity and causing color oversaturation.

3 PATH-INTEGRAL ANALYSIS FOR GUIDANCE SCHEDULING

To identify how guidance is allocated across denoising, we analyze the distribution produced by CFG under the deterministic probability flow ODE (6), without committing to a specific parameterization of $f(x,t)$ and $g(t)$. We first formulate the sampling problem (Section 3.1), then characterize the cumulative effect of constant guidance (Section 3.2) and extend the analysis to time-dependent schedules (Section 3.3). The resulting temporal weight provides the design principle used by DG-CFG in Section 4.

3.1 PROBLEM FORMULATION

Let p_{data} and q_{data} be two probability densities on $\mathbb{R}^n$. In the CFG setting, p_{data} represents the target conditional distribution and q_{data} the reference distribution, typically the unconditional distribution. Consider the forward noising SDE (2) on $[0,T]$. Let p_t denote the marginal density at time t initialized from $p_0 = p_{\mathrm{data}}$; it evolves according to

$$\frac{\partial p_t}{\partial t} = -\nabla \cdot (f p_t) + \frac{1}{2} g^2 \Delta p_t. \tag{7}$$

Similarly, let q_t be the marginal density initialized from $q_0 = q_{\mathrm{data}}$ and governed by the same Fokker–Planck equation. In practical diffusion sampling, reverse processes associated with different data distributions are typically initialized from the same Gaussian prior at time T.

Since the score may become singular near $t = 0$, we therefore introduce an arbitrarily small stopping time $t_0 > 0$ and analyze the reverse-time horizon $[t_0, T]$, consistent with practical discrete samplers that sample to the smallest positive noise level followed by a final denoising step. The guided reverse process maps the terminal reference density at time T to a density $\hat{r}_{t_0}$ at time t_0. All theoretical results are stated for fixed $t_0 > 0$, with p_{t_0} and q_{t_0} serving as small-noise approximations to their respective data distributions.

Classifier-free guidance modifies the deterministic probability flow ODE by replacing the target score with the blended score s_t^ω:

$$dx = \Big(f(x,t) - \frac{1}{2} g^2(t) \left(\omega \nabla_x \log p_t(x) + (1-\omega) \nabla_x \log q_t(x) \right) \Big)\, dt. \tag{8}$$

To analyze (8), define the normalized product reference distribution

$$\tilde{r}_t(x) := \frac{1}{Z_t} p_t^\omega(x)\, q_t^{1-\omega}(x) = \frac{1}{Z_t} r_t(x), \tag{9}$$

where $r_t(x) = p_t^\omega(x) q_t^{1-\omega}(x)$ and $Z_t = \int_{\mathbb{R}^n} r_t(x)\, dx$ is the normalizing constant which is assumed to be finite.

Remark (Integrability). For boundedly supported data and a Gaussian forward kernel $X_t = a_t X_0 + \sigma_t \epsilon$, $\epsilon \sim \mathcal{N}(0, I)$, both log-densities satisfy $\log h_t(x) = -\|x\|^2/(2\sigma_t^2) + O(1 + \|x\|)$, $h_t \in \{p_t, q_t\}$, as $\|x\| \to \infty$ for each fixed $t > 0$ with $\sigma_t > 0$. Thus $p_t^{\omega(t)} q_t^{1-\omega(t)} = \exp\big(-\|x\|^2/(2\sigma_t^2) + O(1 + \|x\|) \big)$ is integrable for every finite guidance strength $\omega(t)$, including $\omega(t) > 1$. Appendix B.1 also establishes finiteness of the expectations used below.

A key identity follows immediately:

$$\nabla_x \log \tilde{r}_t = \nabla_x \log r_t = \omega \nabla_x \log p_t + (1-\omega) \nabla_x \log q_t. \tag{10}$$

Consequently, the guided sampler (8) can be written as $dx = (f - \frac{1}{2} g^2\, \nabla_x \log \tilde{r}_t)\, dt$.

Let $\hat{r}_t$ denote the marginal density transported by the CFG reverse process on $[t_0, T]$, initialized from the terminal reference density $\tilde{r}_T$ at time T. Under the usual terminal condition $p_T \simeq q_T$, this density coincides with the common diffusion prior. Our goal is to characterize $\hat{r}_{t_0}$, the distribution from which deterministic CFG actually samples.

3.2 PATH-INTEGRAL CHARACTERIZATION OF CFG

Theorem 3.2 (CFG distribution under deterministic sampling)**.** *Suppose $p, q \in C^{1,2}([t_0, T] \times \mathbb{R}^n)$ with $p_t(x) > 0$ and $q_t(x) > 0$, and that both densities satisfy the Fokker–Planck equation* (7)*.*

Assume that the characteristic ODE below generates a unique C^1 flow and that all path integrals appearing in the statement are finite. Let $\hat{r}_{t_0}$ be the distribution obtained by evolving the terminal density $\tilde{r}_T$ backward to t_0 via the deterministic ODE $dx/dt = (f - \frac{1}{2}g^2\nabla_x \log \tilde{r}_t)$. Then

$$\hat{r}_{t_0}(x) = \frac{p_{t_0}^{\omega}(x)\, q_{t_0}^{1-\omega}(x)}{Z'} \exp\left(-\frac{\omega(\omega-1)}{2} \int_{t_0}^{T} g^2(s) \left\| \nabla_x \log p_s(X_s^x) - \nabla_x \log q_s(X_s^x) \right\|_2^2 ds \right), \tag{11}$$

where X_s^x solves the characteristic ODE

$$\frac{dX_s^x}{ds} = f(X_s^x, s) - \frac{1}{2} g^2(s)\, \nabla_x \log \tilde{r}_s(X_s^x), \qquad X_{t_0}^x = x, \tag{12}$$

and Z' is a normalization constant.

Under the usual terminal condition $\log p_T = \log q_T$, the distribution can be equivalently written in terms of the log-ratio $L_t(x) = \log p_t(x) - \log q_t(x)$ as

$$\hat{r}_{t_0}(x) = \frac{1}{Z'}\, p_{t_0}(x)\, \exp\left(-\frac{1}{2}(\omega-1) \int_{t_0}^{T} g^2(s) \left(\Delta L_s(X_s^x) + \nabla L_s(X_s^x) \cdot \nabla \log p_s(X_s^x) \right) ds \right). \tag{13}$$

Proof. See Appendix B.1. □

Theorem 3.2 gives an exact analytic representation of the CFG-induced distribution at the stopping time t_0. The first form (11) shows that the naïve product distribution is corrected by an exponential path integral of the score discrepancy along the characteristic trajectory. For $\omega > 1$, trajectories that pass through regions where the target and reference scores disagree strongly are penalized by this factor. Under the usual terminal condition $\log p_T = \log q_T$, the equivalent log-ratio form (13) provides a complementary view: the CFG-induced distribution is the target distribution p_{t_0} reweighted by an exponential path-integral correction whose contribution is controlled by $\omega - 1$.

Several consequences follow. First, deterministic CFG does *not* generally sample from the normalized product distribution $\tilde{r}_{t_0} \propto p_{t_0}^{\omega} q_{t_0}^{1-\omega}$; an additional path-integral correction is required. This correction vanishes in the limiting cases $\omega \in \{0, 1\}$, and it becomes distributionally irrelevant in degenerate settings where the score discrepancy is spatially constant, so that its contribution can be absorbed into the normalization. Second, under the usual terminal condition, Eq. (13) expresses the induced distribution as p_{t_0} reweighted by an exponential path-integral correction. This form separates the target distribution from the cumulative guidance effect and motivates the schedule design in Section 4.

3.3 Time-Dependent Guidance Extension

Theorem 3.3 (CFG with time-dependent guidance)**.** *Suppose $p, q \in C^{1,2}([t_0, T] \times \mathbb{R}^n)$ with $p_t(x) > 0$ and $q_t(x) > 0$, and that both densities satisfy the Fokker–Planck equation* (7)*. Let $\omega \in W^{1,1}([t_0, T])$, define $r_t(x) = p_t^{\omega(t)}(x) q_t^{1-\omega(t)}(x)$ and $Z_t = \int_{\mathbb{R}^n} r_t(x)\, dx$, and set $\tilde{r}_t = Z_t^{-1} r_t$, where $0 < Z_t < \infty$ on $[t_0, T]$. Assume that the characteristic ODE below generates a unique C^1 flow and that all path integrals appearing in the statement are finite. Let $L_t(x) = \log p_t(x) - \log q_t(x)$, and let X_s^x solve*

$$\frac{dX_s^x}{ds} = f(X_s^x, s) - \frac{1}{2} g^2(s)\, \nabla_x \log \tilde{r}_s(X_s^x), \qquad X_{t_0}^x = x. \tag{14}$$

If $\log p_T = \log q_T$ or $\omega(T) = 1$, the distribution obtained by evolving the terminal density $\tilde{r}_T$ backward to t_0 via the deterministic ODE is

$$\hat{r}_{t_0}(x) = \frac{p_{t_0}(x)}{Z'} \exp\left(-\int_{t_0}^{T} \frac{g^2(s)}{2} (\omega(s) - 1) \left(\Delta L_s(X_s^x) + \nabla L_s(X_s^x) \cdot \nabla \log p_s(X_s^x) \right) ds \right), \tag{15}$$

where Z' is a normalization constant.

Proof. See Appendix B.2. □

For general terminal marginals, the right-hand side of (15) includes the additional factor $\exp((\omega(T)-1)L_T(X_T^x))$, with Z' adjusted accordingly; for constant guidance, the corresponding factor in (13) is $\exp((\omega-1)L_T(X_T^x))$. Thus, a schedule satisfying $\omega(T)=1$ removes this boundary term even when $p_T \neq q_T$, under the same initialization from $\tilde{r}_T = p_T$.

Equation (15) shows that time-dependent guidance has the same structural form as constant guidance: the induced distribution is represented as p_{t_0} reweighted by an exponential path-integral correction. The schedule-dependent factor $\omega(t)-1$ controls how strongly each timestep contributes to this correction.

At each fixed noise level s, the operator $\mathcal{L}_{p_s} = \Delta + \nabla \log p_s \cdot \nabla$ is the infinitesimal generator of the overdamped Langevin diffusion with invariant density p_s. Thus, $\mathcal{L}_{p_s} L_s$ represents the instantaneous expected rate of change of L_s under these frozen-time dynamics. In Eq. (15), the log-density correction, up to normalization, is the negative integral of this rate evaluated along the deterministic CFG trajectory, weighted by $g^2(s)(\omega(s)-1)/2$.

Constant guidance is recovered as the special case $\omega(t) \equiv \bar{\omega}$, for which (15) reduces to Theorem 3.2. This observation places standard CFG within a broader time-varying framework. As we shall argue in Section 4 and validate experimentally in Section 5, constant guidance can be suboptimal because the path-integral weight is also modulated by the noise schedule.

Remark 3.4. The preceding analysis does *not* rely on specific choices of $f(x,t)$, $g(t)$, or the data distributions p_{data} and q_{data}. The derivation uses only the forward SDE (2) and a deterministic guided reverse process of the form

$$dx = \Big(f(x,t) - \frac{1}{2} g^2(t) \left(\omega(t) \nabla_x \log p_t + (1-\omega(t)) \nabla_x \log q_t \right) \Big)\, dt. \tag{16}$$

In the usual CFG setting, q_{data} is the unconditional distribution and $p_{\text{data}} = q_{\text{data}}(\cdot \mid c)$ is the conditional distribution for condition c. The form (16) therefore covers deterministic guided samplers with possibly time-dependent guidance strength.

This paper focuses on the variance-preserving SDE (VPSDE) (Song et al., 2021a), with $f(x,t) = -\frac{1}{2}\beta(t)x$ and $g(t) = \sqrt{\beta(t)}$, where $\beta(t)$ is the continuous variance rate of the forward process. The schedule design (Section 4) and experiments (Section 5) use this framework.

4 Distribution-Guided Classifier-Free Guidance

We now develop DG-CFG using the explicit temporal coefficient identified by the path-integral characterization. Under VPSDE, $g^2(s) = \beta(s)$, and the time-dependent analysis shows that the schedule-dependent contribution to the path-integral exponent is

$$\frac{1}{2}\beta(s)\,(\omega(s)-1)\left(\Delta L_s(X_s^x) + \nabla L_s(X_s^x) \cdot \nabla \log p_s(X_s^x) \right). \tag{17}$$

4.1 Motivation: Temporal Guidance Allocation.

With constant guidance $\omega(s) \equiv \bar{\omega}$, the schedule-dependent weight is $\beta(s)(\bar{\omega}-1)$, so an increasing noise rate assigns larger temporal coefficients to high-noise states. A common example is the linear VPSDE rate $\beta(s) = \beta_{\min} + s(\beta_{\max}-\beta_{\min})/T$, with $\beta_{\min} = 0.1$ and $\beta_{\max} = 20$, whose endpoint rates differ by a factor of 200 on $[0,T]$. DG-CFG compensates for the chosen noise rate, whether linear or nonlinear, while keeping the integrated guidance budget fixed.

4.2 Distribution-Guided Classifier-Free Guidance (DG-CFG).

The path-integral characterization identifies $\beta(t)(\omega(t)-1)$ as an explicit temporal coefficient, motivating inverse-rate time balancing. DG-CFG combines this principle with signal-content weighting and error-motivated low-noise attenuation. The individual and complementary effects of these factors are evaluated in the ablation study (Section 5.2.4).

Step 1: Time balancing. We compensate for the diffusion-rate factor by scaling the guidance deviation inversely with $\beta(t)$: $\omega(t) = 1 + C_1 \cdot (\bar{\omega}-1)/\beta(t)$. This removes the diffusion-rate dependence from the explicit coefficient, making $\beta(t)(\omega(t)-1)$ constant in continuous time.

Table 1: Compact toy-model verification over $\bar{\omega} \in \{3, 5, 7, 9\}$. The MAE range summarizes agreement between empirical and theoretical label distributions; the Min-Max-ratio measures balance among the four preferred modes.

Schedule	MAE range↓	Min-Max ($\bar{\omega} = 3$)↑	Min-Max ($\bar{\omega} = 9$)↑
Constant CFG	0.00234–0.00357	0.665	0.099
Time-dependent CFG	0.00112–0.00219	0.959	0.808

Step 2: Signal-content weighting. We further account for signal content through the forward-process coefficient $\sqrt{\bar{\alpha}_t}$. Under $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sigma_t \boldsymbol{\epsilon}$, the factor $\sqrt{\bar{\alpha}_t} = \exp\Big(-\frac{1}{2}\int_0^t \beta(s)\,ds\Big)$ scales the retained data signal: it approaches zero near the pure-noise endpoint $t = T$ and one near the clean endpoint $t = t_0$. Multiplying by this factor de-emphasizes nearly pure-noise states and shifts guidance toward timesteps with greater retained signal content, giving $\omega(t) = 1 + C_2 \cdot (\bar{\omega} - 1) \cdot \sqrt{\bar{\alpha}_t}/\beta(t)$.

Step 3: Low-noise attenuation. Signal-content weighting increases the relative influence of the low-noise regime. Under the noise-prediction parameterization, a prediction error $\delta\boldsymbol{\epsilon}_t$ corresponds to a score-estimation error $\delta\mathbf{s}_t = -\delta\boldsymbol{\epsilon}_t/\sigma_t$. The two score factors in the trajectory-integral term $\nabla L_t \cdot \nabla \log p_t$ introduce a $1/\sigma_t^2$ scaling, motivating multiplication of the schedule by $\sigma_t^2 = 1 - \bar{\alpha}_t$ to attenuate guidance near the low-noise endpoint.

Final schedule. Combining all three factors yields Distribution-Guided Classifier-Free Guidance (DG-CFG):

$$\omega(t) = 1 + C \cdot (\bar{\omega} - 1) \cdot \frac{(1 - \bar{\alpha}_t)\sqrt{\bar{\alpha}_t}}{\beta(t)}, \tag{18}$$

where $\bar{\omega} \geq 1$ is the nominal guidance strength and $C = (T - t_0)\big/\int_{t_0}^{T}\big((1 - \bar{\alpha}_s)\sqrt{\bar{\alpha}_s}/\beta(s)\big)\,ds$ normalizes the integrated guidance deviation to match constant CFG. This normalization changes the temporal allocation of guidance while preserving its total integrated budget, enabling a fair comparison with constant CFG.

The factor $(1 - \bar{\alpha}_t)\sqrt{\bar{\alpha}_t}$ vanishes in the clean- and pure-noise limits, concentrating guidance at intermediate noise levels where meaningful signal and reliable score estimates coexist. At the high-noise endpoint, $\sqrt{\bar{\alpha}_T} \approx 0$ yields $\omega(T) \approx 1$, attenuating the terminal correction in Theorem 3.3. This middle-focused profile combines path-integral time balancing with signal-content and score-reliability considerations, and agrees with prior empirical observations (Kynkäänniemi et al., 2024; Malarz et al., 2025). Section 5.2.4 evaluates the individual and complementary effects of these factors.

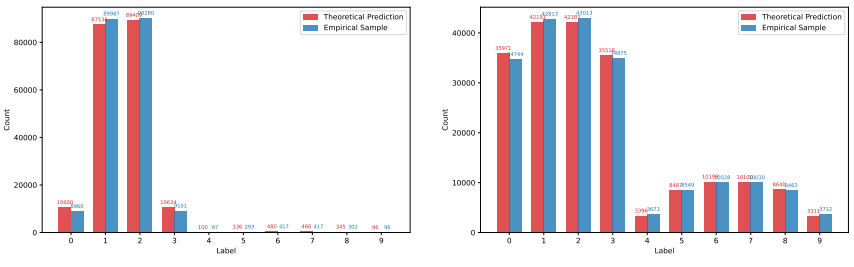
5 Experiments

We validate the path-integral analysis in a toy model with analytic scores and evaluate DG-CFG on Stable Diffusion models.

5.1 Discrete Distribution on a Circle

This experiment aims to validate the preceding theoretical derivations in a controlled model with analytic scores and distributions. The unconditional distribution is uniform over ten equally spaced points on a circle, whereas the conditional distribution doubles the mass of four adjacent points. We compare constant CFG with a time-dependent schedule proportional to $\sqrt{\bar{\alpha}_t}/\beta(t)$ over $\bar{\omega} \in \{3, 5, 7, 9\}$, and compare the empirical outcomes with the path-integral predictions. Complete distribution definitions, schedules, sampling details, and metric definitions are provided in Appendix C.

Table 1 and Figure 1 show close agreement between empirical and predicted label distributions, with MAE ranging from 0.00112 to 0.00357 across both schedules and all guidance strengths, supporting the path-integral predictions.The time-dependent schedule also preserves substantially more balanced coverage shown by Min-Max-ratio. The mode-preservation comparison serves as a simple heuristic illustration; detailed comparative evaluation is provided in the Stable Diffusion experiments below (Section 5.2).



Constant CFG Time-dependent CFG

Figure 1: Toy-model label distributions at strong guidance ($\bar{\omega} = 9$). Bars compare the theoretical target distribution with empirical CFG samples. The complete four-panel comparison is provided in Appendix C.

Table 2: Representative main results with DDIM sampling at 8 NFE. At $\bar{\omega} = 7$, Div and FID summarize the diversity–fidelity trade-off; at $\bar{\omega} = 23$, CLIP, IR, Sat, and FID measure stability under strong guidance. Complete sweeps are reported in Appendix D.2.

		$\bar{\omega} = 7$		$\bar{\omega} = 23$			
Backbone	Method	Div↑	FID↓	CLIP↑	IR↑	Sat	FID↓
SD1.5	Constant CFG	5.40	**67.67**	30.64	−0.330	0.469	82.34
	Interval CFG	**5.95**	70.83	29.10	−0.986	0.462	110.89
	β–CFG	5.44	68.14	30.88	−0.239	0.438	85.38
	DG-CFG (ours)	5.71	67.87	**31.39**	**+0.029**	0.332	**74.53**
SD2.1	Constant CFG	5.01	**67.45**	30.98	−0.092	0.431	79.94
	Interval CFG	**5.49**	72.60	29.80	−0.630	0.422	105.19
	β–CFG	4.99	68.64	31.14	+0.060	0.398	83.11
	DG-CFG (ours)	5.20	68.70	**31.43**	**+0.244**	0.315	**76.43**
SDXL	Constant CFG	4.84	72.86	29.46	−0.717	0.351	105.39
	Interval CFG	**5.62**	79.47	28.17	−1.144	0.326	126.27
	β–CFG	4.75	**72.29**	30.76	−0.222	0.318	92.56
	DG-CFG (ours)	5.00	76.79	**32.00**	**+0.401**	0.241	**75.02**

5.2 Stable Diffusion Models

5.2.1 Setup

We evaluate Stable Diffusion 1.5 (SD1.5), Stable Diffusion 2.1 (SD2.1), and Stable Diffusion XL (SDXL) at their native resolutions, using DDIM (Song et al., 2021a) for the main comparison, ablations, and sampling-budget studies, and Euler for the additional SDXL comparison. Prompts come from MS-COCO 2017 validation (Lin et al., 2014), with empty unconditional prompts. We measure fidelity, alignment, perceptual quality, saturation, and diversity using FID (Heusel et al., 2017), CLIP (Radford et al., 2021), ImageReward (IR), Sat, and Vendi score (Friedman & Dieng, 2023), respectively. Checkpoints and evaluation details are provided in Appendix D.1.

5.2.2 Main Results

We compare Constant CFG, Interval CFG (Kynkäänniemi et al., 2024), β–CFG (Malarz et al., 2025) (using only its temporal scheduling component for consistency across methods), and DG-CFG using the same DDIM sampler with 8 NFE and $\bar{\omega} \in \{3, 7, 11, 15, 19, 23\}$. All non-constant schedules are normalized so that $\int_{t_0}^{T} (\omega(s) - 1)\, ds = (\bar{\omega} - 1)(T - t_0)$, ensuring an equal total guidance budget. Table 2 reports representative moderate- and strong-guidance settings; complete schedule definitions and strength sweeps are provided in Appendix D.2.

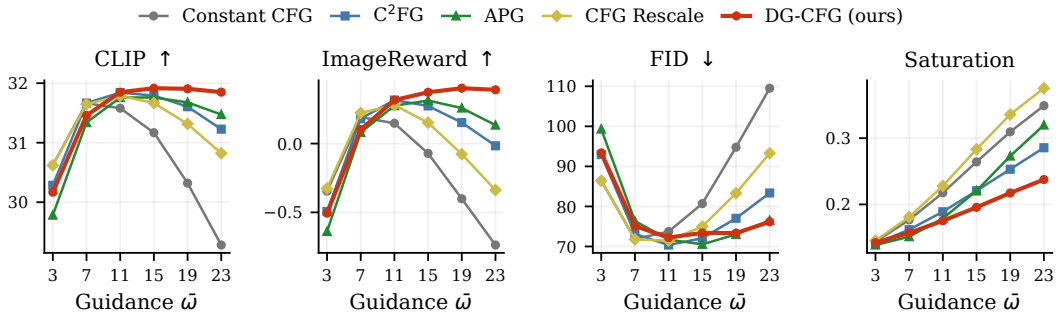


Figure 2: Guidance-strength comparison on SDXL with 8-step Euler sampling over 1K MS-COCO validation prompts, reporting CLIP, ImageReward, FID, and saturation. All methods use matched prompts and seeds.

At moderate guidance ($\bar{\omega} = 7$), DG-CFG increases diversity over Constant CFG and β–CFG on all three backbones while retaining lower FID than the more diversity-oriented Interval CFG. Under strong guidance ($\bar{\omega} = 23$), DG-CFG achieves the highest CLIP and ImageReward and the lowest FID on every backbone, while substantially suppressing the saturation increase observed with the baselines. The complete sweeps in Appendix D.2 show that this stability persists across the evaluated guidance range. Besides, qualitative comparisons across all three backbones in Appendix D.2 show that DG-CFG preserves more natural colors and coherent detail under strong guidance.

5.2.3 Comparison with Additional Guidance Methods

We broaden the comparison to C^{2}FG (Gao et al., 2026), APG (Sadat et al., 2025), and CFG Rescale (Lin et al., 2024), covering both guidance scheduling and prediction modification. Using its default EulerDiscreteScheduler for 8 steps, we evaluate these methods alongside Constant CFG and DG-CFG over the same nominal guidance strengths $\bar{\omega} \in \{3, 7, 11, 15, 19, 23\}$, with matched prompts and seeds across methods. Appendix D.3 provides the settings and complete results.

Figure 2 shows that DG-CFG maintains strong prompt alignment and perceptual quality as guidance increases, achieving the highest CLIP and ImageReward at $\bar{\omega} \in \{15, 19, 23\}$ with comparable FID and controlled saturation. This extends the strong-guidance trend observed with DDIM to the Euler setting.

5.2.4 Ablation Study

We ablate the three factors in the proposed schedule at 8 NFE on SD1.5 and SD2.1. Starting from constant CFG, Medium 1 introduces time balancing by $1/\beta(t)$, and Medium 2 additionally applies signal-content weighting by $\sqrt{\bar{\alpha}_t}$. Medium 3 combines time balancing with low-noise attenuation, while Medium 4 combines signal-content weighting with low-noise attenuation. DG-CFG includes all three factors. Each non-constant schedule is independently normalized to match the integrated guidance deviation of constant CFG.

Table 3 shows a consistent component-wise pattern. Time balancing improves FID and controls saturation under strong guidance; adding signal-content weighting suppresses saturation further but incurs a quality cost; low-noise attenuation then recovers perceptual and distributional quality. Medium 2–4 further show that removing any one of DG-CFG's three components weakens the joint trade-off among IR, saturation, and FID on both backbones. DG-CFG consequently achieves the highest IR with controlled saturation. The complete schedules, all guidance strengths, and 16-sample qualitative comparisons are reported in Appendix D.4.

5.2.5 Robustness Across Sampling Budgets

We evaluate whether DG-CFG's advantage is robust to the sampling budget. On SD1.5 and SD2.1, we fix the strong-guidance setting $\bar{\omega} = 23$ and vary DDIM sampling over $\mathrm{NFE} \in \{4, 8, 16, 32\}$. Across the eight backbone–NFE combinations, DG-CFG achieves the best ImageReward with controlled

Table 3: Component ablation at $\bar{\omega} = 23$ and 8 NFE. Bold denotes the best IR and FID; Sat is reported to assess saturation control under strong guidance. The evaluation protocol and complete results are in Appendices D.1 and D.4.

	SD1.5			SD2.1		
Method	IR↑	Sat	FID↓	IR↑	Sat	FID↓
Constant CFG	−0.330	0.469	82.34	−0.092	0.431	79.94
Medium 1	−0.000	0.305	**73.24**	+0.172	0.290	**74.93**
Medium 2	−0.283	0.259	87.65	−0.158	0.250	91.22
Medium 3	−0.259	0.455	82.68	+0.009	0.417	80.73
Medium 4	−0.141	0.420	82.00	+0.110	0.385	81.08
DG-CFG	**+0.029**	0.332	74.53	**+0.244**	0.315	76.43

saturation in all eight settings, and the best FID and CLIP score in seven. These results show that its advantage is not specific to the 8-NFE configuration used in the main comparison. Table 4 summarizes the results; complete five-metric tables and qualitative comparisons are provided in Appendix D.5.

Table 4: Robustness across sampling budgets on Stable Diffusion 1.5 and Stable Diffusion 2.1 at $\bar{\omega} = 23$. Gains are measured against the strongest baseline for each metric ($\Delta\mathrm{IR} = \mathrm{IR}_{\mathrm{ours}} - \mathrm{IR}_{\mathrm{baseline}}$, $\Delta\mathrm{FID} = \mathrm{FID}_{\mathrm{baseline}} - \mathrm{FID}_{\mathrm{ours}}$); positive values favor DG-CFG. Sat (reference) is the lowest saturation among the three non-DG-CFG methods at the same NFE.

Model	NFE	ΔIR↑	ΔFID↑	Sat (DG-CFG)	Sat (reference)
SD1.5	4	+0.304	+3.62	0.348	0.408
	8	+0.268	+7.81	0.332	0.438
	16	+0.042	+2.40	0.341	0.413
	32	+0.039	+2.84	0.356	0.395
SD2.1	4	+0.230	−4.16	0.322	0.369
	8	+0.184	+3.51	0.315	0.398
	16	+0.072	+0.19	0.327	0.381
	32	+0.020	+0.22	0.340	0.366

6 Conclusion

We introduced DG-CFG, a training-free guidance schedule that redistributes a fixed guidance budget toward informative and reliable stages of denoising. Its design combines path-integral time balancing with signal-content weighting and error-motivated low-noise attenuation, yielding a middle-focused guidance profile without modifying the diffusion model.

Across Stable Diffusion 1.5, Stable Diffusion 2.1, and Stable Diffusion XL, DG-CFG offers a favorable diversity–fidelity trade-off at moderate guidance and improves image quality while controlling saturation under strong guidance. Component ablations support the combined schedule design, and NFE sweeps show that these gains persist across sampling budgets.

Limitations and Future Work. DG-CFG addresses noise-schedule-induced temporal non-uniformity but does not explicitly model the geometry of the conditional–reference score discrepancy. Incorporating this geometry and extending the analysis to flow matching models are promising directions.

AI use statement

In this work, we used generative AI tools for assisting in the writing of proofs, proposing or refining hypotheses, implementing methods, assisting with translation and supporting qualitative and thematic data analysis. We have not used generative AI tools for helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, designing or providing feedback on research methodology or experiments and

interpreting results. Generating synthetic data sets and cleaning or reformatting datasets are not applicable to this work. Additionally, we used generative AI tools for creating or modifying scientific figures or images, drafting parts of a research paper, summarizing or analysing existing literature and editing a research paper to improve readability. We have reviewed all AI-assisted work. All AI-assisted text, code, and figures were manually reviewed and verified by the authors for correctness and consistency with the reported results. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Reproducibility statement

To support reproducibility, Section 4 specifies the complete DG-CFG schedule and its normalization, and Section 5 describes the main experimental setup. Appendix B provides complete proofs of the theoretical results, Appendix C gives the analytic toy-model formulas and sampling protocol, and Appendix D documents the Stable Diffusion checkpoints, generation and evaluation protocols, baseline and ablation settings, the Euler comparison, NFE experiments, and complete results.

References

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

Yogesh Balaji, Seungjun Nah, Xun Huang, Arash Vahdat, Jiaming Song, Qinsheng Zhang, Karsten Kreis, Miika Aittala, Timo Aila, Samuli Laine, et al. ediff-i: Text-to-image diffusion models with an ensemble of expert denoisers. *arXiv preprint arXiv:2211.01324*, 2022.

Arwen Bradley and Preetum Nakkiran. Classifier-free guidance is a predictor-corrector. *arXiv preprint arXiv:2408.09000*, 2024.

Hyungjin Chung, Jeongsol Kim, Geon Yeong Park, Hyelin Nam, and Jong Chul Ye. Cfg++: Manifold-constrained classifier free guidance for diffusion models. In *International Conference on Learning Representations*, volume 2025, pp. 30824–30850, 2025.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Dan Friedman and Adji Bousso Dieng. The vendi score: A diversity evaluation metric for machine learning. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=g97OHbQyk1.

Jiayang Gao, Tianyi Zheng, Jiayang Zou, Fengxiang Yang, Shice Liu, Luyao Fan, Zheyu Zhang, Hao Zhang, Jinwei Chen, Peng-Tao Jiang, Bo Li, and Jia Wang. C^2FG: Control classifier-free guidance via score discrepancy analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 34398–34407, 2026.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021. URL https://openreview.net/forum?id=qw8AKxfYbI.

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

Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *Advances in neural information processing systems*, 35:8633–8646, 2022.

Cheng Jin, Qitan Shi, and Yuantao Gu. Stage-wise dynamics of classifier-free guidance in diffusion models. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=fP0s1TEow3.

Tero Karras, Miika Aittala, Tuomas Kynkäänniemi, Jaakko Lehtinen, Timo Aila, and Samuli Laine. Guiding a diffusion model with a bad version of itself. *Advances in Neural Information Processing Systems*, 37:52996–53021, 2024.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=a-xFK8Ymz5J.

Tuomas Kynkäänniemi, Miika Aittala, Tero Karras, Samuli Laine, Timo Aila, and Jaakko Lehtinen. Applying guidance in a limited interval improves sample and distribution quality in diffusion models. *Advances in Neural Information Processing Systems*, 37:122458–122483, 2024.

Krunoslav Lehman Pavasovic, Jakob Verbeek, Giulio Biroli, and Marc Mezard. Understanding classifier-free guidance: High-dimensional theory and non-linear generalizations. *arXiv e-prints*, pp. arXiv–2502, 2025.

Shanchuan Lin, Bingchen Liu, Jiashi Li, and Xiao Yang. Common diffusion noise schedules and sample steps are flawed. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5392–5399. IEEE, 2024.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pp. 740–755. Springer, 2014.

Haohe Liu, Zehua Chen, Yi Yuan, Xinhao Mei, Xubo Liu, Danilo Mandic, Wenwu Wang, and Mark D Plumbley. Audioldm: Text-to-audio generation with latent diffusion models. *arXiv preprint arXiv:2301.12503*, 2023.

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.

Dawid Malarz, Artur Kasymov, Maciej Zię ba, Jacek Tabor, and Przemysław Spurek. Classifier-free guidance with adaptive scaling. In *ECAI 2025: 28th European Conference on Artificial Intelligence 25-30 October 2025, Bologna, Italy-Including 14th Conference on Prestigious Applications of Intelligent Systems (PAIS 2025) Proceedings*, pp. 435–442. SAGE Publications 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP, 2025.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PMLR, 2021.

Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=FjNys5c7VyY.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 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*, pp. 10684–10695, 2022.

Seyedmorteza Sadat, Otmar Hilliges, and Romann M. Weber. Eliminating oversaturation and artifacts of high guidance scales in diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=e2ONKX6qzJ.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, 35:36479–36494, 2022.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=TIdIXIpzhoI.

Marta Skreta, Tara Akhound-Sadegh, Viktor Ohanesian, Roberto Bondesan, Alan Aspuru-Guzik, Arnaud Doucet, Rob Brekelmans, Alexander Tong, and Kirill Neklyudov. Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=Vhc0KrcqWu.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. pmlr, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a. URL https://openreview.net/forum?id=St1giarCHLP.

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*, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Arash Vahdat, Karsten Kreis, and Jan Kautz. Score-based generative modeling in latent space. *Advances in neural information processing systems*, 34:11287–11302, 2021.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=PzcvxEMzvQC.

A Related Work

A.1 Diffusion Models

Diffusion models generate data by reversing a progressive noising process (Sohl-Dickstein et al., 2015; Ho et al., 2020). Song et al. (2021b) unified diffusion probabilistic models and score-based generative models through continuous-time SDEs and derived the probability flow ODE, which follows deterministic dynamics while preserving the same time marginals. Closely related samplers such as DDIM enable efficient deterministic generation (Song et al., 2021a). Subsequent advances have improved sample quality (Dhariwal & Nichol, 2021; Nichol & Dhariwal, 2021), accelerated inference (Salimans & Ho, 2022; Lu et al., 2022), and extended diffusion models to latent spaces (Rombach et al., 2022; Vahdat et al., 2021).

A.2 Classifier-Free Guidance and Variants

Ho & Salimans (2021) introduced classifier-free guidance (CFG) as an alternative to classifier-based guidance (Dhariwal & Nichol, 2021). CFG learns conditional and unconditional score estimates within a single model, eliminating the need for an external classifier. It is widely used in text-to-image systems (Rombach et al., 2022; Saharia et al., 2022; Ramesh et al., 2022; Balaji et al., 2022), with extensions to video (Ho et al., 2022), audio (Liu et al., 2023), and 3D synthesis (Poole et al., 2023). Several variants address its failure modes: CFG++ applies guidance to the denoised estimate while using the unconditional prediction for re-noising (Chung et al., 2025); autoguidance replaces the unconditional reference with a weaker version of the model (Karras et al., 2024); and adaptive projected guidance decomposes the CFG update relative to the conditional prediction and down-weights its parallel component (Sadat et al., 2025). Dynamic thresholding (Saharia et al.,

2022) mitigates saturation by clipping and rescaling predicted clean images using a percentile-based threshold. CFG rescaling (Lin et al., 2024) adjusts guided predictions using the ratio of conditional to guided standard deviations to reduce overexposure. Dynamic thresholding and CFG rescaling operate in the spatial or amplitude domain, clipping pixel values or renormalizing prediction magnitudes within each sampling step. DG-CFG operates in the temporal domain, redistributing a fixed guidance budget across denoising stages along the PF-ODE trajectory. Section 5.2.3 compares DG-CFG with APG and CFG Rescale.

A.3 Theoretical Understanding of CFG

Recent work has analyzed CFG from several complementary perspectives. Bradley & Nakkiran (2024) showed that CFG does not directly sample a gamma-powered product distribution under DDPM or DDIM and interpreted its SDE limit as a predictor-corrector procedure. Lehman Pavasovic et al. (2025) established recovery of the target conditional distribution in an infinite-dimensional limit and characterized finite-dimensional mean and variance corrections. Jin et al. (2026) identified direction-shift, mode-separation, and concentration stages under multimodal conditionals. Skreta et al. (2025) introduced Feynman–Kac correctors for product-distribution sampling.

A.4 Guidance Schedule Design

Time-dependent CFG schedules redistribute guidance across the sampling trajectory. Interval guidance applies CFG only within a selected noise range, motivated by its limited utility at the highest and lowest noise levels (Kynkäänniemi et al., 2024). β–CFG instead uses a beta-shaped schedule that suppresses guidance near both endpoints (Malarz et al., 2025). C^{2}FG derives bounds on the conditional–unconditional score discrepancy and uses an exponential-decay control function (Gao et al., 2026); we include it in the comparison in Section 5.2.3.

B Proofs of Main Results

This appendix gives the proofs of the two main theorems in Section 3. The proof of Theorem 3.2 begins with an auxiliary lemma characterizing the evolution of the normalized product reference density $\tilde{r}_t$; the proof of Theorem 3.3 then reuses this lemma for time-dependent guidance.

B.1 Proof of Theorem 3.2 (Exact CFG Sampling Distribution)

Integrability under Gaussian perturbations. Suppose both initial distributions are supported in $\{x : \|x\| \le R\}$ and the forward kernel is $X_t = a_t X_0 + \sigma_t \epsilon$, with $\epsilon \sim \mathcal{N}(0, I)$ and $\sigma_t > 0$. Fix $t > 0$, and write $M_t = |a_t| R$ and $c_t = (2\pi\sigma_t^2)^{-n/2}$. For either marginal $h_t \in \{p_t, q_t\}$, expanding the Gaussian kernel gives

$$c_t e^{-\frac{\|x\|^2}{2\sigma_t^2} - \frac{M_t\|x\|}{\sigma_t^2} - \frac{M_t^2}{2\sigma_t^2}} \le h_t(x) \le c_t e^{-\frac{\|x\|^2}{2\sigma_t^2} + \frac{M_t\|x\|}{\sigma_t^2}}. \tag{19}$$

Hence $\log h_t(x) = -\|x\|^2/(2\sigma_t^2) + O(1 + \|x\|)$, so for every finite $\omega(t)$,

$$r_t(x) = p_t(x)^{\omega(t)} q_t(x)^{1-\omega(t)} = \exp\left(-\frac{\|x\|^2}{2\sigma_t^2} + O(1 + \|x\|)\right). \tag{20}$$

Thus $0 < Z_t < \infty$, and $\tilde{r}_t$ has finite moments of every order. Differentiating the Gaussian convolution also yields

$$\nabla \log h_t(x) = -\frac{x - m_{h,t}(x)}{\sigma_t^2}, \qquad m_{h,t}(x) = \mathbb{E}_h[a_t X_0 \mid X_t = x], \qquad \|m_{h,t}(x)\| \le M_t. \tag{21}$$

Consequently, $\|\nabla \log p_t - \nabla \log q_t\| \le 2M_t/\sigma_t^2$ and $|\log(p_t/q_t)| = O(1 + \|x\|)$, ensuring $\mathbb{E}_{\tilde{r}_t}[|V_t|] < \infty$ and $\mathbb{E}_{\tilde{r}_t}[|\log(p_t/q_t)|] < \infty$. These bounds hold uniformly on $[t_0, T]$ when a_t, σ_t are continuous, $\inf_{[t_0,T]} \sigma_t > 0$, and $\omega(t)$ is bounded, as for standard VP and VE perturbations of boundedly supported data with $t_0 > 0$.

We now establish the auxiliary evolution equation used in the density-ratio argument.

Lemma B.1 (Fokker–Planck equation for $\tilde{r}_t$)**.** *Fix $0 < t_0 < T$ and $\omega \in \mathbb{R}$. Suppose $p, q \in C^{1,2}([t_0, T] \times \mathbb{R}^n)$ with $p_t(x) > 0$ and $q_t(x) > 0$, and that both densities satisfy the Fokker–Planck equation (7). Let $r_t = p_t^\omega q_t^{1-\omega}$ and $Z_t = \int_{\mathbb{R}^n} r_t(x)\,dx$, where $0 < Z_t < \infty$ on $[t_0, T]$. For $B_R = \{x : \|x\| \le R\}$, assume the boundary decay condition $\lim_{R\to\infty} \int_{\partial B_R} \big(\|f(x,t)\|\, r_t(x) + \|\nabla r_t(x)\|\big)\, dS(x) = 0$ for every $t \in [t_0, T]$. Then the normalized product distribution $\tilde{r}_t = Z_t^{-1} p_t^\omega q_t^{1-\omega}$ satisfies*

$$\frac{\partial \tilde{r}_t}{\partial t} = -\nabla \cdot (f \tilde{r}_t) + \frac{1}{2} g^2 \Delta \tilde{r}_t - \frac{1}{2} g^2 \big(V_t - \mathbb{E}_{\tilde{r}_t}[V_t]\big)\tilde{r}_t, \tag{22}$$

where the interaction potential is

$$V_t(x) = \omega(\omega - 1)\, \big\|\nabla_x \log p_t(x) - \nabla_x \log q_t(x)\big\|_2^2. \tag{23}$$

Proof. We start from the forward Fokker–Planck equations satisfied by the two marginals:

$$\frac{\partial p_t}{\partial t} = -\nabla \cdot (f p_t) + \frac{1}{2} g^2 \Delta p_t, \tag{24}$$

$$\frac{\partial q_t}{\partial t} = -\nabla \cdot (f q_t) + \frac{1}{2} g^2 \Delta q_t. \tag{25}$$

Step 1: Evolution of the unnormalized product. Let $r_t = p_t^\omega q_t^{1-\omega}$. Applying the chain rule gives

$$\begin{aligned}\frac{\partial r_t}{\partial t} &= \omega p_t^{\omega-1} q_t^{1-\omega} \frac{\partial p_t}{\partial t} + (1-\omega) p_t^\omega q_t^{-\omega} \frac{\partial q_t}{\partial t} \\ &= \omega p_t^{\omega-1} q_t^{1-\omega} \Big[-\nabla \cdot (f p_t) + \frac{1}{2} g^2 \Delta p_t \Big] + (1-\omega) p_t^\omega q_t^{-\omega} \Big[-\nabla \cdot (f q_t) + \frac{1}{2} g^2 \Delta q_t \Big]. \end{aligned} \tag{26}$$

Drift terms. Expanding the divergence terms, we obtain

$$\begin{aligned}&\omega p_t^{\omega-1} q_t^{1-\omega} \big[-\nabla \cdot (f p_t) \big] + (1-\omega) p_t^\omega q_t^{-\omega} \big[-\nabla \cdot (f q_t) \big] \\ &= -\omega p_t^{\omega-1} q_t^{1-\omega} \big[(\nabla \cdot f) p_t + f \cdot \nabla p_t \big] - (1-\omega) p_t^\omega q_t^{-\omega} \big[(\nabla \cdot f) q_t + f \cdot \nabla q_t \big] \\ &= -(\nabla \cdot f)\, p_t^\omega q_t^{1-\omega} - f \cdot \big(\omega p_t^{\omega-1} q_t^{1-\omega} \nabla p_t + (1-\omega) p_t^\omega q_t^{-\omega} \nabla q_t \big) \\ &= -(\nabla \cdot f) r_t - f \cdot \nabla r_t = -\nabla \cdot (f r_t). \end{aligned} \tag{27}$$

Diffusion terms. Define

$$A_t = \frac{1}{2} g^2 \Big[\omega p_t^{\omega-1} q_t^{1-\omega} \Delta p_t + (1-\omega) p_t^\omega q_t^{-\omega} \Delta q_t \Big]. \tag{28}$$

The Laplacian of r_t can be decomposed as

$$\begin{aligned}\Delta r_t = {} & \omega p_t^{\omega-1} q_t^{1-\omega} \Delta p_t + (1-\omega) p_t^\omega q_t^{-\omega} \Delta q_t \\ & + \sum_{i=1}^n \Big(\frac{\partial}{\partial x_i} (\omega p_t^{\omega-1} q_t^{1-\omega}) \frac{\partial p_t}{\partial x_i} + \frac{\partial}{\partial x_i} ((1-\omega) p_t^\omega q_t^{-\omega}) \frac{\partial q_t}{\partial x_i} \Big). \end{aligned} \tag{29}$$

Therefore,

$$A_t = \frac{1}{2} g^2 \Delta r_t - \frac{1}{2} g^2 \sum_{i=1}^n \Big(\frac{\partial}{\partial x_i} (\omega p_t^{\omega-1} q_t^{1-\omega}) \frac{\partial p_t}{\partial x_i} + \frac{\partial}{\partial x_i} ((1-\omega) p_t^\omega q_t^{-\omega}) \frac{\partial q_t}{\partial x_i} \Big). \tag{30}$$

Residual term. Observe that $\omega p_t^{\omega-1} q_t^{1-\omega} = \omega (p_t/q_t)^{\omega-1}$ and $(1-\omega) p_t^\omega q_t^{-\omega} = (1-\omega)(p_t/q_t)^\omega$. Therefore,

$$\frac{\partial}{\partial x_i} \big(\omega p_t^{\omega-1} q_t^{1-\omega} \big) = \omega(\omega-1) \Big(\frac{p_t}{q_t} \Big)^{\omega-2} \frac{\partial}{\partial x_i} \Big(\frac{p_t}{q_t} \Big), \tag{31}$$

$$\frac{\partial}{\partial x_i} \big((1-\omega) p_t^\omega q_t^{-\omega} \big) = (1-\omega)\omega \Big(\frac{p_t}{q_t} \Big)^{\omega-1} \frac{\partial}{\partial x_i} \Big(\frac{p_t}{q_t} \Big) = -\omega(\omega-1) \Big(\frac{p_t}{q_t} \Big)^{\omega-1} \frac{\partial}{\partial x_i} \Big(\frac{p_t}{q_t} \Big). \tag{32}$$

The residual sum is then

$$\sum_{i=1}^{n}\Big(\frac{\partial}{\partial x_i}(\omega p_t^{\omega-1}q_t^{1-\omega})\frac{\partial p_t}{\partial x_i}+\frac{\partial}{\partial x_i}((1-\omega)p_t^{\omega}q_t^{-\omega})\frac{\partial q_t}{\partial x_i}\Big)$$
$$=\sum_{i=1}^{n}\omega(\omega-1)\Big(\frac{p_t}{q_t}\Big)^{\omega-2}\frac{\partial}{\partial x_i}\Big(\frac{p_t}{q_t}\Big)\Big(\frac{\partial p_t}{\partial x_i}-\frac{p_t}{q_t}\frac{\partial q_t}{\partial x_i}\Big). \tag{33}$$

Using $\frac{\partial}{\partial x_i}(p_t/q_t)=\frac{1}{q_t}(\frac{\partial p_t}{\partial x_i}-\frac{p_t}{q_t}\frac{\partial q_t}{\partial x_i})$,

$$\begin{aligned}\text{Residual}&=\sum_{i=1}^{n}\omega(\omega-1)\Big(\frac{p_t}{q_t}\Big)^{\omega-2}\frac{1}{q_t}\Big(\frac{\partial p_t}{\partial x_i}-\frac{p_t}{q_t}\frac{\partial q_t}{\partial x_i}\Big)^2\\&=\sum_{i=1}^{n}\omega(\omega-1)p_t^{\omega-2}q_t^{1-\omega}\Big(\frac{\partial p_t}{\partial x_i}-\frac{p_t}{q_t}\frac{\partial q_t}{\partial x_i}\Big)^2\\&=\sum_{i=1}^{n}\omega(\omega-1)p_t^{\omega}q_t^{1-\omega}\Big(\frac{1}{p_t}\frac{\partial p_t}{\partial x_i}-\frac{1}{q_t}\frac{\partial q_t}{\partial x_i}\Big)^2\\&=\omega(\omega-1)\,r_t\sum_{i=1}^{n}\big[(\nabla_x\log p_t)_i-(\nabla_x\log q_t)_i\big]^2\\&=\omega(\omega-1)\,\|\nabla_x\log p_t-\nabla_x\log q_t\|_2^2\,r_t=V_t\,r_t.\end{aligned} \tag{34}$$

Combining the drift contribution (27), the diffusion decomposition (30), and the residual identity (34) gives

$$\frac{\partial r_t}{\partial t}=-\nabla\cdot(fr_t)+\frac{1}{2}g^2\Delta r_t-\frac{1}{2}g^2V_tr_t. \tag{35}$$

Step 2: Evolution of the normalizing constant. Integrating (35) over $\mathbb{R}^n$ and applying the divergence theorem, using the assumed boundary decay $fr_t\to 0$ and $\nabla r_t\to 0$ as $|x|\to\infty$, yields

$$\frac{dZ_t}{dt}=\int_{\mathbb{R}^n}\frac{\partial r_t}{\partial t}\,dx=-\frac{1}{2}g^2\int_{\mathbb{R}^n}V_tr_t\,dx=-\frac{1}{2}g^2Z_t\,\mathbb{E}_{\tilde{r}_t}[V_t]. \tag{36}$$

Step 3: Evolution of the normalized density. Finally, since $\tilde{r}_t=r_t/Z_t$,

$$\begin{aligned}\frac{\partial\tilde{r}_t}{\partial t}&=\frac{1}{Z_t}\frac{\partial r_t}{\partial t}-\frac{r_t}{Z_t^2}\frac{dZ_t}{dt}\\&=\frac{1}{Z_t}\Big[-\nabla\cdot(fr_t)+\frac{1}{2}g^2\Delta r_t-\frac{1}{2}g^2V_tr_t\Big]-\frac{r_t}{Z_t^2}\Big[-\frac{1}{2}g^2Z_t\,\mathbb{E}_{\tilde{r}_t}[V_t]\Big]\\&=-\nabla\cdot(f\tilde{r}_t)+\frac{1}{2}g^2\Delta\tilde{r}_t-\frac{1}{2}g^2\big(V_t-\mathbb{E}_{\tilde{r}_t}[V_t]\big)\tilde{r}_t,\end{aligned} \tag{37}$$

which proves the lemma. □

Proof of Theorem 3.2. With Lemma B.1 established, we now analyze the density transported by the deterministic guided sampler. The key step is to compare this density with the reference density through the ratio $\rho_t=\hat{r}_t/\tilde{r}_t$, and then solve the resulting transport equation along characteristics.

Step 1: Continuity equation for $\hat{r}_t$. Under the deterministic ODE $dx=(f-\frac{1}{2}g^2\nabla_x\log\tilde{r}_t)\,dt$, the density $\hat{r}_t$ satisfies the continuity equation

$$\begin{aligned}\frac{\partial\hat{r}_t}{\partial t}&=-\nabla\cdot\Big[\big(f-\frac{1}{2}g^2\nabla_x\log\tilde{r}_t\big)\hat{r}_t\Big]\\&=-\nabla\cdot(f\hat{r}_t)+\frac{1}{2}g^2\,\nabla\cdot(\hat{r}_t\nabla_x\log\tilde{r}_t).\end{aligned} \tag{38}$$

Step 2: Equation for the density ratio. Applying the quotient rule to $\rho_t = \hat{r}_t/\tilde{r}_t$ gives

$$\begin{aligned}\frac{\partial \rho_t}{\partial t} &= \frac{1}{\tilde{r}_t}\frac{\partial \hat{r}_t}{\partial t} - \frac{\hat{r}_t}{\tilde{r}_t^2}\frac{\partial \tilde{r}_t}{\partial t} \\ &= \frac{1}{\tilde{r}_t}\Big[-\nabla\cdot(f\hat{r}_t) + \frac{1}{2}g^2\nabla\cdot(\hat{r}_t\nabla_x \log \tilde{r}_t)\Big] \\ &\quad - \frac{\hat{r}_t}{\tilde{r}_t^2}\Big[-\nabla\cdot(f\tilde{r}_t) + \frac{1}{2}g^2\Delta\tilde{r}_t - \frac{1}{2}g^2(V_t - \mathbb{E}_{\tilde{r}_t}[V_t])\tilde{r}_t\Big]. \end{aligned} \tag{39}$$

Drift terms.

$$-\frac{1}{\tilde{r}_t}\nabla\cdot(f\hat{r}_t) + \frac{\hat{r}_t}{\tilde{r}_t^2}\nabla\cdot(f\tilde{r}_t) = -f\cdot\Big(\frac{\nabla\hat{r}_t}{\tilde{r}_t} - \frac{\hat{r}_t\nabla\tilde{r}_t}{\tilde{r}_t^2}\Big) = -f\cdot\nabla\rho_t. \tag{40}$$

Diffusion terms. Since $\hat{r}_t\nabla_x \log \tilde{r}_t = \rho_t\nabla\tilde{r}_t$,

$$\begin{aligned}\frac{1}{\tilde{r}_t}\cdot\frac{1}{2}g^2\nabla\cdot(\hat{r}_t\nabla_x\log\tilde{r}_t) &= \frac{g^2}{2\tilde{r}_t}\nabla\cdot(\rho_t\nabla\tilde{r}_t) \\ &= \frac{1}{2}g^2\nabla\log\tilde{r}_t\cdot\nabla\rho_t + \frac{1}{2}g^2\frac{\rho_t}{\tilde{r}_t}\Delta\tilde{r}_t, \end{aligned} \tag{41}$$

while

$$-\frac{\hat{r}_t}{\tilde{r}_t^2}\cdot\frac{1}{2}g^2\Delta\tilde{r}_t = -\frac{1}{2}g^2\frac{\rho_t}{\tilde{r}_t}\Delta\tilde{r}_t. \tag{42}$$

The $\Delta\tilde{r}_t$ terms cancel, leaving $\frac{1}{2}g^2\nabla\log\tilde{r}_t\cdot\nabla\rho_t$.

Potential term.

$$-\frac{\hat{r}_t}{\tilde{r}_t^2}\cdot\Big[-\frac{1}{2}g^2(V_t - \mathbb{E}_{\tilde{r}_t}[V_t])\tilde{r}_t\Big] = \frac{1}{2}g^2(V_t - \mathbb{E}_{\tilde{r}_t}[V_t])\rho_t. \tag{43}$$

Collecting the drift, diffusion, and potential contributions yields

$$\frac{\partial\rho_t}{\partial t} = -f\cdot\nabla\rho_t + \frac{1}{2}g^2\nabla\log\tilde{r}_t\cdot\nabla\rho_t + \frac{1}{2}g^2(V_t - \mathbb{E}_{\tilde{r}_t}[V_t])\rho_t. \tag{44}$$

Step 3: Solution along characteristics. Let X_s^x solve the characteristic ODE $dX_s^x/ds = f(X_s^x, s) - \frac{1}{2}g^2(s)\nabla_x\log\tilde{r}_s(X_s^x)$ with initial condition $X_{t_0}^x = x$. Along this trajectory, the total derivative of $\rho_s(X_s^x)$ is

$$\begin{aligned}\frac{d}{ds}\rho_s(X_s^x) &= \frac{\partial\rho_s}{\partial s}(X_s^x) + \nabla\rho_s(X_s^x)\cdot\frac{dX_s^x}{ds} \\ &= \frac{\partial\rho_s}{\partial s}(X_s^x) + \nabla\rho_s(X_s^x)\cdot\Big(f(X_s^x, s) - \frac{1}{2}g^2(s)\nabla\log\tilde{r}_s(X_s^x)\Big) \\ &= \frac{1}{2}g^2(s)\big(V_s(X_s^x) - \mathbb{E}_{\tilde{r}_s}[V_s]\big)\rho_s(X_s^x), \end{aligned} \tag{45}$$

where the last equality follows from (44). Integrating from t_0 to T gives

$$\log\frac{\rho_T(X_T^x)}{\rho_{t_0}(X_{t_0}^x)} = \frac{1}{2}\int_{t_0}^{T} g^2(s)\big(V_s(X_s^x) - \mathbb{E}_{\tilde{r}_s}[V_s]\big)\,ds. \tag{46}$$

Because the transported density is initialized from $\tilde{r}_T$, we have $\rho_T = \hat{r}_T/\tilde{r}_T = 1$ at the terminal time. Hence

$$\rho_{t_0}(x) = \exp\Big(-\frac{1}{2}\int_{t_0}^{T} g^2(s)\big(V_s(X_s^x) - \mathbb{E}_{\tilde{r}_s}[V_s]\big)\,ds\Big). \tag{47}$$

Step 4: Product-reference form. Multiplying by $\tilde{r}_{t_0} = Z_{t_0}^{-1}p_{t_0}^{\omega}q_{t_0}^{1-\omega}$ and absorbing all x-independent factors into the normalization constant gives

$$\hat{r}_{t_0}(x) = \frac{1}{Z'}p_{t_0}^{\omega}(x)q_{t_0}^{1-\omega}(x)\,\exp\Big(-\frac{1}{2}\int_{t_0}^{T} g^2(s)\,V_s(X_s^x)\,ds\Big). \tag{48}$$

Step 5: Equivalent form relative to p_{t_0}. It remains to rewrite the product-reference form as the equivalent expression in Theorem 3.2. Let $L_t(x) = \log p_t(x) - \log q_t(x)$, hence $\nabla L_t = \nabla \log p_t - \nabla \log q_t$ and $L_T(X_T^x) = \log p_T(X_T^x) - \log q_T(X_T^x)$.

Rewriting the prefactor. The prefactor $p_{t_0}^{\omega} q_{t_0}^{1-\omega}$ can be written as

$$\begin{aligned} p_{t_0}^{\omega}(x) q_{t_0}^{1-\omega}(x) &= p_{t_0}(x) \Big(\frac{p_{t_0}(x)}{q_{t_0}(x)}\Big)^{\omega-1} \\ &= p_{t_0}(x) e^{(\omega-1)L_{t_0}(x)} \\ &= p_{t_0}(x) \exp\Big((\omega-1)\big(L_T(X_T^x) - \int_{t_0}^{T} \frac{d}{ds} L_s(X_s^x)\, ds\big)\Big) \\ &= p_{t_0}(x) \exp\Big((\omega-1) L_T(X_T^x) - (\omega-1) \int_{t_0}^{T} \frac{d}{ds} L_s(X_s^x)\, ds\Big), \end{aligned} \tag{49}$$

where the integral is evaluated along the characteristic ODE (12).

Computing dL/ds along the characteristic. The Fokker–Planck equations (24)–(25) imply that the log-densities satisfy

$$\frac{\partial \log p_s}{\partial s} = -\nabla \cdot f - f \cdot \nabla \log p_s + \frac{1}{2} g^2 \big(\Delta \log p_s + \|\nabla \log p_s\|^2\big), \tag{50}$$

$$\frac{\partial \log q_s}{\partial s} = -\nabla \cdot f - f \cdot \nabla \log q_s + \frac{1}{2} g^2 \big(\Delta \log q_s + \|\nabla \log q_s\|^2\big). \tag{51}$$

Subtracting (51) from (50) gives

$$\frac{\partial L_s}{\partial s} = -f \cdot \nabla L_s + \frac{1}{2} g^2 \big(\Delta L_s + \|\nabla \log p_s\|^2 - \|\nabla \log q_s\|^2\big). \tag{52}$$

Along the same characteristic ODE, the total derivative of L is

$$\begin{aligned} \frac{dL_s}{ds} &= \frac{\partial L_s}{\partial s} + \nabla L_s \cdot \frac{dX_s}{ds} \\ &= \frac{\partial L_s}{\partial s} + \nabla L_s \cdot \big(f - \frac{1}{2} g^2 \nabla \log \tilde{r}_s\big). \end{aligned} \tag{53}$$

Substituting (52) and $\nabla \log \tilde{r}_s = \omega \nabla \log p_s + (1-\omega) \nabla \log q_s$ gives

$$\begin{aligned} \frac{dL_s}{ds} &= -f \cdot \nabla L_s + \frac{1}{2} g^2 \big(\Delta L_s + \|\nabla \log p_s\|^2 - \|\nabla \log q_s\|^2\big) \\ &\quad + \nabla L_s \cdot f - \frac{1}{2} g^2 \nabla L_s \cdot \big(\omega \nabla \log p_s + (1-\omega) \nabla \log q_s\big) \\ &= \frac{1}{2} g^2 \Big[\Delta L_s + \|\nabla \log p_s\|^2 - \|\nabla \log q_s\|^2 - \omega \|\nabla L_s\|^2 - \nabla L_s \cdot \nabla \log q_s\Big], \end{aligned} \tag{54}$$

where we used $\nabla L_s = \nabla \log p_s - \nabla \log q_s$ and $\nabla L_s \cdot (\omega \nabla \log p_s + (1-\omega) \nabla \log q_s) = \omega \|\nabla L_s\|^2 + \nabla L_s \cdot \nabla \log q_s$.

Using $\|\nabla \log p_s\|^2 - \|\nabla \log q_s\|^2 = \|\nabla L_s\|^2 + 2 \nabla L_s \cdot \nabla \log q_s$, we obtain

$$\boxed{\frac{dL_s}{ds} = \frac{1}{2} g^2 \Big[\big(\Delta L_s + \|\nabla L_s\|^2 + \nabla L_s \cdot \nabla \log q_s\big) - \omega \|\nabla L_s\|^2\Big]}. \tag{55}$$

Combining the exponents. We now combine the rewritten prefactor (49) with the V-integral from Step 4. Since $V_s = \omega(\omega-1)\|\nabla L_s\|^2$, the full exponent becomes

$$\begin{aligned}
&(\omega-1)L_T(X_T^x) - (\omega-1)\int_{t_0}^T \frac{dL_s}{ds}\,ds - \frac{1}{2}\int_{t_0}^T g^2(s)\,\omega(\omega-1)\|\nabla L_s\|^2\,ds \\
&= (\omega-1)L_T(X_T^x) - (\omega-1)\int_{t_0}^T \frac{1}{2}g^2\Big[\Delta L_s + \|\nabla L_s\|^2 + \nabla L_s\cdot\nabla\log q_s - \omega\|\nabla L_s\|^2\Big]ds \\
&\quad - \frac{1}{2}\int_{t_0}^T g^2\,\omega(\omega-1)\|\nabla L_s\|^2\,ds \\
&= (\omega-1)L_T(X_T^x) - \frac{1}{2}(\omega-1)\int_{t_0}^T g^2 \\
&\quad \times\Big[\Delta L_s + \|\nabla L_s\|^2 + \nabla L_s\cdot\nabla\log q_s - \omega\|\nabla L_s\|^2 + \omega\|\nabla L_s\|^2\Big]ds \\
&= (\omega-1)L_T(X_T^x) - \frac{1}{2}(\omega-1)\int_{t_0}^T g^2\Big[\Delta L_s + \|\nabla L_s\|^2 + \nabla L_s\cdot\nabla\log q_s\Big]ds. \qquad (56)
\end{aligned}$$

Finally, because $\|\nabla L_s\|^2+\nabla L_s\cdot\nabla\log q_s = \nabla L_s\cdot(\nabla\log p_s-\nabla\log q_s+\nabla\log q_s) = \nabla L_s\cdot\nabla\log p_s$, the exponent simplifies to

$$(\omega-1)L_T(X_T^x) - \frac{1}{2}(\omega-1)\int_{t_0}^T g^2(s)\Big(\Delta L_s(X_s^x) + \nabla L_s(X_s^x)\cdot\nabla\log p_s(X_s^x)\Big)\,ds. \qquad (57)$$

Under the usual terminal condition $\log p_T = \log q_T$, we have $L_T(X_T^x) = 0$. Substituting the resulting exponent into the expression from Step 4 gives the log-ratio form (13). The product-reference form (11) follows directly from Step 4 after substituting $V_s = \omega(\omega-1)\|\nabla L_s\|^2$ and absorbing $\mathbb{E}_{\tilde{r}_s}[V_s]$ into the normalization constant.

This completes the proof of Theorem 3.2. □

B.2 Proof of Theorem 3.3 (Time-Dependent Guidance)

We now extend the proof to time-dependent guidance, $\omega = \omega(t)$. The time dependence adds a source term involving $\omega'(t)$ to the evolution of the product reference density. After solving the corresponding ratio equation, an integration-by-parts argument along the characteristic ODE converts this extra term into the same path-integral structure as in the constant-guidance case.

Step 1: Evolution of the time-dependent product. For $r_t = p_t^{\omega(t)}q_t^{1-\omega(t)}$, Lemma B.1 supplies the constant-ω spatial terms pointwise in time, while differentiating the time-dependent exponent contributes one additional term:

$$\begin{aligned}
\frac{\partial r_t}{\partial t} &= \omega p_t^{\omega-1}q_t^{1-\omega}\frac{\partial p_t}{\partial t} + (1-\omega)p_t^{\omega}q_t^{-\omega}\frac{\partial q_t}{\partial t} + \omega'(t)\,p_t^{\omega}q_t^{1-\omega}\log\frac{p_t}{q_t} \\
&= -\nabla\cdot(fr_t) + \frac{1}{2}g^2\Delta r_t - \frac{1}{2}g^2V_t r_t + \omega'(t)\log\frac{p_t}{q_t}\,r_t, \qquad (58)
\end{aligned}$$

where $V_t = \omega(t)(\omega(t)-1)\|\nabla_x\log p_t - \nabla_x\log q_t\|^2$, and ω' denotes the weak derivative of ω with respect to t.

Using the normalization calculation in Lemma B.1 and including this additional source term, the Fokker–Planck equation for $\tilde{r}_t$ becomes

$$\begin{aligned}
\frac{\partial\tilde{r}_t}{\partial t} &= -\nabla\cdot(f\tilde{r}_t) + \frac{1}{2}g^2\Delta\tilde{r}_t - \frac{1}{2}g^2\big(V_t - \mathbb{E}_{\tilde{r}_t}[V_t]\big)\tilde{r}_t \\
&\quad + \omega'(t)\Big(\log\frac{p_t}{q_t} - \mathbb{E}_{\tilde{r}_t}\big[\log\frac{p_t}{q_t}\big]\Big)\tilde{r}_t, \qquad (59)
\end{aligned}$$

Step 2: Ratio equation. Repeating the density-ratio calculation from Theorem 3.2 gives

$$\frac{\partial \rho_t}{\partial t} = -f \cdot \nabla \rho_t + \frac{1}{2} g^2 \nabla \log \tilde{r}_t \cdot \nabla \rho_t + \frac{1}{2} g^2 (V_t - \mathbb{E}_{\tilde{r}_t}[V_t]) \rho_t - \omega'(t) \Big(\log \frac{p_t}{q_t} - \mathbb{E}_{\tilde{r}_t} \big[\log \frac{p_t}{q_t} \big] \Big) \rho_t. \tag{60}$$

Step 3: Solution along characteristics. Solving (60) along the same characteristic ODE yields

$$\begin{aligned} \rho_{t_0}(x) &= \exp\left(\begin{array}{l} -\frac{1}{2} \int_{t_0}^{T} g^2(s) \big(V_s(X_s^x) - \mathbb{E}_{\tilde{r}_s}[V_s] \big)\, ds \\ + \int_{t_0}^{T} \omega'(s) \left(\log \frac{p_s(X_s^x)}{q_s(X_s^x)} - \mathbb{E}_{\tilde{r}_s} \left[\log \frac{p_s}{q_s} \right] \right) ds \end{array} \right) \\ &= \frac{1}{Z''} \exp\left(-\frac{1}{2} \int_{t_0}^{T} g^2(s) V_s(X_s^x)\, ds + \int_{t_0}^{T} \omega'(s) \log \frac{p_s(X_s^x)}{q_s(X_s^x)}\, ds \right), \end{aligned} \tag{61}$$

where all terms independent of x are absorbed into Z''. Since $\hat{r}_{t_0} = \tilde{r}_{t_0} \rho_{t_0}$, it remains to simplify the trajectory-dependent exponent.

Step 4: Integration by parts. Let $L_t(x) = \log p_t(x) - \log q_t(x)$. Along the characteristic trajectory,

$$\int_{t_0}^{T} \omega'(s) L_s(X_s^x)\, ds = \big[\omega(s) L_s(X_s^x) \big]_{t_0}^{T} - \int_{t_0}^{T} \omega(s) \frac{d}{ds} L_s(X_s^x)\, ds. \tag{62}$$

The total derivative of L along the characteristic ODE was computed in Appendix B.1, Step 5:

$$\frac{dL_s}{ds} = \frac{1}{2} g^2 \Big[\big(\Delta L_s + \|\nabla L_s\|^2 + \nabla L_s \cdot \nabla \log q_s \big) - \omega(s) \|\nabla L_s\|^2 \Big]. \tag{63}$$

Combining the ω' integral with the V-integral gives

$$\begin{aligned} &\int_{t_0}^{T} \omega'(s) L_s\, ds - \frac{1}{2} \int_{t_0}^{T} \omega(s)(\omega(s) - 1) g^2(s) \|\nabla L_s\|^2\, ds \\ &= \omega(T) L_T(X_T^x) - \omega(t_0) L_{t_0}(x) - \int_{t_0}^{T} \omega(s) \frac{dL_s}{ds}\, ds - \frac{1}{2} \int_{t_0}^{T} (\omega^2(s) - \omega(s)) g^2 \|\nabla L_s\|^2\, ds \\ &= \omega(T) L_T(X_T^x) - \omega(t_0) L_{t_0}(x) - \int_{t_0}^{T} \frac{1}{2} g^2(s) \omega(s) \big(\Delta L_s + \nabla L_s \cdot \nabla \log q_s \big)\, ds \\ &= (\omega(T) - 1) L_T(X_T^x) - (\omega(t_0) - 1) L_{t_0}(x) \\ &\quad - \int_{t_0}^{T} \left[\frac{1}{2} g^2(s) \omega(s) \big(\Delta L_s + \nabla L_s \cdot \nabla \log q_s \big) - \frac{dL_s}{ds} \right] ds \\ &= (\omega(T) - 1) L_T(X_T^x) - (\omega(t_0) - 1) L_{t_0}(x) \\ &\quad - \int_{t_0}^{T} \frac{1}{2} g^2(s) (\omega(s) - 1) \big(\Delta L_s + \nabla L_s \cdot \nabla \log q_s + \|\nabla L_s\|^2 \big)\, ds \\ &= (\omega(T) - 1) L_T(X_T^x) - (\omega(t_0) - 1) L_{t_0}(x) \\ &\quad - \int_{t_0}^{T} \frac{1}{2} g^2(s) (\omega(s) - 1) \big(\Delta L_s + \nabla L_s \cdot \nabla \log p_s \big)\, ds. \end{aligned} \tag{64}$$

Here we used (63).

Substituting this identity into $\hat{r}_{t_0} = \tilde{r}_{t_0} \rho_{t_0}$, the boundary term at t_0 combines with $p_{t_0}^{\omega(t_0)} q_{t_0}^{1-\omega(t_0)}$ to yield p_{t_0}, since $p_{t_0}^{\omega(t_0)} q_{t_0}^{1-\omega(t_0)} e^{-(\omega(t_0)-1) L_{t_0}(x)} = p_{t_0}$. The remaining terminal boundary factor is $e^{(\omega(T)-1) L_T(X_T^x)} = (p_T(X_T^x)/q_T(X_T^x))^{\omega(T)-1}$. This factor is one whenever $\log p_T = \log q_T$ or $\omega(T) = 1$, yielding (15). □

C TOY MODEL: ANALYTIC SCORE FORMULAS

This appendix records the closed-form quantities used in the toy model experiments of Section 5.1. Let $\mathbf{x}^k$ denote the k-th support point, with $K = 10$ points equally spaced on a circle of radius $R = 2$. The unconditional distribution is uniform, $q_0(\mathbf{x}) = \sum_{k=0}^{K-1}(1/K)\delta(\mathbf{x} - \mathbf{x}^k)$, whereas the conditional distribution reweights the same support by likelihoods $\boldsymbol{\ell}$:

$$p_0(\mathbf{x}) = \sum_{k=0}^{K-1} \frac{\pi_k \ell_k}{\sum_j \pi_j \ell_j}\, \delta(\mathbf{x} - \mathbf{x}^k). \tag{65}$$

In the experiments, $\pi_k = 1/K$ and $\boldsymbol{\ell} = [2, 2, 2, 2, 1, 1, 1, 1, 1, 1]$.

Under the VPSDE forward process $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sigma_t \boldsymbol{\epsilon}$ with $\sigma_t^2 = 1 - \bar{\alpha}_t$, the marginal density at time t is a Gaussian mixture:

$$p_t(\mathbf{x}) = \sum_{k=0}^{K-1} w_k\, \mathcal{N}\big(\mathbf{x}; \sqrt{\bar{\alpha}_t}\,\mathbf{x}^k, \sigma_t^2 I\big), \tag{66}$$

where $w_k = \pi_k \ell_k / \sum_j \pi_j \ell_j$ for p_t and $w_k = 1/K$ for q_t.

Score function.

$$\mathbf{score}_t(\mathbf{x}) = \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) = \frac{\sum_k \gamma_k(\mathbf{x})\, \boldsymbol{\mu}_{t,k} - \mathbf{x}}{\sigma_t^2}, \tag{67}$$

where $\boldsymbol{\mu}_{t,k} = \sqrt{\bar{\alpha}_t}\,\mathbf{x}^k$ and

$$\gamma_k(\mathbf{x}) = \frac{w_k\, \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{t,k}, \sigma_t^2 I)}{\sum_j w_j\, \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{t,j}, \sigma_t^2 I)}. \tag{68}$$

Score divergence.

$$\nabla_{\mathbf{x}} \cdot \mathbf{score}_t(\mathbf{x}) = \frac{\mathrm{Tr}\big(\mathrm{Cov}_\gamma[\boldsymbol{\mu}_t]\big)}{\sigma_t^4} - \frac{d}{\sigma_t^2}, \tag{69}$$

where $d = 2$ and $\mathrm{Cov}_\gamma[\boldsymbol{\mu}_t] = \sum_k \gamma_k \boldsymbol{\mu}_{t,k}\boldsymbol{\mu}_{t,k}^\top - (\sum_k \gamma_k \boldsymbol{\mu}_{t,k})(\sum_k \gamma_k \boldsymbol{\mu}_{t,k})^\top$.

Empirical label distribution. Sampling is stopped at the small positive time t_0 rather than at exactly 0. Each generated sample is assigned to the nearest support point $\mathbf{x}^k$, producing the empirical label distribution $\hat{P}_k^{\mathrm{samp}}$. Since t_0 is small, this nearest-neighbor label distribution provides a direct discrete summary of the sampled density around the original support points.

Theoretical distribution computation. For each support point $\mathbf{x}^k$, we compute the theoretical label weight $\hat{P}_k^{\mathrm{th}}$ by averaging the path-integral correction over local noise at time t_0. Specifically, draw $N_{\mathrm{mc}} = 1000$ perturbations $\mathbf{x}_{t_0}^{(k,\ell)} = \sqrt{\bar{\alpha}_{t_0}}\mathbf{x}^k + \sigma_{t_0}\boldsymbol{\epsilon}^{(\ell)}$. For each perturbation, integrate the DDIM characteristic ODE from t_0 to T using 1000 steps. At each step s_j, evaluate the conditional score $\mathbf{score}^{(p)}$, the unconditional score $\mathbf{score}^{(q)}$, their difference $\Delta\mathbf{score} = \mathbf{score}^{(p)} - \mathbf{score}^{(q)}$, and the corresponding divergence difference using (69). The discretized path integral is

$$I^{(k,\ell)} = \frac{1}{2}\sum_j (\omega(s_j) - 1)\, \beta(s_j) \left(\nabla \cdot \Delta\mathbf{score} + \Delta\mathbf{score} \cdot \mathbf{score}^{(p)}\right)\Big|_{\mathbf{x}_{s_j}^{(k,\ell)}} \Delta s, \tag{70}$$

where $\Delta s = T/1000 = 10^{-3}$ is the uniform step size, since $T = 1$. The resulting theoretical label probability is

$$\hat{P}_k^{\mathrm{th}} \propto p_{t_0}(\mathbf{x}^k)\, \frac{1}{N_{\mathrm{mc}}} \sum_{\ell=1}^{N_{\mathrm{mc}}} \exp(-I^{(k,\ell)}), \tag{71}$$

normalized over $k = 0, \ldots, K-1$. This is the discrete counterpart of the continuous theoretical density at time t_0, averaged over a small neighborhood of each support point, and is the quantity compared with $\hat{P}_k^{\mathrm{samp}}$ in Section 5.1.

C.1 Experimental Protocol and Complete Results

We use a discrete linear DDPM noise schedule (Ho et al., 2020) with 1000 steps, with β_k linearly spaced from 10^{-4} to 0.02 and $\bar{\alpha}_k = \prod_{j=1}^{k}(1-\beta_j)$. These discrete arrays determine the toy-model perturbations and guidance schedule. We use $T = 1$ and $t_0 = 0.001$.

The unconditional data distribution q_{data} is uniform over $K = 10$ equally spaced point masses on a circle of radius $R = 2$. The conditional distribution p_{data} assigns likelihood weights $\boldsymbol{\ell} = [2, 2, 2, 2, 1, 1, 1, 1, 1, 1]$ to the same support and then renormalizes, so that points 0–3 receive twice the mass of points 4–9. Under the forward Gaussian perturbation, both p_t and q_t become Gaussian mixtures, and their scores and score divergences admit closed-form expressions.

We compare two guidance strategies at $\bar{\omega} \in \{3, 5, 7, 9\}$:

- **Constant CFG:** $\omega(t) \equiv \bar{\omega}$.
- **Time-dependent CFG:** $\omega(t) = 1 + C_2 \cdot (\bar{\omega} - 1) \cdot \sqrt{\bar{\alpha}_t}/\beta(t)$, with $C = (T - t_0)/\int_{t_0}^{T} \sqrt{\bar{\alpha}_t}/\beta(t)\, dt$. (The $(1 - \bar{\alpha}_t)$ correction is omitted here because the toy model uses analytic score functions; the $1/\sigma_t^2$ error amplification that motivates Step 3 does not arise.)

For each setting, we draw $N = 200{,}000$ samples using 50-step DDIM sampling from $\mathcal{N}(\mathbf{0}, I)$. Each sample is assigned to its nearest data point, yielding the empirical label distribution $\hat{P}_k^{\text{samp}}$. The theoretical label probabilities $\hat{P}_k^{\text{th}}$ are computed from Theorem 3.2 for constant guidance and from Theorem 3.3 for time-dependent guidance, evaluated at each support point $\mathbf{x}^k$. Agreement is measured by mean absolute error (MAE), $\frac{1}{10}\sum_{k=0}^{9} |\hat{P}_k^{\text{samp}} - \hat{P}_k^{\text{th}}|$.

As an additional diagnostic, we compute the *Min-Max-ratio* from the empirical sample distribution over the four high-weight points ($k = 0, 1, 2, 3$), defined as $\min_{k\in\{0,1,2,3\}} \hat{P}_k^{\text{samp}} \,/\, \max_{k\in\{0,1,2,3\}} \hat{P}_k^{\text{samp}}$. This metric captures mode imbalance within the conditioned subset: a value near 1 indicates balanced coverage of the four preferred modes, whereas a value near 0 indicates that at least one preferred mode has nearly disappeared.

Complete results. Table 5 and Figure 3 report the complete results. Across both schedules and all guidance strengths, MAE remains below 0.0036, supporting the theoretical predictions. Higher Min-Max-ratios under time-dependent guidance provide an additional illustration of mode preservation.

Table 5: Complete toy-model verification ($N = 200{,}000$, 50-step DDIM). MAE compares empirical and theoretical label distributions; the Min-Max-ratio measures balance among the four preferred modes.

Schedule	$\bar{\omega}$	MAE↓	Min-Max-ratio↑
Constant CFG	3	0.00234	0.665
Constant CFG	5	0.00321	0.370
Constant CFG	7	0.00357	0.193
Constant CFG	9	0.00332	0.099
Time-dependent CFG	3	0.00112	0.959
Time-dependent CFG	5	0.00149	0.981
Time-dependent CFG	7	0.00181	0.898
Time-dependent CFG	9	0.00219	0.808

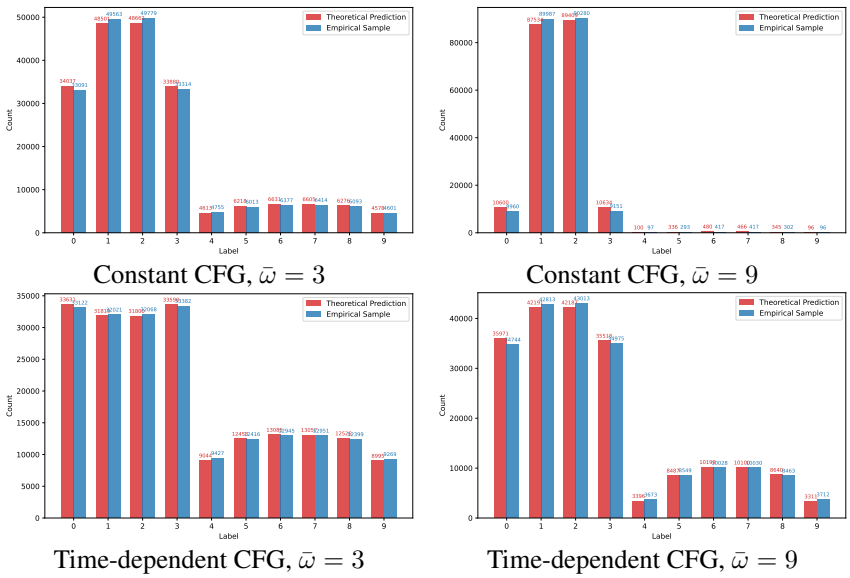


Figure 3: **Toy-model label distributions under constant and time-dependent guidance.** Top and bottom rows show constant and time-dependent CFG, respectively. Bars compare theoretical and empirical distributions ($N = 200{,}000$, 50-step DDIM).

D STABLE DIFFUSION EXPERIMENTS: PROTOCOL AND COMPLETE RESULTS

D.1 STABLE DIFFUSION EVALUATION PROTOCOL

Checkpoints and generation. We use the RunwayML SD1.5, SD2.1-base, and SDXL-base-1.0 checkpoints. For each method and backbone, we use the same randomly selected $N_{\text{gen}} = 1{,}000$ prompt–image pairs from the MS-COCO 2017 validation set and generate one image per prompt. Images are produced at the native evaluation resolution of each model: 512×512 for SD1.5 and SD2.1, and 1024×1024 for SDXL.

Noise schedule. For SD1.5, SD2.1, and SDXL, DG-CFG uses the pretrained checkpoint scheduler's `betas` and `alphas_cumprod` arrays at each sampled timestep. These checkpoints use a `scaled_linear` schedule: $\sqrt{\beta_k}$ is linearly spaced between $\sqrt{0.00085}$ and $\sqrt{0.012}$ over 1000 training timesteps, making β_k quadratic in the timestep index. We substitute these discrete values into Eq. (18) and normalize the guidance budget as above.

Metrics. We evaluate overall image quality with Fréchet Inception Distance (FID), prompt alignment with CLIP score, perceptual quality with ImageReward (IR), and color over-saturation with Sat. For diversity, we select $N_{\text{div}} = 100$ prompts, generate 16 images per prompt from different initializations, compute the Vendi score for each prompt, and average across prompts. CLIP, IR, and Sat are averaged over the N_{gen} generated images. Under the same evaluators, real MS-COCO images yield $\text{CLIP} = 30.57$, $\text{IR} = 0.38$, and $\text{Sat} = 0.33$.

D.2 COMPLETE MAIN-COMPARISON RESULTS

This subsection reports the complete guidance-strength sweeps and qualitative grids summarized in Section 5.2. All methods use DDIM sampling at 8 NFE. The compared schedules are

- **Constant CFG:** $\omega(t) \equiv \bar{\omega}$.
- **Interval CFG (Kynkäänniemi et al., 2024):** $\omega(t) = 1 + (\bar{\omega} - 1)/0.6$ for $t \in [0.2T, 0.8T]$, and $\omega(t) = 1$ otherwise.
- **β–CFG (Malarz et al., 2025):** $\omega(t) = \bar{\omega} \cdot 6 \cdot (t/T) \cdot (1 - t/T)$.
- **DG-CFG (ours):** $\omega(t) = 1 + C(\bar{\omega} - 1)(1 - \bar{\alpha}_t)\sqrt{\bar{\alpha}_t}/\beta(t)$, with C defined in Section 4.

All non-constant schedules are normalized to satisfy $\int_{t_0}^{T} (\omega(s) - 1)\, ds = (\bar{\omega} - 1)(T - t_0)$.

Tables 6–8 report the complete quantitative results, and Figures 4–6 provide representative samples. We summarize the main observations below.

Diversity–fidelity trade-off. Across all three backbones, DG-CFG provides a favorable diversity–fidelity trade-off. In the low-to-moderate guidance regime, where outputs remain broadly comparable in quality, it generally preserves more diversity than constant CFG while maintaining competitive FID and ImageReward. Interval CFG can produce higher Vendi scores in some settings, but often with weaker fidelity or perceptual quality. At stronger guidance, Vendi alone becomes less informative because color distortion, structural failures, and texture artifacts can disperse feature embeddings and spuriously inflate the score without adding meaningful semantic variation. Under the quality-aware assessment, DG-CFG more consistently preserves meaningful variation while avoiding the degradation observed in the baselines.

Image-quality improvement at high guidance. The advantage of DG-CFG is most pronounced in the high-guidance regime. As the nominal guidance strength increases, constant and heuristic schedules progressively lose perceptual quality, prompt alignment, or distributional fidelity, whereas DG-CFG remains comparatively stable across all three backbones. The contrast is strongest on SDXL: baseline performance deteriorates sharply under strong guidance, while DG-CFG maintains high ImageReward and CLIP scores together with competitive FID.

Saturation control. The Sat metric captures the color over-saturation commonly induced by strong guidance. Across all three backbones, Sat increases steadily with the guidance strength under constant and interval CFG, approaching or exceeding the real-image reference at high guidance. By

contrast, DG-CFG exhibits a substantially weaker increase and remains at or below the reference level throughout the evaluated guidance range.

Qualitative visualization. Figures 4–6 show representative samples from all three backbones using the same prompts and random seeds. As guidance becomes stronger, the baselines progressively develop posterized colors, washed-out textures, and less coherent object boundaries. DG-CFG preserves more natural color balance, texture detail, and object structure throughout the evaluated guidance range. These visual results are consistent with the quantitative findings, confirming that the stability of DG-CFG translates into more natural and coherent samples rather than merely improved aggregate metrics.

Table 6: Main comparison on Stable Diffusion 1.5 with DDIM sampling at 8 NFE. Metrics are computed over 1K MS-COCO validation prompts, except Div (Vendi score), which is averaged over 100 prompts with 16 samples per prompt. The real-image reference values are CLIP=30.57, IR=0.38, and Sat=0.33.

Metric	Method	$\bar{\omega} = 3$	$\bar{\omega} = 7$	$\bar{\omega} = 11$	$\bar{\omega} = 15$	$\bar{\omega} = 19$	$\bar{\omega} = 23$
CLIP↑	Constant CFG	30.15	**31.29**	31.38	31.29	31.01	30.64
	Interval CFG	**30.39**	30.95	30.75	30.32	29.78	29.10
	β–CFG	30.34	31.21	**31.42**	31.28	31.18	30.88
	DG-CFG (ours)	29.95	31.08	31.27	**31.32**	**31.29**	**31.39**
IR↑	Constant CFG	−0.503	−0.104	−0.032	−0.061	−0.177	−0.330
	Interval CFG	−0.463	−0.255	−0.308	−0.518	−0.758	−0.986
	β–CFG	**−0.437**	**−0.103**	**−0.023**	−0.022	−0.104	−0.239
	DG-CFG (ours)	−0.522	−0.115	−0.026	**+0.013**	**+0.024**	**+0.029**
Sat	Constant CFG	0.212	0.258	0.317	0.379	0.430	0.469
	Interval CFG	0.220	0.275	0.338	0.392	0.431	0.462
	β–CFG	0.214	0.256	0.308	0.359	0.401	0.438
	DG-CFG (ours)	0.206	0.229	0.256	0.284	0.310	0.332
FID↓	Constant CFG	74.44	**67.67**	**68.73**	71.57	74.94	82.34
	Interval CFG	**72.46**	70.83	76.84	86.60	99.07	110.89
	β–CFG	73.14	68.14	70.73	73.07	77.92	85.38
	DG-CFG (ours)	77.26	67.87	68.79	**70.62**	**72.54**	**74.53**
Div↑	Constant CFG	6.54	5.40	5.18	5.27	5.43	5.67
	Interval CFG	6.47	5.95	6.04	6.28	6.61	6.80
	β–CFG	6.47	5.44	5.17	5.26	5.37	5.57
	DG-CFG (ours)	6.81	5.71	5.34	5.22	5.17	5.24

Table 7: Main comparison on Stable Diffusion 2.1 with DDIM sampling at 8 NFE. Metrics are computed over 1K MS-COCO validation prompts, except Div (Vendi score), which is averaged over 100 prompts with 16 samples per prompt. The real-image reference values are CLIP=30.57, IR=0.38, and Sat=0.33.

Metric	Method	$\bar{\omega} = 3$	$\bar{\omega} = 7$	$\bar{\omega} = 11$	$\bar{\omega} = 15$	$\bar{\omega} = 19$	$\bar{\omega} = 23$
CLIP↑	Constant CFG	30.54	**31.48**	31.48	31.48	31.25	30.98
	Interval CFG	**30.71**	31.02	30.93	30.54	30.12	29.80
	β–CFG	30.62	31.36	**31.51**	**31.55**	31.40	31.14
	DG-CFG (ours)	30.36	31.23	31.36	31.38	**31.49**	**31.43**
IR↑	Constant CFG	−0.264	**+0.130**	+0.203	+0.176	+0.078	−0.092
	Interval CFG	−0.235	−0.055	−0.101	−0.275	−0.456	−0.630
	β–CFG	**−0.212**	+0.119	+0.203	+0.214	+0.142	+0.060
	DG-CFG (ours)	−0.322	+0.091	**+0.209**	**+0.238**	**+0.245**	**+0.244**
Sat	Constant CFG	0.207	0.248	0.299	0.351	0.396	0.431
	Interval CFG	0.214	0.262	0.315	0.361	0.396	0.422
	β–CFG	0.209	0.248	0.291	0.334	0.372	0.398
	DG-CFG (ours)	0.203	0.222	0.247	0.272	0.295	0.315
FID↓	Constant CFG	73.40	**67.45**	70.10	72.63	74.79	79.94
	Interval CFG	**71.29**	72.60	77.43	86.10	95.66	105.19
	β–CFG	71.79	68.64	71.36	73.27	76.73	83.11
	DG-CFG (ours)	75.54	68.70	**70.08**	**71.98**	**74.05**	**76.43**
Div↑	Constant CFG	5.99	5.01	4.81	4.87	5.08	5.33
	Interval CFG	6.04	5.49	5.66	5.86	6.18	6.38
	β–CFG	5.98	4.99	4.85	4.89	5.01	5.24
	DG-CFG (ours)	6.26	5.20	4.97	4.88	4.85	4.83

Table 8: Main comparison on Stable Diffusion XL with DDIM sampling at 8 NFE. Metrics are computed over 1K MS-COCO validation prompts, except Div (Vendi score), which is averaged over 100 prompts with 16 samples per prompt. The real-image reference values are CLIP=30.57, IR=0.38, and Sat=0.33.

Metric	Method	$\bar{\omega} = 3$	$\bar{\omega} = 7$	$\bar{\omega} = 11$	$\bar{\omega} = 15$	$\bar{\omega} = 19$	$\bar{\omega} = 23$
CLIP↑	Constant CFG	30.68	31.82	31.75	31.34	30.51	29.46
	Interval CFG	30.73	31.32	30.82	29.97	29.04	28.17
	β–CFG	**30.86**	**31.96**	**32.06**	31.90	31.44	30.76
	DG-CFG (ours)	30.23	31.65	32.03	**32.10**	**32.10**	**32.00**
IR↑	Constant CFG	−0.349	+0.205	+0.159	−0.045	−0.383	−0.717
	Interval CFG	−0.344	−0.115	−0.304	−0.604	−0.899	−1.144
	β–CFG	**−0.289**	**+0.279**	**+0.350**	+0.247	+0.053	−0.222
	DG-CFG (ours)	−0.530	+0.102	+0.321	**+0.392**	**+0.417**	**+0.401**
Sat	Constant CFG	0.147	0.180	0.221	0.267	0.313	0.351
	Interval CFG	0.153	0.183	0.220	0.259	0.293	0.326
	β–CFG	0.150	0.182	0.219	0.259	0.292	0.318
	DG-CFG (ours)	0.146	0.161	0.179	0.200	0.221	0.241
FID↓	Constant CFG	91.33	72.86	73.18	79.01	91.75	105.39
	Interval CFG	**88.45**	79.47	85.78	100.31	113.81	126.27
	β–CFG	88.83	**72.29**	**70.65**	74.59	82.42	92.56
	DG-CFG (ours)	99.47	76.79	73.46	**73.49**	**73.57**	**75.02**
Div↑	Constant CFG	5.53	4.84	5.04	5.58	6.14	6.70
	Interval CFG	5.77	5.62	6.08	6.56	6.87	7.06
	β–CFG	5.49	4.75	4.72	5.01	5.51	5.94
	DG-CFG (ours)	5.79	5.00	4.74	4.63	4.62	4.70

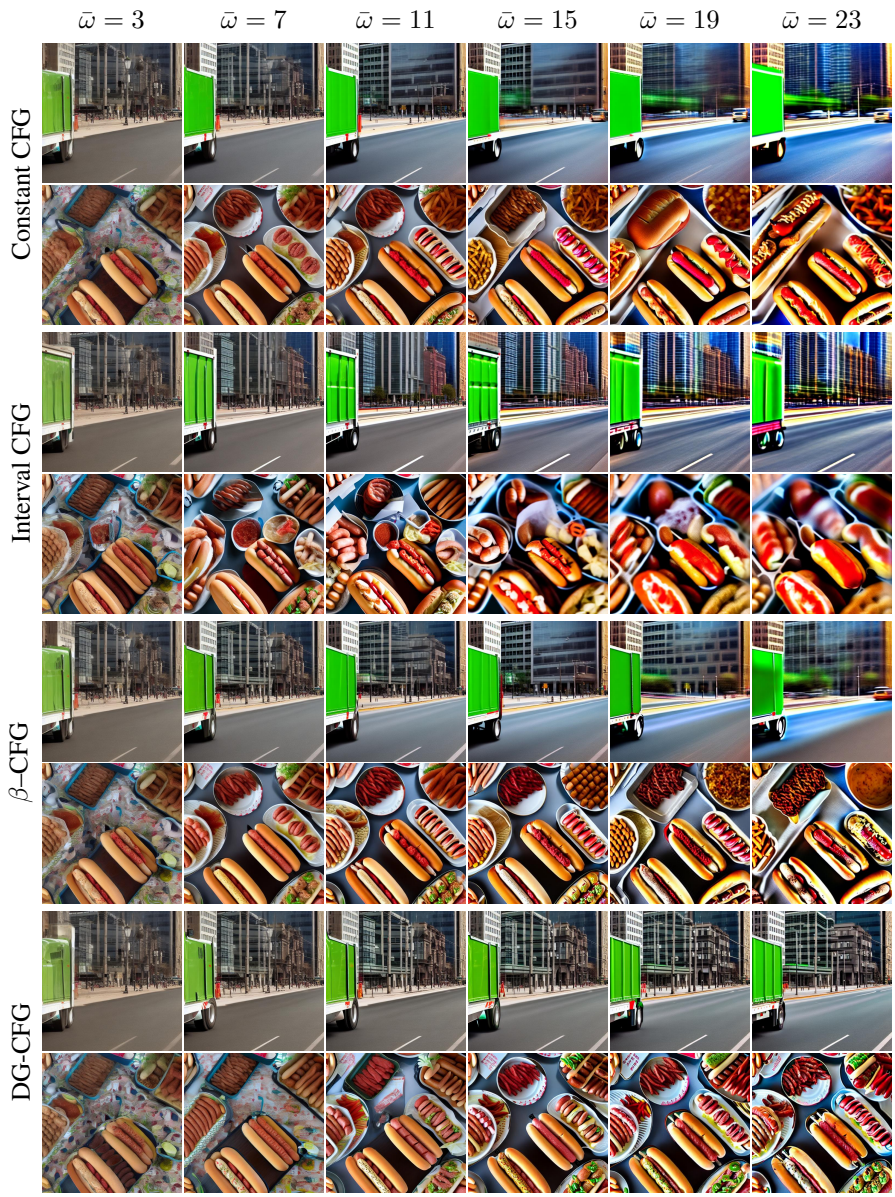


Figure 4: Qualitative comparison on Stable Diffusion 1.5 (DDIM, 8 NFE). Columns vary the nominal guidance strength $\bar{\omega} \in \{3, 7, 11, 15, 19, 23\}$; rows are grouped by method, with two prompts per method. The upper prompt is *a large green truck on a city street*, and the lower prompt is *a table topped with trays full of hot dogs*.

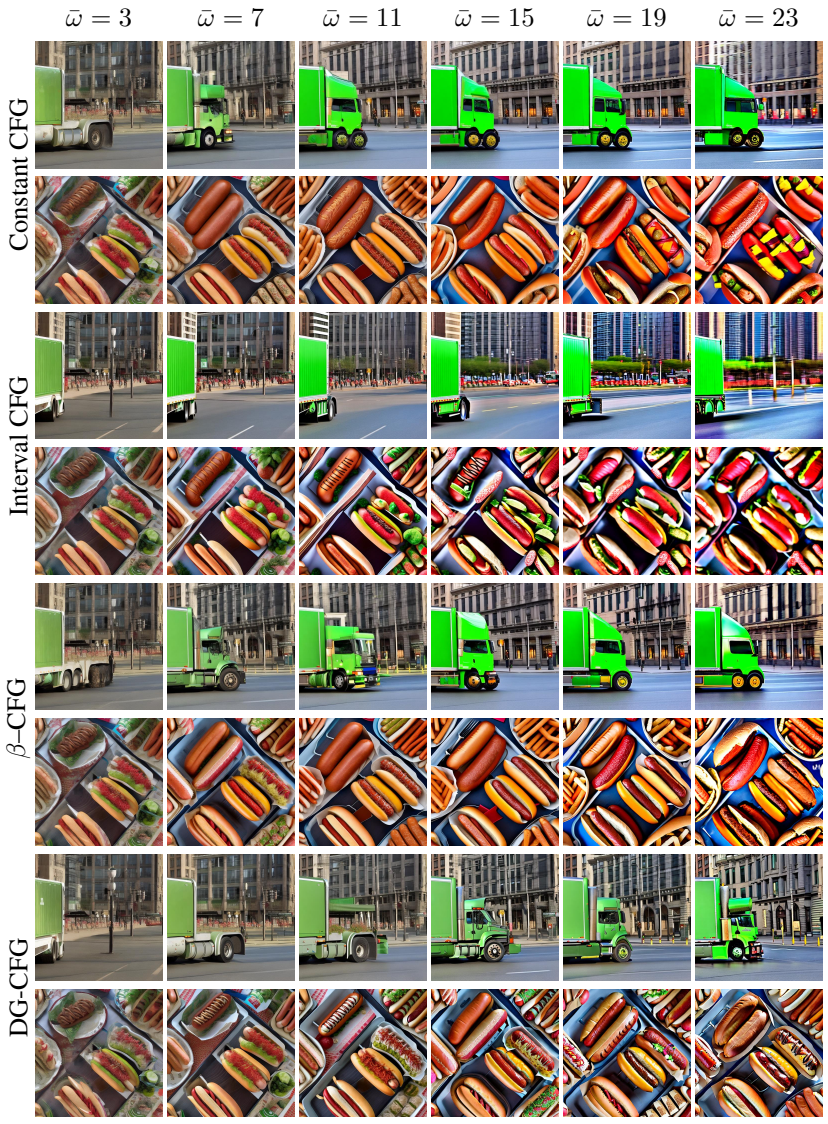


Figure 5: Qualitative comparison on Stable Diffusion 2.1 (DDIM, 8 NFE). Columns vary the nominal guidance strength $\bar{\omega} \in \{3, 7, 11, 15, 19, 23\}$; rows are grouped by method, with two prompts per method. The prompts and random seeds are identical to Figure 4: the upper prompt is *a large green truck on a city street*, and the lower prompt is *a table topped with trays full of hot dogs*.

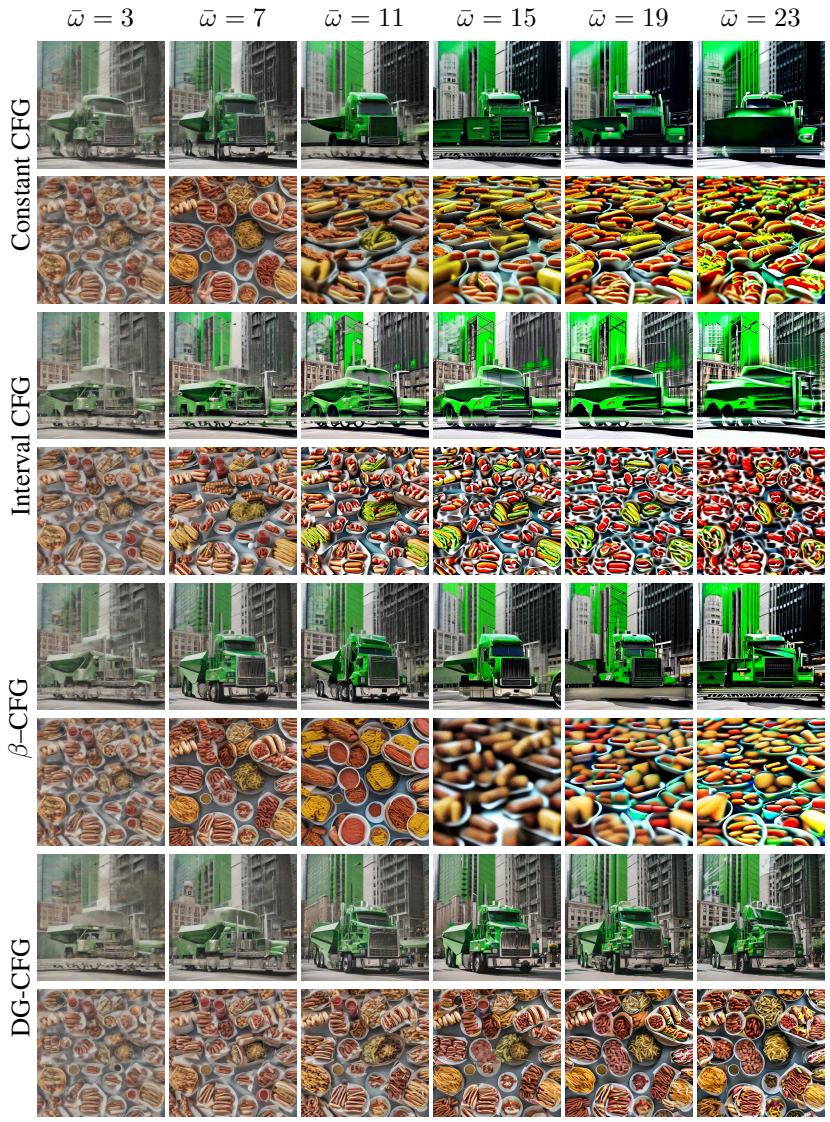


Figure 6: Qualitative comparison on Stable Diffusion XL (DDIM, 8 NFE). Columns vary the nominal guidance strength $\bar{\omega} \in \{3, 7, 11, 15, 19, 23\}$; rows are grouped by method, with two prompts per method. The prompts and random seeds are identical to Figure 4: the upper prompt is *a large green truck on a city street*, and the lower prompt is *a table topped with trays full of hot dogs*.

D.3 Additional Guidance Comparisons with Euler Sampling

We use SDXL-base-1.0 with `EulerDiscreteScheduler`, 8 sampling steps, and 1024×1024 output resolution. All five methods share the same 1,000 MS-COCO 2017 validation prompt–image pairs and per-prompt seeds (base seed 0), with an empty negative prompt. We sweep nominal guidance strengths $\bar{\omega} \in \{3, 7, 11, 15, 19, 23\}$. APG and CFG Rescale use a constant nominal guidance scale; APG uses $\eta = 0$, norm threshold 15, and momentum -0.5, while CFG Rescale uses a rescaling factor of 0.7. We report CLIP, ImageReward, FID, and saturation using the evaluation protocol in Appendix D.1.

Table 9 reports the complete sweep. DG-CFG achieves the highest CLIP and ImageReward at the three strongest guidance settings, while C^2FG and APG offer better FID at intermediate strengths. DG-CFG also limits the growth in saturation: from $\bar{\omega} = 7$ to 23, Sat rises from 0.157 to 0.238, compared with 0.177 to 0.349 for Constant CFG.

Table 9: Complete SDXL comparison with 8-step Euler sampling. Bold denotes the best CLIP, IR, and FID at each nominal guidance strength; Sat is reported without a preferred direction.

Metric	Method	$\bar{\omega} = 3$	$\bar{\omega} = 7$	$\bar{\omega} = 11$	$\bar{\omega} = 15$	$\bar{\omega} = 19$	$\bar{\omega} = 23$
CLIP↑	Constant CFG	**30.623**	31.663	31.580	31.168	30.322	29.281
	C^2FG	30.285	**31.666**	31.842	31.791	31.605	31.227
	APG	29.783	31.339	31.759	31.760	31.677	31.473
	CFG Rescale	30.619	31.648	31.787	31.667	31.315	30.822
	DG-CFG (ours)	30.168	31.457	**31.849**	**31.915**	**31.906**	**31.849**
IR↑	Constant CFG	-0.346	0.192	0.147	-0.071	-0.401	-0.737
	C^2FG	-0.494	0.181	**0.316**	0.274	0.153	-0.016
	APG	-0.639	0.082	0.275	0.313	0.258	0.134
	CFG Rescale	**-0.326**	**0.223**	0.280	0.152	-0.077	-0.338
	DG-CFG (ours)	-0.505	0.102	**0.316**	**0.373**	**0.402**	**0.391**
FID↓	Constant CFG	**86.41**	71.87	73.72	80.71	94.78	109.52
	C^2FG	92.96	73.15	**70.36**	72.05	77.03	83.36
	APG	99.31	76.35	71.71	**70.53**	**73.01**	76.50
	CFG Rescale	**86.41**	**71.76**	71.45	74.99	83.28	93.29
	DG-CFG (ours)	93.37	75.09	72.23	73.38	73.34	**76.15**
Sat	Constant CFG	0.144	0.177	0.218	0.264	0.310	0.349
	C^2FG	0.141	0.162	0.189	0.221	0.253	0.286
	APG	0.139	0.151	0.179	0.221	0.273	0.320
	CFG Rescale	0.145	0.181	0.228	0.283	0.336	0.375
	DG-CFG (ours)	0.142	0.157	0.175	0.196	0.217	0.238

D.4 Complete Ablation Results

This subsection provides the complete quantitative sweeps, schedule definitions, component-wise analysis, and qualitative comparisons summarized in Section 5.2.4.

To isolate the contribution of each design component, we compare six schedule variants at 8 NFE. We perform complete ablations on SD1.5 and SD2.1 over $\bar{\omega} \in \{3, 7, 11, 15, 19, 23\}$, allowing us to assess whether the component-wise effects remain consistent across model generations. Each variant is normalized independently so that its integrated guidance deviation matches that of constant CFG.

- **Constant CFG:** $\omega(t) \equiv \bar{\omega}$ (baseline).
- **Medium 1:** $\omega(t) = 1 + C_1 \cdot (\bar{\omega} - 1)/\beta(t)$ (time-balance only).
- **Medium 2:** $\omega(t) = 1 + C_2 \cdot (\bar{\omega} - 1) \cdot \sqrt{\bar{\alpha}_t}/\beta(t)$ (+ signal-content weighting).
- **Medium 3:** $\omega(t) = 1 + C_3 \cdot (\bar{\omega} - 1) \cdot (1 - \bar{\alpha}_t)/\beta(t)$ (time balancing and low-noise attenuation).
- **Medium 4:** $\omega(t) = 1 + C_4 \cdot (\bar{\omega} - 1) \cdot (1 - \bar{\alpha}_t)\sqrt{\bar{\alpha}_t}$ (signal-content weighting and low-noise attenuation).
- **DG-CFG (ours):** $\omega(t) = 1 + C \cdot (\bar{\omega} - 1) \cdot (1 - \bar{\alpha}_t)\sqrt{\bar{\alpha}_t}/\beta(t)$ (all three factors).

Tables 10 and 11 report the complete SD1.5 and SD2.1 results, respectively. We analyze the components in the same order as the schedule construction in Section 4, focusing on how each factor changes diversity, fidelity, perceptual quality, and saturation as guidance becomes stronger.

Step 1: time-balance (Medium 1 vs. constant). Normalizing by $\beta(t)$ improves the behavior of the sampler across the evaluated guidance ranges on both backbones. Medium 1 generally preserves more sample variation than constant CFG in the quality-comparable regime, while also moderating the saturation drift that becomes increasingly visible under stronger guidance. These results support inverse-rate time balancing as a practical choice for temporal guidance allocation.

Step 2: Signal-content weighting (Medium 2 vs. Medium 1). Adding $\sqrt{\bar{\alpha}_t}$ further shifts guidance toward timesteps with stronger signal content. This increases raw diversity scores and further limits the growth of saturation, but does not preserve image quality. Across both backbones, Medium 2 tends to lose prompt alignment, perceptual quality, and distributional fidelity relative to Medium 1. These results show that the additional shift toward low noise trades image quality for greater sample spread and lower saturation.

Step 3: Low-noise attenuation (DG-CFG vs. Medium 2). Adding the $(1 - \bar{\alpha}_t)$ factor improves IR and FID relative to Medium 2 while retaining much of the saturation control gained from the earlier reweighting, consistent with the motivation for attenuating low-noise guidance. It also remains stable as guidance becomes stronger, providing greater meaningful variation in the high-guidance regime while preserving perceptual quality and distributional fidelity.

Component removal variants (Medium 2–4). Medium 2–4 each remove one of DG-CFG's three components. Under strong guidance, every removal weakens the joint trade-off among IR, saturation, and FID on both backbones, supporting their complementary roles in the final schedule.

The qualitative comparisons of the progressive variants on SD1.5 and SD2.1 in Figures 7 and 8, respectively, reinforce these findings. Constant CFG produces pronounced color over-saturation and limited fine-grained detail. Medium 1 improves fine-detail quality and sample diversity while moderating saturation. Medium 2 controls saturation further, but at the cost of unnatural textures and structural artifacts. DG-CFG provides a favorable balance, preserving meaningful variation and fine detail while keeping color saturation controlled and producing more coherent textures and structures.

Taken together, the ablation results on SD1.5 and SD2.1 support the combined schedule design and show a consistent quality–saturation trade-off across both backbones.

Table 10: Component-wise ablation of the DG-CFG schedule on Stable Diffusion 1.5 at 8 NFE. Medium 1 and Medium 2 form the progressive schedule; Medium 3 and Medium 4 omit signal-content weighting and time balancing from DG-CFG, respectively. Metrics are computed over 1K MS-COCO validation prompts, except Div (Vendi score), which is averaged over 100 prompts with 16 samples per prompt. The real-image reference values are CLIP=30.57, IR=0.38, and Sat=0.33.

Metric	Method	$\bar{\omega} = 3$	$\bar{\omega} = 7$	$\bar{\omega} = 11$	$\bar{\omega} = 15$	$\bar{\omega} = 19$	$\bar{\omega} = 23$
CLIP↑	Constant CFG	30.15	**31.29**	31.38	31.29	31.01	30.64
	Medium 1	29.53	30.88	31.25	31.22	31.25	31.19
	Medium 2	28.94	30.29	30.47	30.51	30.31	30.23
	Medium 3	30.26	31.28	31.39	**31.33**	31.11	30.77
	Medium 4	**30.28**	31.21	**31.40**	31.30	31.26	31.02
	DG-CFG	29.95	31.08	31.27	31.32	**31.29**	**31.39**
IR↑	Constant CFG	−0.503	−0.104	−0.032	−0.061	−0.177	−0.330
	Medium 1	−0.655	−0.203	−0.051	−0.021	+0.001	−0.000
	Medium 2	−0.823	−0.377	−0.263	−0.232	−0.257	−0.283
	Medium 3	−0.463	**−0.089**	**−0.007**	−0.014	−0.104	−0.259
	Medium 4	**−0.458**	−0.099	−0.019	+0.003	−0.049	−0.141
	DG-CFG	−0.522	−0.115	−0.026	**+0.013**	**+0.024**	**+0.029**
Sat	Constant CFG	0.212	0.258	0.317	0.379	0.430	0.469
	Medium 1	0.202	0.219	0.239	0.260	0.282	0.305
	Medium 2	0.199	0.209	0.220	0.234	0.247	0.259
	Medium 3	0.213	0.256	0.312	0.369	0.417	0.455
	Medium 4	0.212	0.251	0.298	0.346	0.386	0.420
	DG-CFG	0.206	0.229	0.256	0.284	0.310	0.332
FID↓	Constant CFG	74.44	67.67	68.73	71.57	74.94	82.34
	Medium 1	82.34	69.39	**68.24**	**69.19**	**71.15**	**73.24**
	Medium 2	90.77	75.87	75.56	77.64	82.97	87.65
	Medium 3	**73.31**	**67.63**	69.41	72.22	75.85	82.68
	Medium 4	73.44	68.00	70.45	72.38	76.29	82.00
	DG-CFG	77.26	67.87	68.79	70.62	72.54	74.53
Div↑	Constant CFG	6.54	5.40	5.18	5.27	5.43	5.67
	Medium 1	7.08	5.86	5.47	5.33	5.25	5.26
	Medium 2	**7.54**	**6.53**	**6.10**	**5.98**	**5.82**	**5.80**
	Medium 3	6.48	5.40	5.15	5.18	5.31	5.55
	Medium 4	6.51	5.42	5.18	5.20	5.32	5.43
	DG-CFG	6.81	5.71	5.34	5.22	5.17	5.24

Table 11: Component-wise ablation of the DG-CFG schedule on Stable Diffusion 2.1 at 8 NFE. Medium 1 and Medium 2 form the progressive schedule; Medium 3 and Medium 4 omit signal-content weighting and time balancing from DG-CFG, respectively. Metrics are computed over 1K MS-COCO validation prompts, except Div (Vendi score), which is averaged over 100 prompts with 16 samples per prompt. The real-image reference values are CLIP=30.57, IR=0.38, and Sat=0.33.

Metric	Method	$\bar{\omega} = 3$	$\bar{\omega} = 7$	$\bar{\omega} = 11$	$\bar{\omega} = 15$	$\bar{\omega} = 19$	$\bar{\omega} = 23$
CLIP↑	Constant CFG	30.54	**31.48**	31.48	31.48	31.25	30.98
	Medium 1	30.03	31.14	31.33	31.35	31.34	31.28
	Medium 2	29.47	30.51	30.63	30.60	30.43	30.25
	Medium 3	**30.61**	31.47	**31.52**	31.46	31.38	31.03
	Medium 4	30.58	31.34	31.49	**31.52**	31.41	31.24
	DG-CFG	30.36	31.23	31.36	31.38	**31.49**	**31.43**
IR↑	Constant CFG	−0.264	+0.130	+0.203	+0.176	+0.078	−0.092
	Medium 1	−0.457	+0.022	+0.151	+0.198	+0.210	+0.172
	Medium 2	−0.610	−0.187	−0.066	−0.052	−0.108	−0.158
	Medium 3	**−0.231**	**+0.165**	**+0.233**	+0.213	+0.147	+0.009
	Medium 4	−0.232	+0.111	+0.220	**+0.243**	+0.193	+0.110
	DG-CFG	−0.322	+0.091	+0.209	+0.238	**+0.245**	**+0.244**
Sat	Constant CFG	0.207	0.248	0.299	0.351	0.396	0.431
	Medium 1	0.200	0.213	0.231	0.251	0.271	0.290
	Medium 2	0.198	0.205	0.214	0.226	0.238	0.250
	Medium 3	0.208	0.247	0.294	0.344	0.385	0.417
	Medium 4	0.207	0.243	0.283	0.323	0.359	0.385
	DG-CFG	0.203	0.222	0.247	0.272	0.295	0.315
FID↓	Constant CFG	73.40	**67.45**	70.10	72.63	74.79	79.94
	Medium 1	80.42	68.73	**68.58**	**70.21**	**72.70**	**74.93**
	Medium 2	86.85	74.51	74.71	78.63	84.63	91.22
	Medium 3	**72.39**	68.15	70.25	72.14	75.98	80.73
	Medium 4	72.43	69.04	71.04	73.46	76.65	81.08
	DG-CFG	75.54	68.70	70.08	71.98	74.05	76.43
Div↑	Constant CFG	5.99	5.01	4.81	4.87	5.08	5.33
	Medium 1	6.57	5.38	5.07	4.94	4.94	4.92
	Medium 2	**7.01**	**6.00**	**5.67**	**5.59**	**5.49**	**5.46**
	Medium 3	5.95	4.95	4.81	4.81	5.02	5.21
	Medium 4	6.01	5.03	4.83	4.83	4.93	5.14
	DG-CFG	6.26	5.20	4.97	4.88	4.85	4.83

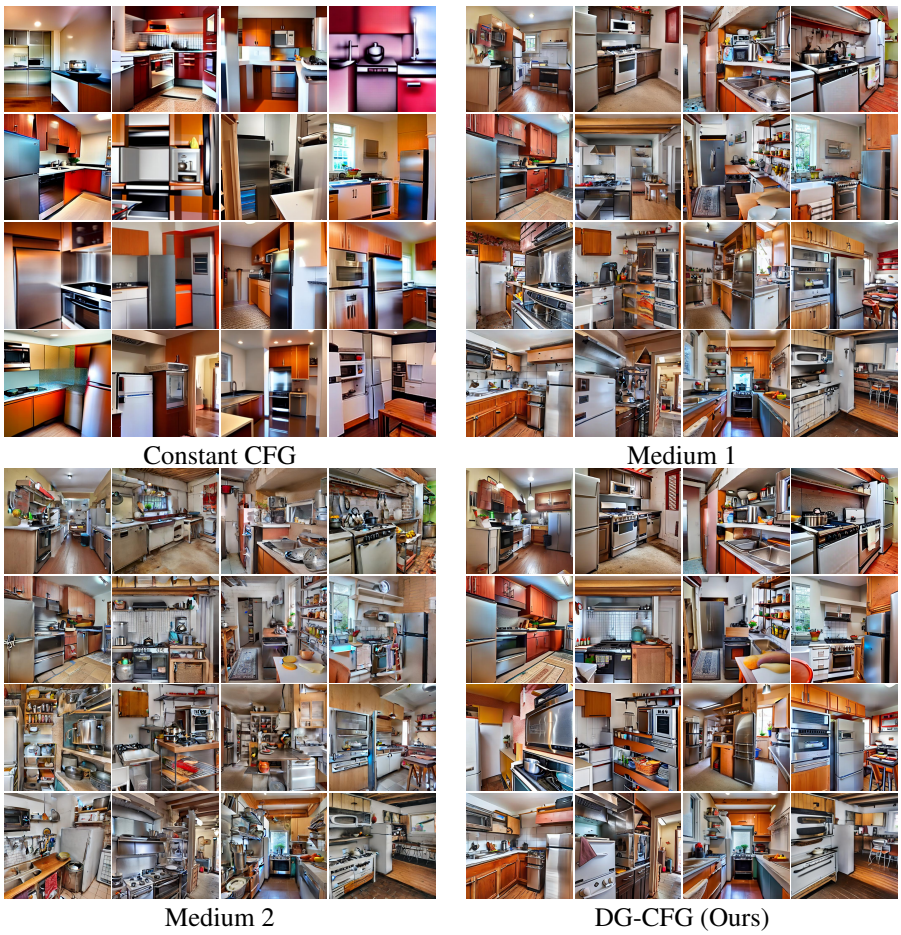


Figure 7: **Qualitative ablation on Stable Diffusion 1.5 at $\bar{\omega} = 23$.** Each quadrant shows 16 samples generated from the same prompt with different random seeds. **Top-left:** constant CFG. **Top-right:** Medium 1 (time balancing). **Bottom-left:** Medium 2 (time balancing plus signal-content weighting). **Bottom-right:** DG-CFG (all three components). The prompt is *a narrow kitchen filled with appliances and cooking utensils*.



Figure 8: **Qualitative ablation on Stable Diffusion 2.1 at $\bar{\omega} = 23$.** Each quadrant shows 16 samples generated from the same prompt with different random seeds. **Top-left:** constant CFG. **Top-right:** Medium 1 (time balancing). **Bottom-left:** Medium 2 (time balancing plus signal-content weighting). **Bottom-right:** DG-CFG (all three components). The prompt is *a narrow kitchen filled with appliances and cooking utensils*.

D.5 Complete Results Across Sampling Budgets

This subsection provides the complete quantitative and qualitative results summarized in Section 5.2.5.

We next examine whether the advantage of DG-CFG persists as the numerical integration becomes more accurate and whether the trend transfers across backbones. Given the cost of repeated sampling, we focus this analysis on SD1.5 and SD2.1. We fix $\bar{\omega} = 23$, the high-guidance regime where color over-saturation is pronounced, and evaluate all four methods across NFE $\in \{4, 8, 16, 32\}$ on both backbones.

Tables 12 and 13 summarize the results. DG-CFG achieves the best ImageReward and lowest saturation in all eight backbone–NFE combinations, and the best FID and CLIP score in seven. The consistent gains across budgets show that DG-CFG is not tied to a particular discretization level.

The raw Div values further underscore the need for a quality-conditioned interpretation of diversity. Several baselines attain higher Div scores than DG-CFG despite substantially worse FID and IR. As NFE increases and these artifacts recede, their Div scores often decline, indicating that low-NFE distortions in global structure and fine detail can artificially disperse feature embeddings and inflate the Vendi score.

The saturation trend is consistent across backbones. At every NFE, DG-CFG moderates the saturation increase induced by strong guidance on both SD1.5 and SD2.1. Constant CFG remains markedly over-saturated, whereas Interval CFG exhibits the most severe saturation at the lowest NFE. The qualitative comparisons in Figure 9 corroborate these results: increasing NFE improves structural quality for all methods, while DG-CFG maintains more natural color and coherent detail across budgets.

Table 12: Evaluation across NFE settings on Stable Diffusion 1.5 at $\bar{\omega} = 23$ using DDIM sampling. The evaluation protocol is given in Appendix D.1.

Metric	Method	NFE$=4$	NFE$=8$	NFE$=16$	NFE$=32$
CLIP↑	Constant	26.79	30.64	31.50	31.58
	Interval CFG	24.84	29.10	31.00	31.45
	β–CFG	26.11	30.88	31.57	31.62
	DG-CFG (ours)	**27.89**	**31.39**	**31.60**	**31.64**
IR↑	Constant	-1.618	-0.330	+0.180	+0.287
	Interval CFG	-1.948	-0.986	-0.128	+0.185
	β–CFG	-1.738	-0.239	+0.227	+0.299
	DG-CFG (ours)	**-1.314**	**+0.029**	**+0.269**	**+0.338**
Sat	Constant	0.408	0.469	0.455	0.450
	Interval CFG	0.500	0.462	0.413	0.395
	β–CFG	0.439	0.438	0.415	0.412
	DG-CFG (ours)	0.348	0.332	0.341	0.356
FID↓	Constant	137.16	82.34	73.00	72.79
	Interval CFG	165.71	110.89	81.61	75.43
	β–CFG	150.37	85.38	73.71	73.45
	DG-CFG (ours)	**133.54**	**74.53**	**70.60**	**69.95**
Div↑	Constant	7.20	5.67	4.80	4.45
	Interval CFG	7.16	6.80	5.77	5.19
	β–CFG	7.16	5.57	4.65	4.53
	DG-CFG (ours)	6.79	5.24	4.76	4.72

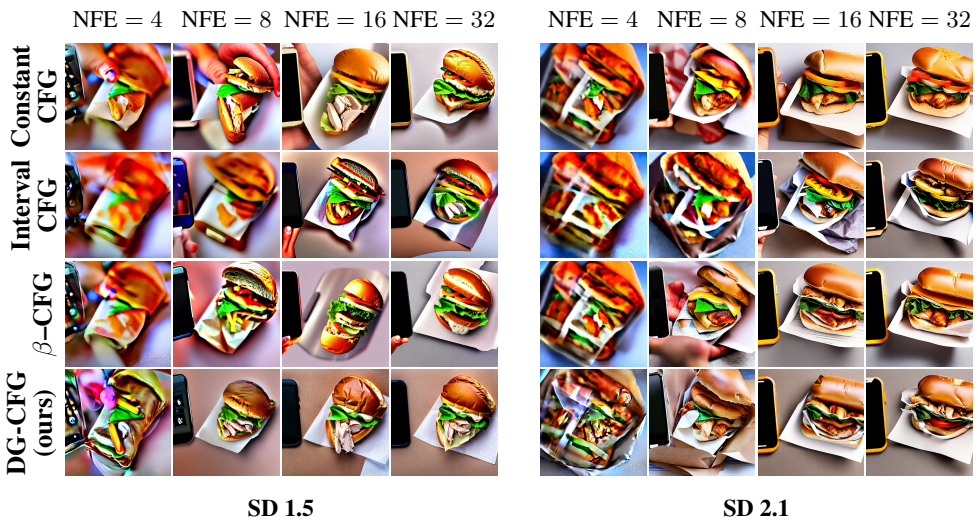


Figure 9: Qualitative comparison across NFE settings on Stable Diffusion 1.5 and Stable Diffusion 2.1 at $\bar{\omega} = 23$. Rows correspond to guidance methods and columns correspond to NFE $\in \{4, 8, 16, 32\}$ within each backbone block. The left block shows SD1.5 and the right block shows SD2.1. The prompt is *a chicken sandwich in a wrapper near a cell phone*.

Table 13: Evaluation across NFE settings on Stable Diffusion 2.1 at $\bar{\omega} = 23$ using DDIM sampling. The evaluation protocol is given in Appendix D.1.

Metric	Method	NFE= 4	NFE= 8	NFE= 16	NFE= 32
CLIP↑	Constant	27.854	30.977	**31.664**	31.656
	Interval CFG	26.460	29.802	31.058	31.444
	β–CFG	27.475	31.141	31.613	31.639
	DG-CFG (ours)	**28.348**	**31.427**	31.655	**31.657**
IR↑	Constant	−1.325	−0.092	+0.409	+0.469
	Interval CFG	−1.687	−0.630	+0.082	+0.364
	β–CFG	−1.466	+0.060	+0.416	+0.516
	DG-CFG (ours)	**−1.095**	**+0.244**	**+0.488**	**+0.536**
Sat	Constant	0.369	0.431	0.413	0.408
	Interval CFG	0.454	0.422	0.382	0.366
	β–CFG	0.395	0.398	0.381	0.381
	DG-CFG (ours)	0.322	0.315	0.327	0.340
FID↓	Constant	**125.43**	79.94	72.12	72.60
	Interval CFG	147.68	105.19	82.19	76.67
	β–CFG	135.48	83.11	74.38	73.91
	DG-CFG (ours)	129.59	**76.43**	**71.93**	**72.38**
Div↑	Constant	6.88	5.33	4.44	4.14
	Interval CFG	7.15	6.38	5.39	4.77
	β–CFG	6.91	5.24	4.40	4.27
	DG-CFG (ours)	6.46	4.83	4.41	4.35