DISENTANGLING DENSITY FROM SEMANTICS IN CLASSIFIER-FREE GUIDANCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Classifier-Free Guidance (CFG) is essential for conditional diffusion and flow-matching models. However, operating at high guidance scales, where prompt alignment is strongest, causes two distinct failure modes: (i) visual oversaturation (color burnout) and (ii) an unintended coupling between prompt alignment and sample density, which forces trajectories into high-density mode centers and suppresses fine structural details. Because the clean image estimate and noise prediction directions are not aligned during generation, previous methods that correct only one of these two directions leave the other distortion uncorrected. To resolve both issues simultaneously, we introduce *Density-Decoupled CFG* (D^2-CFG), a training-free framework that projects guidance updates orthogonally to *both* directions. By decoupling semantic steering from typicality control, D^2-CFG mitigates density drift into mode centers. This preserves the high-dimensional typical-set structure of the latent space, generating highly detailed images without the color oversaturation induced by standard high guidance. This allows increasing guidance without visual burnout and gives independent, explicit control over sample density via a quantile parameter q. Across SDXL and SD3.5, D^2-CFG achieves superior text alignment, natural contrast, and, at the highest guidance scales, improved distribution fidelity.

1 INTRODUCTION

Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021b) have established themselves as the state of the art in generative modeling, achieving remarkable success across applications in the visual domain. Beyond unconditional synthesis, a fundamental practical application lies in conditional generation (e.g., class-conditional and text-to-image synthesis), where faithfully steering generation according to input conditions is essential.

For this purpose, Classifier-Free Guidance (CFG; Ho & Salimans, 2021) has become the standard across modern diffusion and flow-matching models (Rombach et al., 2022; Saharia et al., 2022; Peebles & Xie, 2023): by linearly extrapolating between the conditional and unconditional predictions, CFG improves prompt alignment and visual quality. Despite its universal adoption, CFG exhibits well-known failures at moderate to high guidance scales, where the alignment is maximized (Li et al., 2026): (i) severe visual oversaturation and contrast burnout; and (ii) an unintended entanglement between prompt alignment and trajectory density, where increasing guidance prevents independent, explicit control over image density, and thus of image details (Xu et al., 2026; Karczewski et al., 2025a; Guth et al., 2025).

Historically, these effects have been treated as isolated challenges: for instance, Sadat et al. (2025a); Chung et al. (2025) focus primarily on oversaturation, while Karras et al. (2024) targets better exploration of the image density.

To address these challenges, we analyze the guidance vector associated with CFG from a geometric perspective. For visual oversaturation (i), recent work (Sadat et al., 2025a) showed that guidance acts as an amplitude gain multiplier on the clean image estimate, which can be mitigated by rescaling or eliminating the content-parallel component of the denoiser guidance update. However, this clean-space projection leaves the score-space density drift completely unaddressed. Our key observation for (ii) is that standard guidance simultaneously injects an uncalibrated force along the noise/score direction pushing trajectories into excessively high-density regions and suppressing fine details. We

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

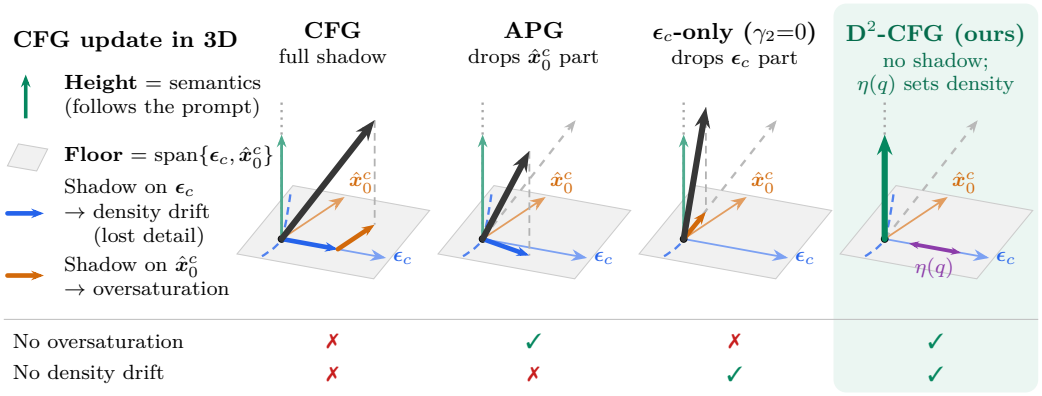


Figure 1: **Overview of D^2-CFG.** The CFG update $w\Delta\boldsymbol{\epsilon}$ has a component orthogonal to the plane $\mathcal{V}_2 = \text{span}\{\boldsymbol{\epsilon}_c, \hat{\boldsymbol{x}}_0^c\}$ (height: semantic steering) and a component inside it (shadow). The shadow along $\hat{\boldsymbol{x}}_0^c$ inflates the amplitude of the clean estimate (oversaturation) and the shadow along $\boldsymbol{\epsilon}_c$ pushes the sample across iso-density contours (dashed; density drift, lost detail). Because the two directions are not orthogonal, removing one component leaves part of the other, so single-axis projections fix one distortion each. D^2-CFG removes the whole shadow and re-assigns the density direction to an explicit dial $\eta(q)$.

show that the score-parallel component of guidance is the only first-order contribution to this unintended density forcing by tracking the log-density evolution along the probability-flow ODE (Song et al., 2021b; Karczewski et al., 2025a). Therefore, because the clean estimate and the score are non-collinear during generation, enforcing either constraint alone is fundamentally incomplete.

Based on this insight, we introduce a joint 2-plane subspace decomposition that yields two disentangled controls: (i) pure semantic steering along the level sets of the data manifold via the guidance weight w, and (ii) an independent dial to control the density of the generated image (modulating image details); see Fig. 1. We refer to our method as Density-Decoupled Classifier-Free Guidance (D^2-CFG).

We validate D^2-CFG across diffusion (SDXL) and flow-matching (SD3.5) models: as a training-free drop-in replacement for CFG, it preserves density controllability, image quality, and prompt alignment at extreme guidance scales without visual burnout.

All in all, the main contributions of this paper are summarized as follows:

- We show that standard CFG couples semantic steering with two geometric distortions: an uncalibrated noise/score-space density drift and a clean-space amplitude gain causing oversaturation.
- We introduce **Density-Decoupled CFG (D^2-CFG)**, where D^2 reflects *dual decoupling*: a training-free projection of the guidance update orthogonally to both the conditional score and the clean image estimate, isolating guidance from both density drift and amplitude blowout.
- For flow-matching models, where the guided norms grow at every step, we complement the projection with a *spherical retraction* that restores the norms of the guided estimates after each step, at no extra network cost.
- We validate on 2D benchmarks, and on MS-COCO and Pick-a-Pic using SDXL and SD3.5, that D^2-CFG scales reliably to high guidance scales, improving alignment and density control without visual burnout.

2 Background and Related Work

2.1 Diffusion Models

The goal of conditional diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021b) is to sample from a target conditional distribution $p(\boldsymbol{x}_0|c)$ on $\mathbb{R}^D$, where $c \sim p(c)$ is a conditioning signal (e.g., a text prompt). For $t \in [0, 1]$, the data $\boldsymbol{x}_0 \sim p_{\text{data}}$ is corrupted by a continuous noising process:

$$p_{t|0}(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}\big(\boldsymbol{x}_t; \alpha_t \boldsymbol{x}_0, \sigma_t^2 \mathbf{I}\big), \tag{1}$$

where $\alpha_t, \sigma_t > 0$ specify the noise schedule with boundary conditions $\alpha_0 = 1, \sigma_0 = 0$ and $\alpha_1 = 0, \sigma_1 = 1$.

Training and sampling. At inference, sampling corresponds to iteratively reversing the noising process in (1) by estimating the clean data posterior expectation $\hat{\boldsymbol{x}}_0 \approx \mathbb{E}[\boldsymbol{x}_0 \mid \boldsymbol{x}_t, c]$. To enable both conditional and unconditional generation, a neural network denoiser $\hat{\boldsymbol{x}}_\theta(\boldsymbol{x}_t, t, c')$ is jointly trained by minimizing the weighted regression loss (Vincent, 2011; Kingma et al., 2021):

$$\mathcal{L}(\theta) \;=\; \mathbb{E}_{t,\boldsymbol{x}_0,\boldsymbol{\epsilon},c'}\Big[\lambda(t)\,\big\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_\theta(\boldsymbol{x}_t, t, c')\big\|^2\Big], \qquad c' \in \{c, \varnothing\}, \tag{2}$$

where $\lambda(t)$ is a noise-dependent weighting that equalizes the loss across noise levels, and conditions are dropped ($c \to \varnothing$) with probability p_{uncond} (Ho & Salimans, 2021). By Tweedie's formula (Raphan & Simoncelli, 2011; Robbins, 1956), the denoiser prediction $\hat{\boldsymbol{x}}_\theta$ and the noise prediction $\boldsymbol{\epsilon}_\theta$ are related via the exact affine correspondence:

$$\hat{\boldsymbol{x}}_0^c \;\coloneqq\; \hat{\boldsymbol{x}}_\theta(\boldsymbol{x}_t, t, c) \;=\; \frac{\boldsymbol{x}_t - \sigma_t\,\boldsymbol{\epsilon}_c}{\alpha_t}, \qquad \text{where } \boldsymbol{\epsilon}_c \;\approx\; -\sigma_t \nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t|c). \tag{3}$$

Similarly, we denote the unconditional denoiser estimate by $\hat{\boldsymbol{x}}_0^{\varnothing} \coloneqq \hat{\boldsymbol{x}}_\theta(\boldsymbol{x}_t, t, \varnothing) = (\boldsymbol{x}_t - \sigma_t \boldsymbol{\epsilon}_u)/\alpha_t$. Once trained, the model draws samples from $\boldsymbol{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ by integrating either the reverse SDE or, as we do throughout, the deterministic probability-flow ODE (Song et al., 2021b) backward from $t = 1$ to 0:

$$d\boldsymbol{x}_t \;=\; \boldsymbol{u}_t(\boldsymbol{x}_t)\,dt, \qquad \boldsymbol{u}_t(\boldsymbol{x}_t) \;\coloneqq\; -\frac{1}{2}\beta(t)\left(\boldsymbol{x}_t - \frac{1}{\sigma_t}\boldsymbol{\epsilon}_c\right), \tag{4}$$

where $\beta(t) \;\coloneqq\; -2\frac{d}{dt}\log\alpha_t$ (VP schedule, $\alpha_t^2 + \sigma_t^2 = 1$). In practice, (4) is discretized with solvers such as Euler, DDIM (Song et al., 2021a), or Heun (Karras et al., 2022), and text-to-image models (Rombach et al., 2022) run both processes in the latent space of a pretrained autoencoder.

2.2 Classifier-Free Guidance and Related Work

To eliminate external classifiers in diffusion guidance (Dhariwal & Nichol, 2021), Ho & Salimans (2021) introduced Classifier-Free Guidance (CFG), which aims to sample from an implicit tilted distribution (at $t = 0$) given by $p_{\tilde{w}}(\boldsymbol{x}_0|c) \propto p(\boldsymbol{x}_0)^{1-\tilde{w}} p(\boldsymbol{x}_0|c)^{\tilde{w}} \propto p(\boldsymbol{x}_0)p(c|\boldsymbol{x}_0)^{\tilde{w}}$, where $\tilde{w} \geq 1$ acts as an inverse temperature on classifier likelihood (Bradley & Nakkiran, 2025). To target this without an auxiliary classifier, CFG linearly extrapolates conditional and unconditional models at inference using[1] $w > 0$:

$$\boldsymbol{\epsilon}_{\text{cfg}} \;=\; \boldsymbol{\epsilon}_c + w\,\Delta\boldsymbol{\epsilon}, \qquad \hat{\boldsymbol{x}}_0^{\text{cfg}} \;=\; \hat{\boldsymbol{x}}_0^c + w\left(\hat{\boldsymbol{x}}_0^c - \hat{\boldsymbol{x}}_0^{\varnothing}\right), \qquad \text{where } \Delta\boldsymbol{\epsilon} \coloneqq \boldsymbol{\epsilon}_c - \boldsymbol{\epsilon}_u. \tag{5}$$

However, diffusing this tilted distribution differs from taking the product of diffused marginals for $t > 0$ (Du et al., 2023; Bradley & Nakkiran, 2025): as Section 3 shows, $w > 0$ in (5) entangles prompt steering with (i) amplitude inflation of $\hat{\boldsymbol{x}}_0^{\text{cfg}}$ (oversaturation) and (ii) uncalibrated density forcing along $\boldsymbol{\epsilon}_c$.

CFG Variants and Oversaturation Control. While numerous works optimize the extrapolation in (5) via dynamic schedules or intervals (Sadat et al., 2024; Zheng & Lan, 2024; Kynkäänniemi et al., 2024; Malarz et al., 2025; Wang et al., 2026), learned weights (Galashov et al., 2026; Pokle et al., 2026), or alternative conditioning signals instead of unconditional evaluation in (5) (Karras et al., 2024; Koulischer et al., 2025; Sadat et al., 2025b), standard guidance fundamentally couples prompt alignment with unintended additional effects, particularly at high scales (Li et al., 2026). Xie et al. (2026) further showed that this coupling distorts evaluation: an inflated effective guidance scale raises preference metrics even as saturation and distribution quality degrade. Several works explored geometric approaches to mitigate these issues, such as magnitude rescaling (CFG-Rescale; Lin et al., 2024), SVD-based tangential damping (TCFG; Kwon et al., 2025), manifold-constrained projections (CFG++; Chung et al., 2025), or clean-denoiser rescaling (APG; Sadat et al., 2025a). However, each targets a single isolated effect, leaving score-space density forcing unaddressed; in contrast, D^2-CFG applies a closed-form projection orthogonal to the joint 2-plane $\mathcal{V}_2 = \text{span}\{\boldsymbol{\epsilon}_c, \hat{\boldsymbol{x}}_0^c\}$, preventing dynamic-range burnout and restoring explicit likelihood control without inflating the effective guidance scale.

[1]Here $w = \tilde{w} - 1$, where $\tilde{w} \geq 1$ denotes the standard guidance scale. All guidance values reported in this paper (e.g., $w = 7.5$ or $w = 15$) follow this convention, i.e., they correspond to a standard scale of $\tilde{w} = w + 1$.

Likelihood Control and Density-Guided Sampling. A complementary line of work seeks to control the likelihood region where generated samples land in (4), which ultimately defines the characteristics of the generated images (Karczewski et al., 2025b; Guth et al., 2025; Zilberstein et al., 2026). For instance, Density-Guided Sampling (DGS; Karczewski et al., 2025a) leverages continuous-time likelihood tracking along probability-flow ODEs to enforce target density rates in unguided models, exploiting the property that, in high dimensions, probability mass concentrates on a thin spherical shell—the typical set—rather than at mode centers (Vershynin, 2019). Related to DGS, TSR (Xu et al., 2026) proposes a rescaled score formulation that enables traversing the density landscape to generate diverse image types; in contrast to DGS, their analysis is based on variance rescaling. Lastly, it is worth mentioning particle-based samplers (Du et al., 2023; Skreta et al., 2025; Hu et al., 2025), which aim to achieve better exploration and diversity during generation. Nevertheless, existing density dials fail when directly applied to CFG in (5) because the rescaled score is fundamentally entangled with the CFG guidance weight, with the parasitic effect growing dramatically at high guidance scales.

3 D^2-CFG: Decoupling Guidance into Independent Geometric Controls

We first address oversaturation (i) by decomposing the guidance update into components parallel and orthogonal to $\hat{\boldsymbol{x}}_0^c$ (Section 3.1); we then show, via the probability-flow ODE, that the score-parallel component alone drives density drift (ii) (Section 3.2); finally, we combine both into the joint 2-plane projection **D^2-CFG** (Section 3.3, Fig. 1).

3.1 Disentangling CFG: The Denoiser Axis

To analyze the visual oversaturation issue (i), we follow Sadat et al. (2025a) and inspect CFG through the clean denoiser perspective. Writing the clean-space update in (5) as $w\Delta\hat{\boldsymbol{x}}_0$ with $\Delta\hat{\boldsymbol{x}}_0 \coloneqq \hat{\boldsymbol{x}}_0^c - \hat{\boldsymbol{x}}_0^{\varnothing}$, and decomposing it into components parallel and perpendicular to $\hat{\boldsymbol{x}}_0^c$, yields:

$$\hat{\boldsymbol{x}}_0^{\text{cfg}} \;=\; \underbrace{\big(1 + w\,\rho_t\big)}_{\text{content amplitude gain}}\;\hat{\boldsymbol{x}}_0^c \;+\; w\,\Delta\hat{\boldsymbol{x}}_{0,\perp}, \tag{6}$$

where $\rho_t \coloneqq \frac{\langle \Delta\hat{\boldsymbol{x}}_0,\, \hat{\boldsymbol{x}}_0^c\rangle}{\|\hat{\boldsymbol{x}}_0^c\|^2}$ is the parallel gain factor, $\Delta\hat{\boldsymbol{x}}_{0,\parallel} = \rho_t \hat{\boldsymbol{x}}_0^c$, and $\Delta\hat{\boldsymbol{x}}_{0,\perp}$ satisfies $\langle \Delta\hat{\boldsymbol{x}}_{0,\perp}, \hat{\boldsymbol{x}}_0^c\rangle = 0$.

Addressing the oversaturation. Because conditional and unconditional clean estimates share a similar coarse composition, their difference is typically positively aligned with the image itself, resulting in $\rho_t > 0$ (Proposition 1). As observed by Sadat et al. (2025a), the prefactor $(1 + w\rho_t)$ in (6) acts as an artificial amplitude multiplier that inflates the dynamic range of $\hat{\boldsymbol{x}}_0^{\text{cfg}}$ as w increases, driving the unnatural contrast and oversaturation artifacts characteristic of high CFG. Downscaling or eliminating the parallel component ($\Delta\hat{\boldsymbol{x}}_0 = \Delta\hat{\boldsymbol{x}}_{0,\perp} + \kappa\,\Delta\hat{\boldsymbol{x}}_{0,\parallel}$ with $\kappa \le 1$) directly alleviates oversaturation. Still, it leaves the entanglement between prompt alignment and trajectory density unaddressed, as we show next.

3.2 Disentangling CFG: The Likelihood Axis

Just as guidance along the clean estimate $\hat{\boldsymbol{x}}_0^c$ inflates signal amplitude, guidance along the score direction $\boldsymbol{\epsilon}_c$ injects an uncalibrated force on sample log-density, entangling prompt steering with probability density drift. Under the probability-flow ODE (4), the instantaneous change of variables theorem (Chen et al., 2018) describes the time evolution of a sample's log-density along a guided velocity field $\tilde{\boldsymbol{u}}_t$ (Karczewski et al., 2025a):

$$\frac{d}{dt}\log p_t(\boldsymbol{x}_t \mid \boldsymbol{c}) \;=\; \underbrace{-\operatorname{div}\boldsymbol{u}_t(\boldsymbol{x}_t)}_{\text{unperturbed rate}} \;+\; \underbrace{\big\langle \nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t \mid \boldsymbol{c}),\; \tilde{\boldsymbol{u}}_t(\boldsymbol{x}_t) - \boldsymbol{u}_t(\boldsymbol{x}_t)\big\rangle}_{\text{guidance-induced density forcing } R}. \tag{7}$$

The forcing term R in (7) is an inner product with the score $-\boldsymbol{\epsilon}_c/\sigma_t$ from (3): only the component of the perturbation along the score changes the log-density. In standard CFG, steering by $w\Delta\boldsymbol{\epsilon}$

injects the velocity perturbation $\tilde{\boldsymbol{u}}_t - \boldsymbol{u}_t = \frac{\beta(t)}{2\sigma_t} w \Delta\boldsymbol{\epsilon}$ (Appendix D), yielding the density forcing rate:

$$R(\boldsymbol{x}_t, t; w) \;=\; -\frac{\beta(t) w}{2\sigma_t^2} \langle \boldsymbol{\epsilon}_c, \Delta\boldsymbol{\epsilon} \rangle. \tag{8}$$

Like the gain factor ρ_t in (6), $\langle \boldsymbol{\epsilon}_c, \Delta\boldsymbol{\epsilon} \rangle$ acts as an uncalibrated density multiplier: $\Delta\boldsymbol{\epsilon}$ is chosen for prompt alignment alone, yet it shifts the density at a rate that scales with w.

Geometric Score Decomposition. Let $\hat{\boldsymbol{e}}_c \coloneqq \boldsymbol{\epsilon}_c / \|\boldsymbol{\epsilon}_c\|$ denote the unit vector along the conditional score direction, and decompose $\Delta\boldsymbol{\epsilon} = \Delta\boldsymbol{\epsilon}_{\parallel} + \Delta\boldsymbol{\epsilon}_{\perp}$, where $\Delta\boldsymbol{\epsilon}_{\parallel} \coloneqq \langle \Delta\boldsymbol{\epsilon}, \hat{\boldsymbol{e}}_c \rangle \hat{\boldsymbol{e}}_c$ and $\langle \Delta\boldsymbol{\epsilon}_{\perp}, \boldsymbol{\epsilon}_c \rangle = 0$. Because $-\boldsymbol{\epsilon}_c$ points along the gradient of steepest probability increase, the parallel component $\Delta\boldsymbol{\epsilon}_{\parallel}$ moves the sample across level sets into higher-density regions than intended, whereas the orthogonal component $\Delta\boldsymbol{\epsilon}_{\perp}$ lies on the tangent plane of the iso-density contours (Fig. 1).

Decoupling Likelihood Control from Prompt Alignment. This guidance–density coupling explains why combining standard CFG with likelihood-regulating methods (Karczewski et al., 2025a; Xu et al., 2026) breaks down at large w. These methods rescale the score by a factor $\eta_t > 0$ that prescribes a target density rate (Appendix D); guiding the rescaled score as $\tilde{\boldsymbol{\epsilon}}_t = \eta_t(\boldsymbol{\epsilon}_c + w\Delta\boldsymbol{\epsilon})$ splits the trajectory density rate into a linear interference relation:

$$\frac{d}{dt} \log p_t(\boldsymbol{x}_t \mid \boldsymbol{c}) \;=\; -\operatorname{div} \boldsymbol{u}_t(\boldsymbol{x}_t) \;+\; \underbrace{R_{\text{target}}(\eta_t)}_{\text{intended density dial}} \;+\; \underbrace{\eta_t \, R(\boldsymbol{x}_t, t; w)}_{\text{CFG interference}}, \tag{9}$$

where $R_{\text{target}}(\eta_t) \coloneqq -\frac{\beta(t)(\eta_t - 1)}{2\sigma_t^2} \|\boldsymbol{\epsilon}_c\|^2$ is the intended likelihood shift. Because $R(\boldsymbol{x}_t, t; w)$ is linear in w, increasing guidance directly corrupts the target density rate unless $\langle \boldsymbol{\epsilon}_c, \Delta\boldsymbol{\epsilon} \rangle = 0$.

Linear diffusion model. We consider a simplified case in Appendix E.1: the class-conditional and unconditional data are each Gaussian, so the Bayes-optimal denoisers are affine (Li et al., 2026). A class differs from the population in two ways, its mean is shifted and its spread is narrower, and each difference produces one distortion. (i) The shifted mean makes the guidance update point along the clean estimate, so guidance inflates the image amplitude: this is the gain ρ_t in (6). (ii) The narrower spread makes the guidance update point along the score, so guidance pulls samples toward the class center: this is the forcing R in (8). Proposition 1 shows the score alignment is positive whenever the class covariance is smaller than the population's and the clean-space alignment is positive for well-separated class means, so both are systematic biases, not fluctuations. The mean term vanishes when the means coincide (up to a non-positive covariance term), the density term when the covariances coincide; projecting onto $\mathcal{V}_2^{\perp}$ removes both.

3.3 D^2-CFG: Synthesizing a Decoupled Guidance

Sections 3.1 and 3.2 show that CFG distorts sampling through two components of the guidance update: (i) its component along the clean estimate $\hat{\boldsymbol{x}}_0^c$, which sets the amplitude gain ρ_t in (6), and (ii) its component along the score $\boldsymbol{\epsilon}_c$, which sets the density forcing R in (8). We therefore seek the guidance vector closest to $\Delta\boldsymbol{\epsilon}$ that cancels both:

$$\boldsymbol{g}^{\star} \;=\; \arg\min_{\boldsymbol{g}} \tfrac{1}{2} \|\boldsymbol{g} - \Delta\boldsymbol{\epsilon}\|^2 \quad \text{s.t.} \quad \langle \boldsymbol{g}, \boldsymbol{\epsilon}_c \rangle = 0, \quad \langle \boldsymbol{g}, \hat{\boldsymbol{x}}_0^c \rangle = 0. \tag{10}$$

The solution is the orthogonal projection of $\Delta\boldsymbol{\epsilon}$ onto the complement of the 2-plane $\mathcal{V}_2 \coloneqq \operatorname{span}\{\boldsymbol{\epsilon}_c, \hat{\boldsymbol{x}}_0^c\}$ (Proposition 2, Appendix G). Note that the plane is determined by the two constraints, and any basis of $\mathcal{V}_2$ yields the same projection.[2] Since $\boldsymbol{\epsilon}_c$ and $\hat{\boldsymbol{x}}_0^c$ are non-collinear for $t > 0$ (Appendix H.1), the two constraints are independent: satisfying one does not satisfy the other. This is why projecting onto a single axis leaves one distortion uncorrected (Appendix C.1).

Closed-form solution. Since $\boldsymbol{\epsilon}_c$ and $\hat{\boldsymbol{x}}_0^c$ are not orthogonal, subtracting the two one-dimensional projections of $\Delta\boldsymbol{\epsilon}$ separately does not yield $\boldsymbol{g}^{\star}$: removing the component along $\hat{\boldsymbol{x}}_0^c$ re-introduces a

[2] By Tweedie's formula, $\mathcal{V}_2 = \operatorname{span}\{\boldsymbol{\epsilon}_c, \boldsymbol{x}_t\}$, the plane of every unguided sampler step (Proposition 3, Appendix G).

component along $\boldsymbol{\epsilon}_c$, proportional to their cosine. We therefore build an orthonormal basis $\{\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2\}$ of $\mathcal{V}_2$ via Gram–Schmidt, orthogonalizing the score direction first:

$$\hat{\boldsymbol{e}}_1 \;\coloneqq\; \frac{\boldsymbol{\epsilon}_c}{\|\boldsymbol{\epsilon}_c\|}, \qquad \hat{\boldsymbol{e}}_2 \;\coloneqq\; \frac{\hat{\boldsymbol{x}}_0^c - \langle \hat{\boldsymbol{x}}_0^c, \hat{\boldsymbol{e}}_1 \rangle \, \hat{\boldsymbol{e}}_1}{\|\hat{\boldsymbol{x}}_0^c - \langle \hat{\boldsymbol{x}}_0^c, \hat{\boldsymbol{e}}_1 \rangle \, \hat{\boldsymbol{e}}_1\|}. \tag{11}$$

The solution of (10) and its relaxation read

$$\Delta\boldsymbol{\epsilon}_\perp^{\text{joint}} \;\coloneqq\; \Delta\boldsymbol{\epsilon} \;-\; \gamma_1 \, \langle \Delta\boldsymbol{\epsilon}, \hat{\boldsymbol{e}}_1 \rangle \, \hat{\boldsymbol{e}}_1 \;-\; \gamma_2 \, \langle \Delta\boldsymbol{\epsilon}, \hat{\boldsymbol{e}}_2 \rangle \, \hat{\boldsymbol{e}}_2, \tag{12}$$

where $\gamma_1, \gamma_2 \in [0, 1]$ and $\gamma_1 = \gamma_2 = 1$ recovers $\boldsymbol{g}^\star$. Because $\hat{\boldsymbol{e}}_2 \perp \boldsymbol{\epsilon}_c$ by construction, fixing $\gamma_1 = 1$ enforces $\langle \boldsymbol{g}, \boldsymbol{\epsilon}_c \rangle = 0$ for any γ_2. Relaxing $\gamma_2 < 1$ replaces the second constraint with a controlled amplitude gain; since amplitude along $\hat{\boldsymbol{x}}_0^c$ governs both saturation and contrast, γ_2 is the sole projection hyperparameter trading off the two, and $\gamma_2 \in [0.4, 0.7]$ balances them on both SDXL and SD3.5 (sweeps in Appendix H; analysis of the relaxation in Appendix G).

The D^2-CFG Guidance Vector. With score-parallel density forcing removed, we apply a density factor $\eta_t(q)$ (Karczewski et al., 2025a) directly to $\boldsymbol{\epsilon}_c$ for explicit likelihood control (Appendix D for details), yielding the unified D^2-CFG prediction noise:

$$\boldsymbol{\epsilon}_{\mathrm{D}^2}(\boldsymbol{x}_t, t, c) \;=\; \eta_t(q) \, \boldsymbol{\epsilon}_c(\boldsymbol{x}_t, t, c) \;+\; w \, \Delta\boldsymbol{\epsilon}_\perp^{\text{joint}}(\boldsymbol{x}_t, t, c), \tag{13}$$

or equivalently in clean Tweedie space, $\hat{\boldsymbol{x}}_0^{\mathrm{D}^2} = \frac{\boldsymbol{x}_t}{\alpha_t} + \eta_t(q)\big(\hat{\boldsymbol{x}}_0^c - \frac{\boldsymbol{x}_t}{\alpha_t}\big) + w\Delta_\perp$ ($\Delta_\perp$: Appendix E). Here $w \geq 0$ controls prompt alignment and $q \in (0, 1)$ is a target density quantile: $\eta_t(q) = 1 + D\sqrt{2/\tilde{D}}\,\Phi^{-1}(q)/\|\boldsymbol{\epsilon}_c\|^2$ rescales the score by $\Phi^{-1}(q)$, where Φ is the CDF of $\mathcal{N}(0, 1)$, the asymptotic law of the normalized log-density rate in high dimension (Appendix D); q is thus the target quantile of that rate (exactly when $\tilde{D} = D$), and $q < 0.5$ lowers the sample density. Here D is the data dimension and $\tilde{D} \leq D$ an effective dimension that handles the growing dimensionality of modern latent spaces: the correction would vanish as D grows, and $\tilde{D}$ keeps the dial numerically effective (Appendix D.2 for details). $\eta_t(q)$ is active for $\mathrm{logSNR}(t) \leq 1$ and 1 otherwise. Alg. 1 summarizes the procedure; the construction extends to CFG++, where it cancels the score-parallel leak exactly (Appendix I).

The two constraints fix the *direction* of the update but not the *length* of the guided estimates: (10) controls the two norms only to first order. Expanding (13) at $\boldsymbol{g}^\star$ ($\gamma_2 = 1, \eta_t = 1$),

$$\|\boldsymbol{\epsilon}_{\mathrm{D}^2}\|^2 = \|\boldsymbol{\epsilon}_c\|^2 + w^2\|\Delta\boldsymbol{\epsilon}_\perp^{\text{joint}}\|^2, \qquad \|\hat{\boldsymbol{x}}_0^{\mathrm{D}^2}\|^2 = \|\hat{\boldsymbol{x}}_0^c\|^2 + \big(\tfrac{\sigma_t}{\alpha_t}\big)^2 w^2 \|\Delta\boldsymbol{\epsilon}_\perp^{\text{joint}}\|^2, \tag{14}$$

where the cross terms vanish by (10), and a quadratic term remains that no choice of basis or γ removes, since it depends only on the length of the update.

Flow matching: the spherical retraction. By (3), a change in the noise prediction moves the clean estimate by σ_t/α_t times as much; this is the prefactor in (14), and in diffusion it vanishes as $t \to 0$, so the clean estimate stops inflating at low noise; empirically, the projection alone suffices there (Section 4.2). In flow matching (with $t = 0$ at noise and $t = 1$ at data) the two endpoints are $\boldsymbol{x}_t + (1 - t)\boldsymbol{v}$ and $\boldsymbol{x}_t - t\boldsymbol{v}$, so a change in $\boldsymbol{v}$ moves them by $(1 - t)$ and t times as much; these coefficients sum to one and never vanish together, so at every Euler step one of the two endpoints inflates by a term of order w^2, which accumulates; empirically, this makes the retraction necessary (Table 9). We therefore complete the projection with a *spherical retraction*: after each update, we rescale the guided clean and noise estimates back to their target norms, $\|\hat{\boldsymbol{x}}_0^c\|$ and $\eta_t(q)\|\boldsymbol{\epsilon}_c\|$, which removes the quadratic term in (14) exactly at no extra network cost (Appendices F and G).

4 Numerical Experiments

We ask (a) whether D^2-CFG decouples *prompt alignment* from *log-density control*, and (b) whether, at high guidance ($w \geq 10$), it improves the FID–CLIP frontier while preventing oversaturation.

4.1 Validation on an Exact-Likelihood 2D Benchmark

We evaluate on the 2D fractal benchmark from Karras et al. (2024), a 1016-component Gaussian mixture with analytical density p_{gt} whose classes (A, B) share a trunk but diverge at sub-branches

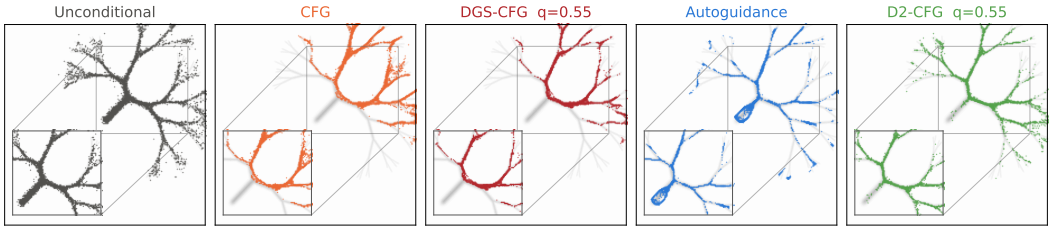


Figure 2: **Class-A samples at high guidance** ($w = 4$**;** ground truth in gray; insets zoom into shared trunk). CFG scatters off the manifold, DGSCFG drops modes, and autoguidance requires an auxiliary network. D^2-CFG stays on the manifold and preserves broad coverage using only CFG's native signal.

(details in Appendix C.3). In 2D the plane $\mathcal{V}_2$ is all of $\mathbb{R}^2$, so the toys use the score-axis projection alone ($\gamma_1 = 1$, $\gamma_2 = 0$). Standard CFG over-truncates this distribution and drops ground-truth modes, which Karras et al. (2024) mitigate with an auxiliary lower-capacity model; our approach needs no extra network. At $w = 4$ (Fig. 2), D^2-CFG stays on the class-A manifold while preserving broad mode coverage. We provide a quantitative evaluation sweeping w and q (along with an additional 6-mode Gaussian mixture benchmark) in Appendices C.3 and C.4.

4.2 Text-to-Image Generation

Setup and metrics. We evaluate on SDXL (Podell et al., 2024) (30-step DDIM) for diffusion and SD3.5-Medium (Esser et al., 2024) (28-step Euler ODE) for flow matching across 5k prompt-paired images from MS-COCO 2014 (Lin et al., 2014) and Pick-a-Pic (Kirstain et al., 2023). Throughout all evaluations, D^2-CFG fixes $\gamma_1 = 1$ and varies only γ_2, using $q = 0.30$ by default to preserve fine textures on complex scenes (q-sweeps in Section 4.2.2). We report FID (Heusel et al., 2017), CLIP Score (Hessel et al., 2021), saturation (Sadat et al., 2025a), k-NN Precision/Recall, and preference proxies (ImageReward, PickScore, HPSv2; Appendix C).

Table 1: **Quantitative evaluation on MS-COCO 5k with fixed $q = 0.30$ (SDXL; $\gamma_1 = 1$). Best** and second-best per guidance scale. At $w = 20.0$, D^2-CFG achieves the best FID, CLIP, Precision, and ImageReward ($\gamma_2 = 0.4$) and the best Recall ($\gamma_2 = 1.0$), outperforming DGSAPG at matched $\gamma_2 = 1.0$ on all but saturation.

Scale	Method	FID ↓	CLIP ↑	Prec. ↑	Rec. ↑	Sat. ↓	ImageReward ↑
$w = 7.5$	Standard CFG	**23.23**	27.00	**0.735**	0.615	22.49%	0.7981
	DGSCFG	23.30	26.98	**0.735**	0.616	23.04%	**0.8120**
	TSR	23.28	27.00	0.731	0.616	22.50%	0.7988
	DGSAPG ($\gamma_2 = 1.0$)	23.79	**27.06**	0.723	0.610	**18.14%**	0.7539
	DGSAPG ($\gamma_2 = 0.4$)	23.34	27.01	**0.735**	0.616	21.48%	0.8045
	D^2-CFG ($\gamma_2 = 1.0$, Ours)	23.78	27.04	0.720	0.611	18.62%	0.7672
	D^2-CFG ($\gamma_2 = 0.4$, Ours)	23.35	27.00	0.731	**0.628**	21.87%	0.8110
$w = 20.0$	Standard CFG	25.29	27.29	0.683	0.596	32.44%	0.8838
	DGSCFG	25.08	27.31	0.692	0.593	32.90%	0.8847
	TSR	25.42	27.29	0.693	0.593	32.42%	0.8826
	DGSAPG ($\gamma_2 = 1.0$)	25.23	27.27	0.689	0.599	**28.00%**	0.8814
	DGSAPG ($\gamma_2 = 0.4$)	25.32	27.31	0.678	0.594	31.12%	0.8850
	D^2-CFG ($\gamma_2 = 1.0$, Ours)	24.81	27.32	0.696	**0.605**	28.34%	0.8934
	D^2-CFG ($\gamma_2 = 0.4$, Ours)	**24.79**	**27.34**	**0.706**	0.591	31.40%	**0.8994**

4.2.1 Comparison with Baselines

We evaluate D^2-CFG against standard CFG, Density-Guided Sampling (DGS) (Karczewski et al., 2025a) applied directly to CFG (DGSCFG), DGSAPG, a baseline combining DGS with APG (Sadat et al., 2025a), and TSR (Xu et al., 2026) across varying guidance regimes; we also include CFG-Zero* (Fan et al., 2025) for flow matching. We also compare against TCFG (Kwon et al., 2025) in Appendix C.1 (Fig. 10); since it damps the tangential part of the guidance update and therefore operates at a different effective guidance scale, we compare it along the FID–CLIP frontier rather than at matched w. DGSAPG applies APG's projection along $\hat{\boldsymbol{x}}_0^c$ (without its momentum and norm threshold) with the same γ_2 as D^2-CFG and no score-space projection; TSR uses its defaults ($k = 0.93$, $\sigma = 3.0$).

SDXL. At moderate guidance ($w = 7.5$), all methods yield comparable FID vs. CLIP, with standard CFG achieving the best FID (Fig. 3 and Table 1). Full clean-space projection ($\gamma_2 = 1.0$) strongly suppresses oversaturation in both D^2-CFG and DGSAPG, while $\gamma_2 = 0.4$ balances CLIP, contrast, and mode coverage (Recall) at a similar FID (more w's in Table 2, Appendix C.1).

Human preference rises with guidance up to $w \approx 20$ for every method (ImageReward from ≈ 0.80 at $w = 7.5$ to ≈ 0.88–0.90 at $w = 15$–20) and falls at $w = 25$ (Table 2). These automated proxies are biased toward saturated images (Xie et al., 2026); our gains coincide with *lower* saturation (Fig. 3),

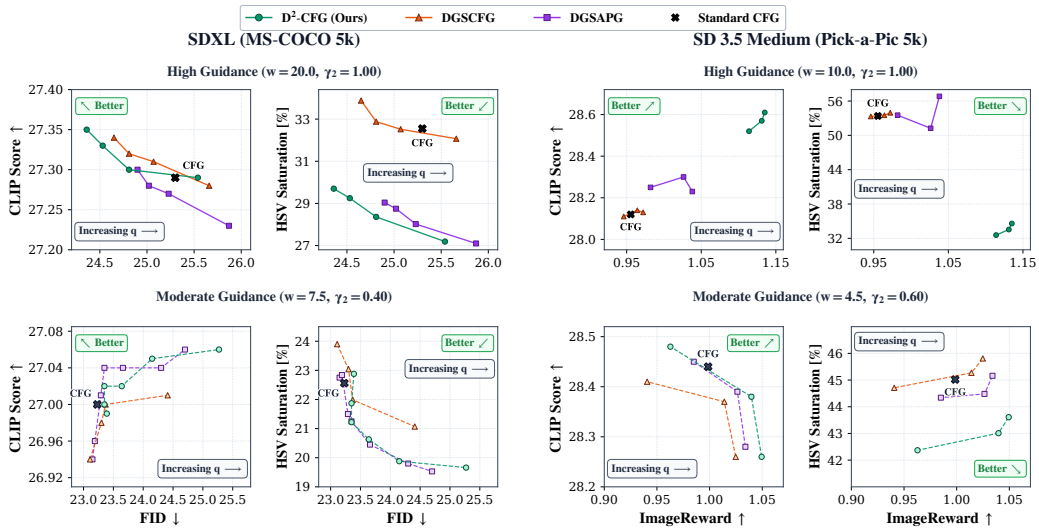


Figure 3: **Guidance Pareto frontiers across density quantiles** q on SDXL (MS-COCO 5k, left) and SD 3.5 Medium (Pick-a-Pic 5k, right) at high (top, solid) and moderate guidance (bottom, dashed); w and γ_2 as in the panel titles, matched between D^2-CFG and DGSAPG (w and the best γ_2 differ between the two models: flow-matching models operate at smaller guidance scales and prefer $\gamma_2 = 0.6$ over 0.4); $q \in [0.05, 0.95]$. At high guidance, D^2-CFG extends the FID–CLIP and saturation frontiers on SDXL and dominates on SD 3.5; baselines without retraction (Table 9).

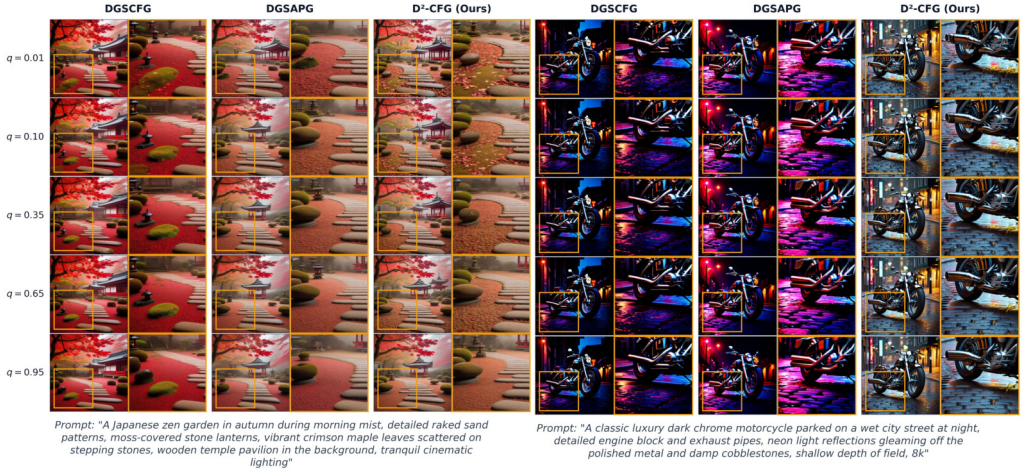


Figure 4: **Qualitative comparison across density quantiles** q **at high guidance** ($w = 15.0$**; SDXL left, SD3.5 right).** For each prompt, we compare DGSCFG, DGSAPG, and D^2-CFG across $q \in \{0.01, 0.10, 0.35, 0.65, 0.95\}$. Lower q adds detail and complexity to the background.

the opposite pattern. Standard CFG, however, pays for this gain: at $w \geq 15$, unconstrained latent growth causes over-saturation and drops in precision and recall, and DGSCFG inherits the saturation and degraded FID because it lacks clean-space orthogonalization. D^2-CFG retains the preference gain (best ImageReward at every $w \geq 10$) without this cost.

Crucially, DGSAPG matches D^2-CFG on saturation and CLIP at matched γ_2, but leaves ϵ_c uncorrected, so its density dial drifts toward higher density; removing this interference ($\gamma_1 = 1$) is what lets D^2-CFG outperform it in FID, CLIP, Precision, Recall, and ImageReward at high guidance ($w \geq 15$) for $\gamma_2 = 1.0$, and in all but Recall at $w = 20$ for $\gamma_2 = 0.4$ (Table 1, Fig. 3; seeds in Tables 5, 7).

SD3.5. Flow-matching baselines degrade as guidance grows (Fig. 3, right; Tables 6–8): at $w = 10$ CFG, DGSCFG, and DGSAPG already reach $\approx 50\%$ saturation or more, and at $w = 15$ they collapse to ImageReward ≤ 0.71, the undamped norm growth of (14). D^2-CFG stays stable across the whole q range and reaches ImageReward $= 1.17$ at $w = 15$ with $\gamma_2 = 0.6$, at lower saturation than every baseline. Table 9 (Appendix F) separates the two ingredients: the retraction alone lifts

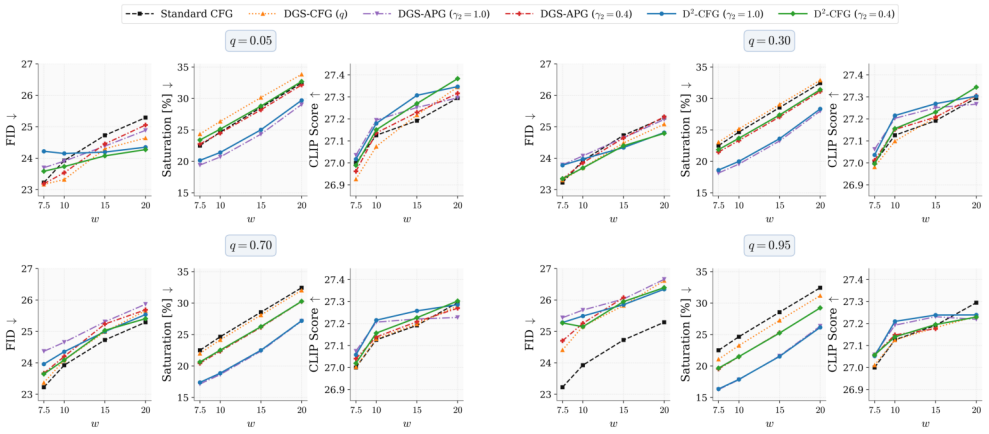


Figure 5: **Guidance sweeps across quantiles q on MS-COCO 5k (SDXL).** Across all regimes, D^2-CFG suppresses saturation blowout, with $\gamma_2 = 0.4$ (green) optimizing FID and CLIP at high guidance ($w = 20$).

standard CFG to 1.12, and the projection adds a consistent gain on all five metrics on top of it, whereas without retraction every variant burns out. On MS-COCO, $\gamma_2 = 1.0$ gives the best CLIP and saturation at every w and $\gamma_2 = 0.4$ the best FID at $w \geq 10$; at $w = 7.5$–10, D^2-CFG trails DGSCFG by 0.6–1.7 FID (at $w = 4.5$ it is 0.1–0.3 ahead). The choice of γ_2 matters more on SD3.5 than on SDXL: at $w = 15$, $\gamma_2 = 1.0$ removes too much contrast (FID 33.3), whereas $\gamma_2 \in \{0.4, 0.6\}$ gives 27.5–28.3.

4.2.2 Ablation over q

We examine the interaction between w and q in Figs. 4 and 5 (more in Appendices B and H). Qualitatively (Fig. 4) and quantitatively (Fig. 5), traversing $q \in [0.01, 0.95]$ suggests a *perception-density* relationship: mode-center trajectories ($q = 0.95$) yield smooth, low-variance images free of color blowout but oversmooth fine textures (elevating FID), whereas lower-density shells ($q \leq 0.30$) synthesize rich high-frequency detail, optimizing both FID and CLIP when paired with $\gamma_2 = 0.4$.

5 Discussion

In this work, we introduced **D^2-CFG**, a training-free inference framework that enables operating in high-guidance regimes ($w \geq 10$) across 2D synthetic distributions, SDXL, and SD3.5, achieving stronger prompt alignment while reducing dynamic-range oversaturation and preserving explicit control over sample log-density.

Geometrically, our method unifies two previously separate threads in the CFG literature (Section 2.2). Standard CFG introduces two entangled geometric failure modes: a score-parallel density drift toward uncalibrated density regions, and a clean-space amplitude gain that drives saturation. Existing methods address these challenges only in isolation: clean projections (e.g., APG) leave score drift uncorrected, while score-space scaling (e.g., DGS, TSR) leaves dynamic range blowout untouched, and naively combining both single-axis corrections (DGSAPG) leaves a residual score-parallel leak because $\hat{\boldsymbol{x}}_0^c$ and $\boldsymbol{\epsilon}_c$ are non-collinear. In contrast, D^2-CFG eliminates both undesired effects simultaneously via a closed-form projection onto the orthogonal complement of the joint 2-plane $\mathcal{V}_2 = \mathrm{span}\{\boldsymbol{\epsilon}_c, \hat{\boldsymbol{x}}_0^c\}$, leaving guidance scale w and density quantile q as two decoupled knobs governing prompt alignment and density control.

Limitations and Future Work. We target undistilled diffusion and flow models at moderate to high guidance, where the two distortions of CFG are largest; *guidance-distilled students* are a natural extension. We evaluate with standard automated metrics and preference proxies, which a human study would complement. We use a constant γ_2 along the trajectory; a step-dependent schedule that keeps the full projection during the early, layout-setting steps and relaxes it at low noise, where the amplitude term sets contrast, is a natural extension. Since the projection changes the guidance direction, D^2-CFG samples from a distribution other than the tilted one of standard CFG; characterizing it is an interesting direction. Finally, image restoration is a natural application, connecting density control with the perception-distortion trade-off.

AI USE STATEMENT

In this work, we used generative AI tools (large language models) for the following tasks with required disclosure: providing feedback on the experimental methodology (suggesting ablations, baselines, and seed-variation checks, which the authors then designed, ran, and selected from), and verifying the proofs for propositions whose statements and arguments were formulated by the authors (Appendix G). We have not used generative AI tools to generate synthetic datasets, to develop the theoretical framework, to formulate mathematical claims, to propose or refine hypotheses, to implement methods, or to interpret results; the remaining required-disclosure tasks (translation, dataset cleaning and reformatting, qualitative or thematic data analysis) are not applicable to this work. Additionally, we used generative AI tools for editing the paper to improve readability, drafting parts of the text, creating and modifying schematic figures, and checking the consistency of reported numbers between tables and text throughout the paper. We have reviewed all AI-assisted work: every proof was checked line by line by the authors, all AI-edited or AI-drafted text was read, revised, and approved by the authors, all reported numbers were verified against the experiment logs, and all interpretations of results are the authors' own. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work proposes a training-free modification on conditional sampling for pretrained text-to-image models. It involves no human subjects, introduces no new dataset, and uses only publicly available models (SDXL, SD3.5-Medium) and datasets (MS-COCO, Pick-a-Pic, ImageNet). As with any guidance method, D^2-CFG inherits the biases and potential for misuse of the underlying generative models, and it does not add safeguards beyond those of the models themselves. We are not aware of conflicts of interest related to this work.

REPRODUCIBILITY STATEMENT

The method is fully specified by Eqs. (11)– (13) and by Algorithms 1 (diffusion) and 2 (flow matching) in Appendices A and F. All hyperparameters used in the experiments (q, γ_1, γ_2, the effective dimension $\tilde{D}$, the logSNR gate, samplers, and step counts) are stated in Section 4.2 and in the "Models" paragraph of Appendix A. The 2D benchmarks are described in Appendices C.3 and C.4, including the network, training, and integration settings. Complete proofs and assumptions for all propositions are given in Appendices D, E.1, G, and I. Seed variation for the main comparisons is reported in Tables 5 and 7.

REFERENCES

P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *J. Mach. Learn. Res.*, 2025.

Arwen Bradley and Preetum Nakkiran. Classifier-free guidance is a predictor-corrector. *Trans. Mach. Learn. Res (TMLR).*, 2025.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in Neural Inf. Process. Syst. (NIPS)*, 31, 2018.

Hyungjin Chung, Jeongsol Kim, Geon Yeong Park, Hyelin Nam, and Jong Chul Ye. CFG++: Manifold-constrained classifier free guidance for diffusion models. In *Intl. Conf. Learn. Repr. (ICLR)*, volume 2025, pp. 30824–30850, 2025.

Peter De Jong. A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, 75(2):261–277, 1987.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. *Advances in Neural Inf. Process. Syst. (NIPS)*, 34:8780–8794, 2021.

Yilun Du, Conor Durkan, Robin Strudel, Joshua B Tenenbaum, Sander Dieleman, Rob Fergus, Jascha Sohl-Dickstein, Arnaud Doucet, and Will Sussman Grathwohl. Reduce, reuse, recycle: Compositional generation with energy-based diffusion models and MCMC. In *Intl. Conf. on Mach. Learn. (ICML)*, pp. 8489–8510. PMLR, 2023.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Intl. Conf. on Mach. Learn. (ICML)*, 2024.

Weichen Fan, Amber Yijia Zheng, Raymond A Yeh, and Ziwei Liu. CFG-zero*: Improved classifier-free guidance for flow matching models. *arXiv preprint arXiv:2503.18886*, 2025.

Alexandre Galashov, Ashwini Pokle, Arnaud Doucet, Arthur Gretton, Mauricio Delbracio, and Valentin De Bortoli. Learn to guide your diffusion model. *Intl. Conf. Learn. Repr. (ICLR)*, 2026.

Florentin Guth, Zahra Kadkhodaie, and Eero P Simoncelli. Learning normalized image densities via dual score matching. *Advances in Neural Inf. Process. Syst. (NIPS)*, 2025.

Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. CLIPScore: A reference-free evaluation metric for image captioning. In *Proceedings of the 2021 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 2021.

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 Inf. Process. Syst. (NIPS)*, 30, 2017.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Inf. Process. Syst. (NIPS)*, 33:6840–6851, 2020.

Yuyang Hu, Kangfu Mei, Mojtaba Sahraee-Ardakan, Ulugbek S Kamilov, Peyman Milanfar, and Mauricio Delbracio. Kernel density steering: Inference-time scaling via mode seeking for image restoration. *Advances in Neural Inf. Process. Syst. (NIPS)*, 2025.

Rafał Karczewski, Markus Heinonen, and Vikas Garg. Devil is in the details: Density guidance for detail-aware generation with flow models. In *Intl. Conf. on Mach. Learn. (ICML)*, 2025a.

Rafał Karczewski, Markus Heinonen, and Vikas Garg. Diffusion models as cartoonists: The curious case of high density regions. In *Intl. Conf. Learn. Repr. (ICLR)*, volume 2025, pp. 12882–12915, 2025b.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in Neural Inf. Process. Syst. (NIPS)*, 35:26565–26577, 2022.

Tero Karras, Miika Aittala, Tuomas Kynkäänniemi, Jaakko Lehtinen, Janne Hellsten, and Timo Aila. Guiding a diffusion model with a bad version of itself. In *Advances in Neural Inf. Process. Syst. (NIPS)*, 2024.

Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in Neural Inf. Process. Syst. (NIPS)*, 34:21696–21707, 2021.

Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. *Advances in Neural Inf. Process. Syst. (NIPS)*, 2023.

Felix Koulischer, Johannes Deleu, Gabriel Raya, Thomas Demeester, and Luca Ambrogioni. Dynamic negative guidance of diffusion models. In *Intl. Conf. Learn. Repr. (ICLR)*, 2025.

Mingi Kwon, Jaeseok Jeong, Yi Ting Hsiao, Youngjung Uh, et al. TCFG: Tangential damping classifier-free guidance. In *Proceedings of the IEEE/CVF Int. Conf. Comput. Vis. Pattern Recogn. (CVPR)*, 2025.

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. In *Advances in Neural Inf. Process. Syst. (NIPS)*, 2024.

Xiang Li, Rongrong Wang, and Qing Qu. Towards understanding the mechanisms of classifier-free guidance. *Advances in Neural Inf. Process. Syst. (NIPS)*, 2026.

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)*, 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 Conf. Comp. Vision (ECCV)*, pp. 740–755. Springer, 2014.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *Intl. Conf. Learn. Repr. (ICLR)*, 2023.

Dawid Malarz, Artur Kasymov, Maciej Zięba, Jacek Tabor, and Przemysław Spurek. Classifier-free guidance with adaptive scaling. *Frontiers in Artificial Intelligence and Applications (ECAI 2025)*, 413:4119–4126, 2025. doi: 10.3233/FAIA250836.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pp. 4172–4182. IEEE, 2023.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *Intl. Conf. Learn. Repr. (ICLR)*, volume 2024, pp. 1862–1874, 2024.

Ashwini Pokle, Alexandre Galashov, Arnaud Doucet, Mauricio Delbracio, and Valentin De Bortoli. Adversarial learning of classifier-free guidance schedules. *arXiv preprint arXiv:2608.14038*, 2026.

Martin Raphan and Eero P Simoncelli. Least squares estimation without priors or supervision. *Neural Comp.*, 23(2):374–420, 2011.

Herbert E Robbins. An empirical Bayes approach to statistics. In *Proc. Third Berkeley Symp. Math. Statist. Prob.*, pp. 388–394. Springer, 1956.

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 Int. Conf. Comput. Vis. Pattern Recogn. (CVPR)*, pp. 10674–10685, 2022.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *Int. J. Comp. Vision*, 115:211–252, 2015.

Seyedmorteza Sadat, Jakob Buhmann, Derek Bradley, Otmar Hilliges, and Romann Weber. Cads: Unleashing the diversity of diffusion models through condition-annealed sampling. In *Intl. Conf. Learn. Repr. (ICLR)*, volume 2024, pp. 23723–23755, 2024.

Seyedmorteza Sadat, Otmar Hilliges, and Romann M Weber. Eliminating oversaturation and artifacts of high guidance scales in diffusion models. In *Intl. Conf. Learn. Repr. (ICLR)*, 2025a.

Seyedmorteza Sadat, Manuel Kansy, Otmar Hilliges, and Romann Weber. No training, no problem: Rethinking classifier-free guidance for diffusion models. In *Intl. Conf. Learn. Repr. (ICLR)*, volume 2025, pp. 76833–76858, 2025b.

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 Inf. Process. Syst. (NIPS)*, 35:36479–36494, 2022.

Marta Skreta, Tara Akhound-Sadegh, Viktor Ohanesian, Roberto Bondesan, Alán Aspuru-Guzik, Arnaud Doucet, Rob Brekelmans, Alexander Tong, and Kirill Neklyudov. Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. *Intl. Conf. on Mach. Learn. (ICML)*, 2025.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Intl. Conf. on Mach. Learn. (ICML)*, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *Intl. Conf. Learn. Repr. (ICLR)*, 2021a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *Intl. Conf. Learn. Repr. (ICLR)*, 2021b.

Roman Vershynin. High-dimensional probability. *Cambridge Series in Statistical and Probabilistic Mathematics*, 47, 2019.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

Hanyang Wang, Yiyang Liu, Jiawei Chi, Fangfu Liu, Ran Xue, and Yueqi Duan. CFG-Ctrl: Control-based classifier-free diffusion guidance. *Proceedings of the IEEE/CVF Int. Conf. Comput. Vis. Pattern Recogn. (CVPR)*, 2026.

Dian Xie, Shitong Shao, Lichen Bai, Zikai Zhou, Bojun Cheng, Shuo Yang, Jun Wu, and Zeke Xie. Guidance matters: Rethinking the evaluation pitfall for text-to-image generation. In *Intl. Conf. Learn. Repr. (ICLR)*, 2026.

Yanbo Xu, Yu Wu, Sungjae Park, Zhizhuo Zhou, and Shubham Tulsiani. Temporal score rescaling for temperature sampling in diffusion and flow models. *Intl. Conf. on Mach. Learn. (ICML)*, 2026.

Candi Zheng and Yuan Lan. Characteristic guidance: Non-linear correction for diffusion model at large guidance scale. *Intl. Conf. on Mach. Learn. (ICML)*, 2024.

Nicolas Zilberstein, Santiago Segarra, Eero Simoncelli, and Florentin Guth. Learning normalized energy models for linear inverse problems. *Intl. Conf. on Mach. Learn. (ICML)*, 2026.

APPENDIX CONTENTS

A Algorithm and Implementation Details **14**

B Additional Qualitative Results **15**

C Extended Quantitative Evaluations and 2D Benchmarks **18**

D Continuous Likelihood Tracking and the Density Dial **25**

E Denoiser Perspective of D^2-CFG **26**

F Flow Matching Formulation of D^2-CFG **29**

G Why the 2-Plane: Constraints, Uniqueness, and Relaxation **32**

H Additional Ablations and Parameter Sweeps **33**

I D^2-CFG++ **41**

A ALGORITHM AND IMPLEMENTATION DETAILS

We summarize the pseudocode of our method in Alg. 1.

Algorithm 1 D^2-CFG Sampling Algorithm (Probability-Flow ODE / DDIM)

Require: Conditional and unconditional score networks $\boldsymbol{\epsilon}_\theta(\cdot, t, c)$ and $\boldsymbol{\epsilon}_\theta(\cdot, t, \varnothing)$; noise schedule $\{\alpha_t, \sigma_t\}_{t=0}^{T}$; guidance scale $w \geq 0$; density quantile $q \in (0, 1)$; effective dimension $\tilde{D}$ (D for SDXL); clean-space projection weight $\gamma_2 \in [0, 1]$ (with $\gamma_1 = 1$ fixed).

1: Initialize latent state $\boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.
2: **for** $k = T$ **downto** 1 **do**
3: Evaluate conditional and unconditional noise predictions: $\boldsymbol{\epsilon}_c \leftarrow \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_{t_k}, t_k, c), \quad \boldsymbol{\epsilon}_u \leftarrow \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_{t_k}, t_k, \varnothing)$
4: Compute clean image estimate and guidance difference: $\hat{\boldsymbol{x}}_0^c \leftarrow \frac{\boldsymbol{x}_{t_k} - \sigma_{t_k}\boldsymbol{\epsilon}_c}{\alpha_{t_k}}, \quad \Delta\boldsymbol{\epsilon} \leftarrow \boldsymbol{\epsilon}_c - \boldsymbol{\epsilon}_u$
5: Construct orthonormal basis for $\mathcal{V}_2 = \mathrm{span}\{\boldsymbol{\epsilon}_c, \hat{\boldsymbol{x}}_0^c\}$ via Gram–Schmidt (score direction first; the order matters for $\gamma_2 < 1$):
6: $\hat{\boldsymbol{e}}_1 \leftarrow \frac{\boldsymbol{\epsilon}_c}{\|\boldsymbol{\epsilon}_c\|}, \qquad \tilde{\boldsymbol{e}}_2 \leftarrow \hat{\boldsymbol{x}}_0^c - \langle \hat{\boldsymbol{x}}_0^c, \hat{\boldsymbol{e}}_1\rangle \hat{\boldsymbol{e}}_1, \qquad \hat{\boldsymbol{e}}_2 \leftarrow \frac{\tilde{\boldsymbol{e}}_2}{\|\tilde{\boldsymbol{e}}_2\|}$
7: Project guidance update onto $\mathcal{V}_2^\perp$ ($\gamma_1 = 1$): $\Delta\boldsymbol{\epsilon}_\perp^{\mathrm{joint}} \leftarrow \Delta\boldsymbol{\epsilon} - \langle \Delta\boldsymbol{\epsilon}, \hat{\boldsymbol{e}}_1\rangle \hat{\boldsymbol{e}}_1 - \gamma_2 \langle \Delta\boldsymbol{\epsilon}, \hat{\boldsymbol{e}}_2\rangle \hat{\boldsymbol{e}}_2$
8: Compute density scale factor: $\eta_{t_k}(q) \leftarrow 1 + \frac{D\sqrt{2/\tilde{D}}\,\Phi^{-1}(q)}{\|\boldsymbol{\epsilon}_c\|^2}$ if $\mathrm{logSNR}(t_k) \leq 1$, else $\eta_{t_k}(q) \leftarrow 1$
9: Build guided noise prediction and step ODE: $\boldsymbol{\epsilon}_{\mathrm{D}^2} \leftarrow \eta_{t_k}(q)\boldsymbol{\epsilon}_c + w\,\Delta\boldsymbol{\epsilon}_\perp^{\mathrm{joint}}, \quad \boldsymbol{x}_{t_{k-1}} \leftarrow \mathrm{ODEStep}(\boldsymbol{x}_{t_k}, \boldsymbol{\epsilon}_{\mathrm{D}^2})$
10: **end for**
11: **return** Generated sample $\boldsymbol{x}_0$.

Models. SDXL-base-1.0 is a 2.6B UNet on $4 \times 128 \times 128$ latents ($D = 65{,}536$), sampled with 30 DDIM (Song et al., 2021a) steps over $w \in [7.5, 25.0]$; SD3.5 Medium is a 2B MMDiT on $16 \times 128 \times 128$ latents ($D = 262{,}144$), sampled with 28 Euler steps over $w \in [4.5, 15.0]$ ($w = 20$ in Fig. 16) at 1024×1024 resolution. Per-step Gram–Schmidt orthogonalization requires two inner products in latent space, adding negligible runtime and memory overhead.

B Additional Qualitative Results

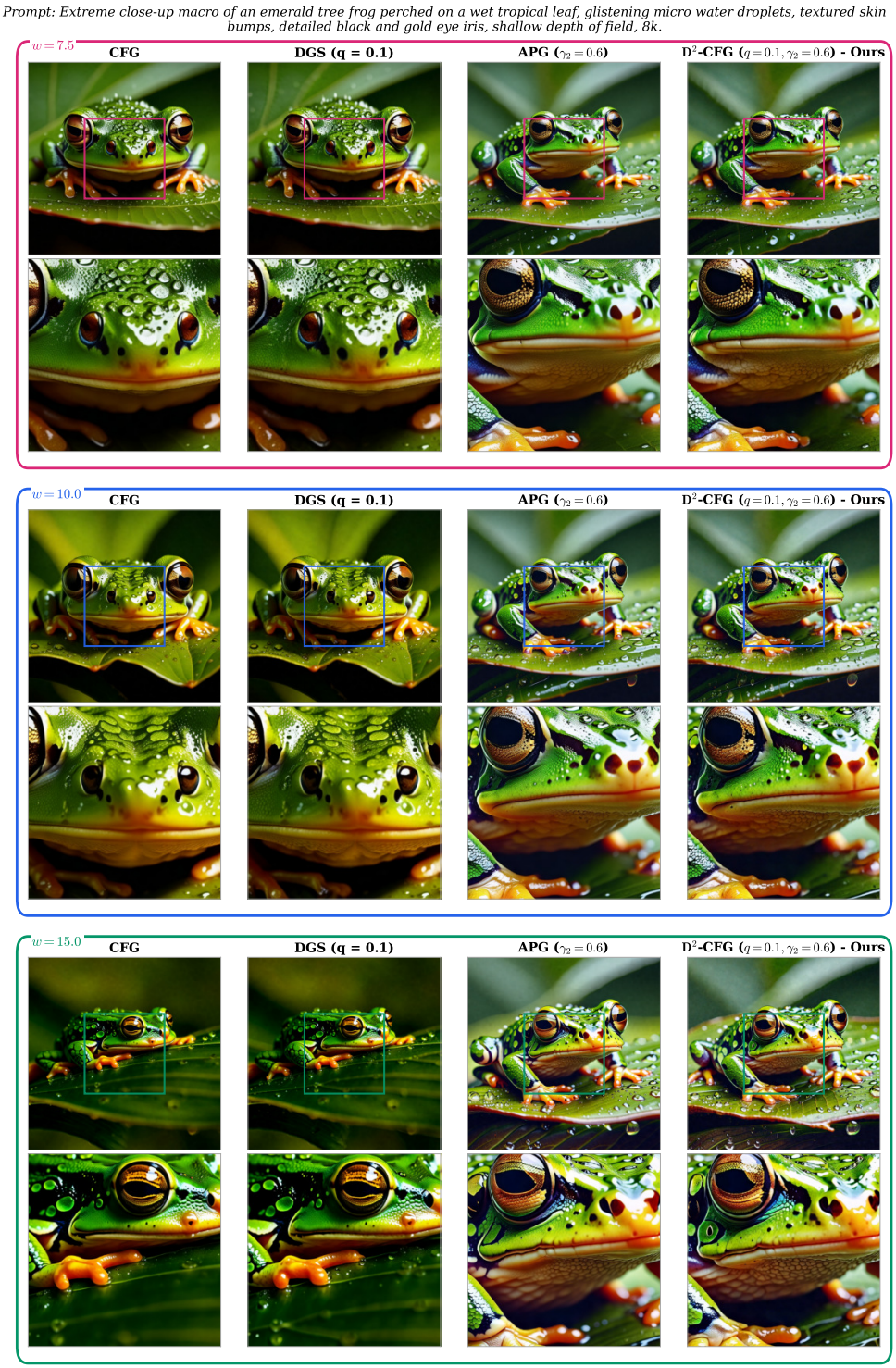


Figure 6: **Qualitative comparison on SD3.5** ($w \in \{7.5, 10, 15\}$**; DGS and D^2-CFG at** $q = 0.1$**).** Standard CFG and DGS suffer from severe oversaturation and prompt misalignment as w grows. APG and D^2-CFG both prevent burnout; D^2-CFG additionally exposes the density dial, which at $q = 0.1$ keeps fine skin texture at every guidance scale.

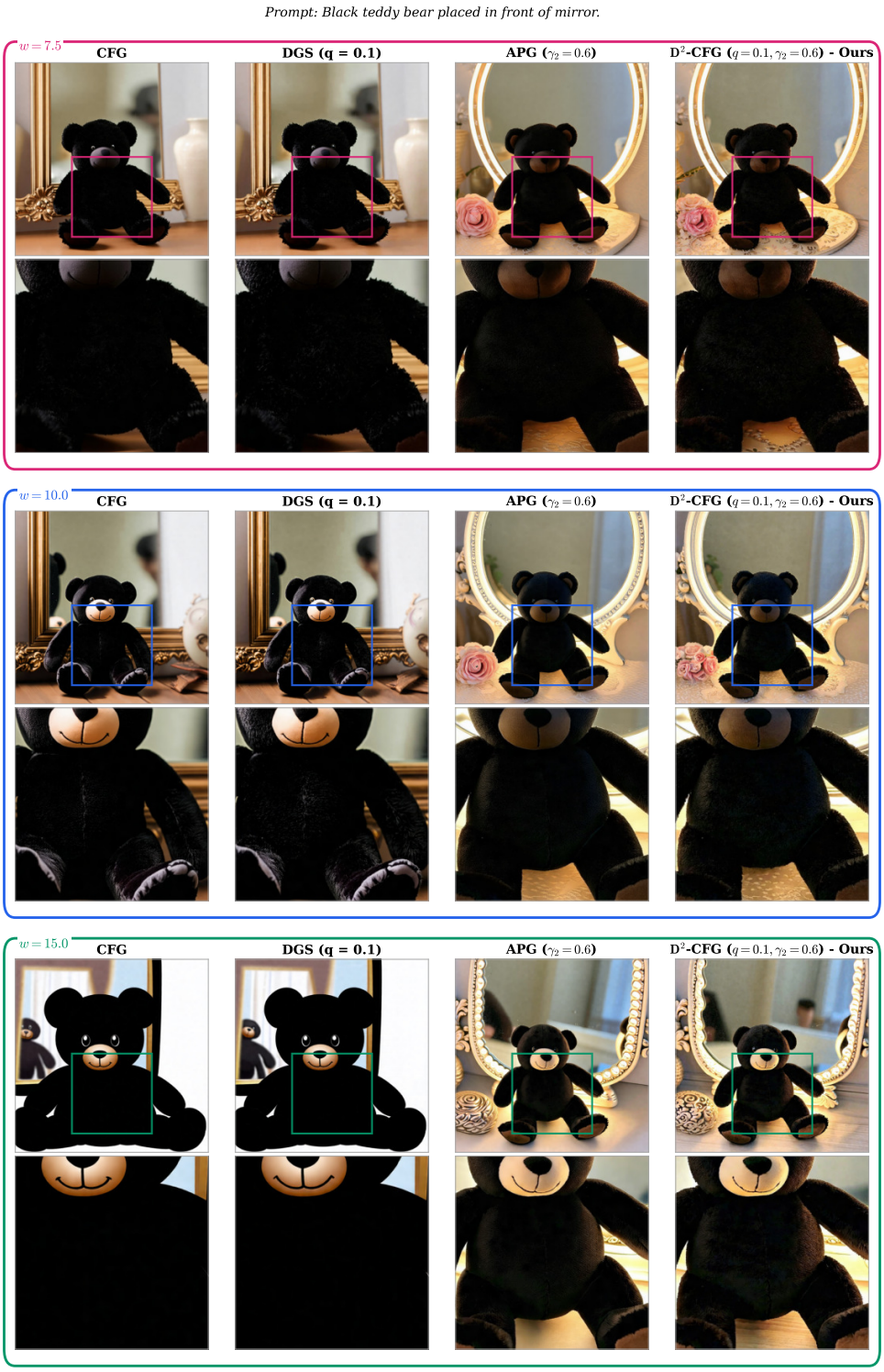


Figure 7: **Additional qualitative comparison on SD3.5** ($w \in \{7.5, 10, 15\}$**, Prompt 2; DGS and D^2-CFG at** $q = 0.1$**).** Standard CFG and DGS suffer from oversaturation and structural distortion as w grows, whereas D^2-CFG removes chromatic burnout at every guidance scale while keeping fine texture.

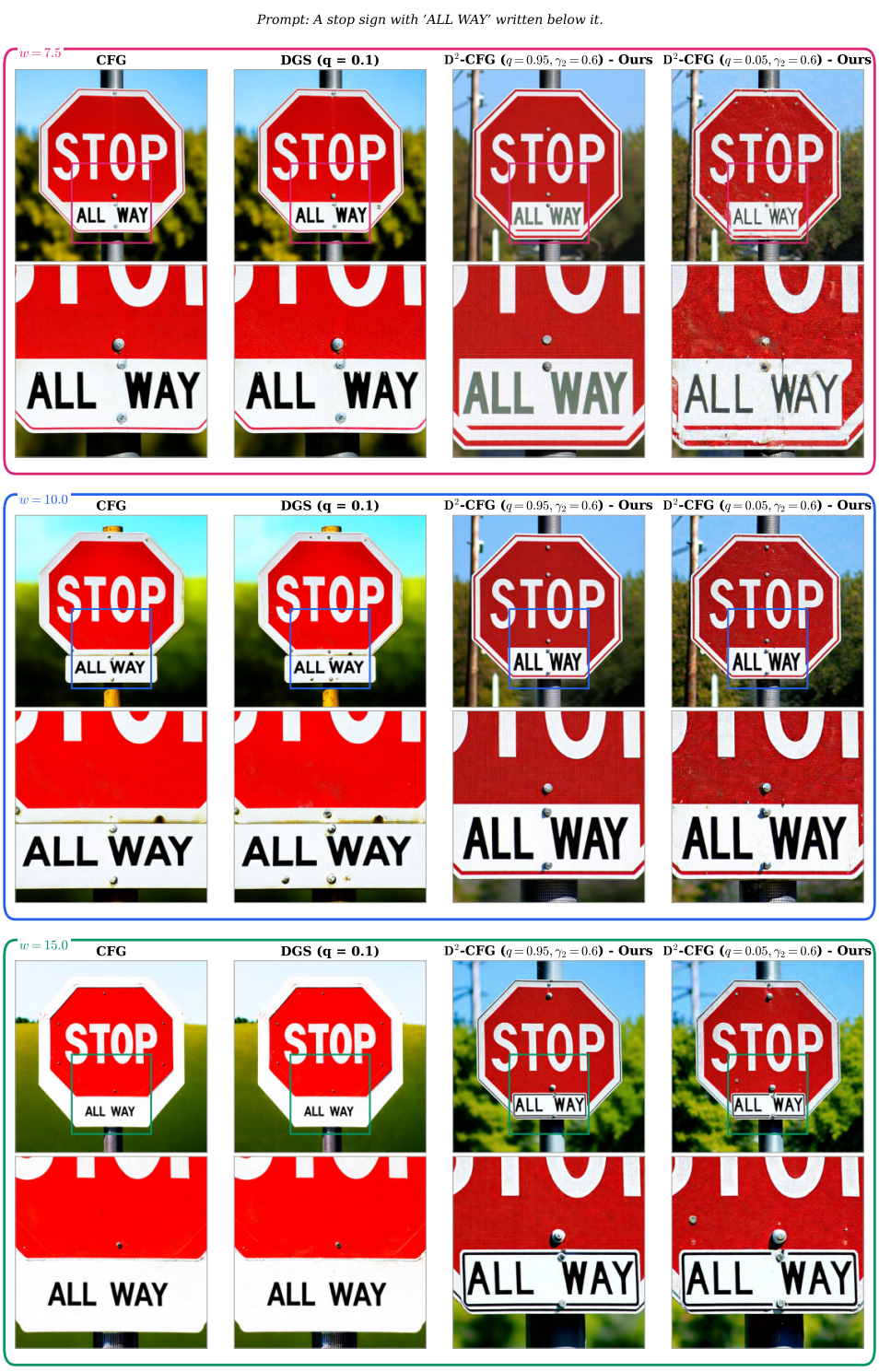


Figure 8: **Additional qualitative comparison on SD3.5** ($w \in \{7.5, 10, 15\}$**, Prompt 3; DGS at** $q = 0.1$**,** **D^2-CFG at** $q \in \{0.95, 0.05\}$ **with** $\gamma_2 = 0.6$**).** D^2-CFG preserves natural scene illumination without contrast clipping at every guidance scale; lowering q from 0.95 to 0.05 adds fine texture.

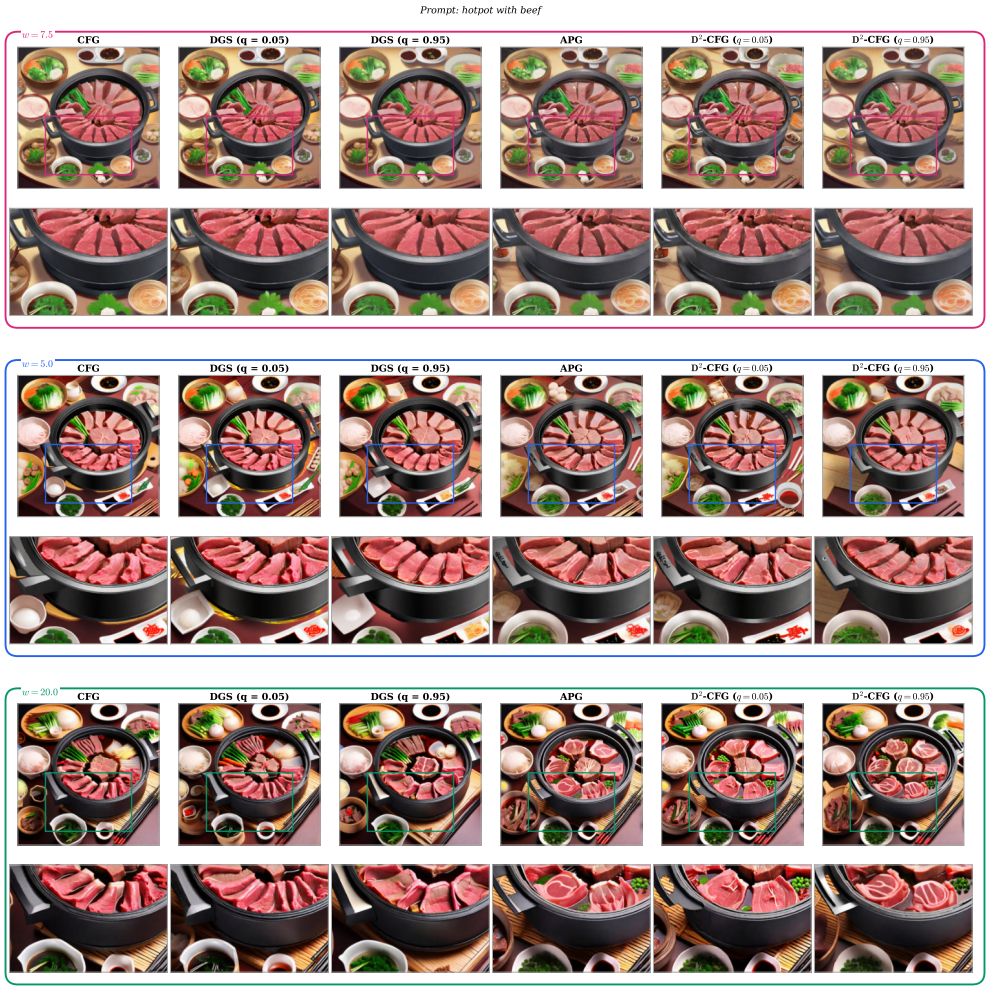


Figure 9: **Qualitative comparison on SDXL with Pick-a-Pic prompts (Kirstain et al., 2023).** For $w \geq 15$, DGS entangles prompt alignment with the density dial, whereas D^2-CFG cleanly modulates structural detail across q without contrast clipping.

C Extended Quantitative Evaluations and 2D Benchmarks

In this section, we expand on the quantitative text-to-image benchmarks across guidance scales and provide full details on the exact-likelihood 2D synthetic evaluations.

C.1 SDXL Quantitative Evaluation

We expand on additional guidance weights in Table 2 (MS-COCO) and Table 3 (Pick-a-Pic), and visualize the corresponding Pareto frontiers in Fig. 10. Across all tables, DGSAPG denotes the default setting ($\gamma_2 = 1.0$) and TSR uses $k = 0.93, \sigma = 3.0$ from the original papers, while we additionally include the relaxed variant DGSAPG ($\gamma_2 = 0.4$) for a matched comparison. TSR reduces to standard CFG for $k = \sigma = 1$; at its default setting the rescaling is mild, which is why its rows closely track standard CFG.

Table 2: **Quantitative evaluation on MS-COCO 5k with fixed $q = 0.30$ (SDXL).** We compare Standard CFG against DGSCFG, TSR, DGSAPG, DGSAPG ($\gamma_2 = 0.4$), and D^2-CFG ($\gamma_2 = 1.0, 0.4$). The **best** and second-best results per guidance scale are bolded and underlined, respectively.

Scale	Method	FID ↓	CLIP ↑	Prec ↑	Rec ↑	Sat. ↓	Cont.	IQA ↑	ImageReward ↑
$w = 7.5$	Standard CFG	**23.23**	27.00	**0.735**	0.615	22.49%	0.2002	91.97	0.7981
	DGSCFG	23.30	26.98	**0.735**	0.616	23.04%	0.2026	**92.09**	**0.8120**
	TSR	23.28	27.00	0.731	0.616	22.50%	0.2002	91.98	0.7988
	DGSAPG	23.79	**27.06**	0.723	0.610	**18.14%**	0.1611	89.90	0.7539
	DGSAPG ($\gamma_2 = 0.4$)	23.34	27.01	**0.735**	0.616	21.48%	0.1903	91.65	0.8045
	D^2-CFG ($\gamma_2 = 1.0$)	23.78	27.04	0.720	0.611	18.62%	0.1628	89.75	0.7672
	D^2-CFG ($\gamma_2 = 0.4$)	23.35	27.00	0.731	**0.628**	21.87%	0.1918	91.57	0.8110
$w = 10.0$	Standard CFG	23.92	27.13	0.719	0.611	24.65%	0.2146	92.83	0.8516
	DGSCFG	**23.69**	27.10	0.722	0.612	25.15%	0.2167	**92.84**	0.8591
	DGSAPG	24.07	27.20	0.721	0.602	**19.56%**	0.1760	91.14	0.8420
	DGSAPG ($\gamma_2 = 0.4$)	23.85	27.15	0.724	**0.618**	23.31%	0.2048	92.51	0.8569
	D^2-CFG ($\gamma_2 = 1.0$)	23.97	**27.22**	**0.732**	0.606	19.99%	0.1777	90.92	0.8411
	D^2-CFG ($\gamma_2 = 0.4$)	**23.69**	27.16	0.726	0.616	23.67%	0.2060	92.33	**0.8703**
$w = 15.0$	Standard CFG	24.73	27.19	0.705	0.601	28.55%	0.2382	**93.83**	0.8832
	DGSCFG	24.48	27.21	0.718	0.604	29.04%	0.2399	93.80	0.8845
	TSR	24.81	27.20	0.706	0.605	28.56%	0.2382	93.81	0.8828
	DGSAPG	24.65	27.25	0.710	0.599	**23.24%**	0.2081	92.91	0.8794
	DGSAPG ($\gamma_2 = 0.4$)	24.64	27.21	0.709	0.595	27.07%	0.2302	93.60	0.8794
	D^2-CFG ($\gamma_2 = 1.0$)	**24.34**	**27.27**	0.719	**0.607**	23.60%	0.2094	92.65	**0.8927**
	D^2-CFG ($\gamma_2 = 0.4$)	24.39	27.23	**0.723**	0.604	27.37%	0.2310	93.36	0.8880
$w = 20.0$	Standard CFG	25.29	27.29	0.683	0.596	32.44%	0.2583	**94.38**	0.8838
	DGSCFG	25.08	27.31	0.692	0.593	32.90%	0.2596	94.30	0.8847
	TSR	25.42	27.29	0.693	0.593	32.42%	0.2583	94.36	0.8826
	DGSAPG	25.23	27.27	0.689	0.599	**28.00%**	0.2384	93.84	0.8814
	DGSAPG ($\gamma_2 = 0.4$)	25.32	27.31	0.678	0.594	31.12%	0.2525	94.27	0.8850
	D^2-CFG ($\gamma_2 = 1.0$)	24.81	27.32	0.696	**0.605**	28.34%	0.2393	93.68	0.8934
	D^2-CFG ($\gamma_2 = 0.4$)	**24.79**	**27.34**	**0.706**	0.591	31.40%	0.2531	94.00	**0.8994**
$w = 25.0$	Standard CFG	26.57	27.32	0.659	0.585	36.14%	0.2751	**94.50**	0.8494
	DGSCFG	26.51	27.35	0.660	0.580	36.47%	0.2762	94.46	0.8505
	TSR	26.54	27.33	0.662	0.579	36.15%	0.2751	94.43	0.8492
	DGSAPG	26.79	27.30	0.654	0.588	**33.11%**	0.2630	94.11	0.8355
	DGSAPG ($\gamma_2 = 0.4$)	26.38	27.32	0.666	0.582	35.36%	0.2719	94.47	0.8566
	D^2-CFG ($\gamma_2 = 1.0$)	25.99	27.33	0.666	**0.600**	33.42%	0.2636	93.78	0.8420
	D^2-CFG ($\gamma_2 = 0.4$)	**25.81**	**27.40**	**0.671**	0.593	35.57%	0.2720	94.20	**0.8647**

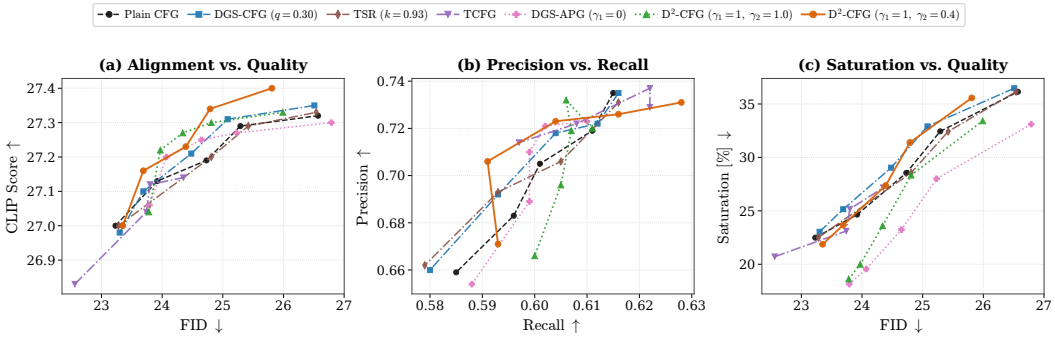


Figure 10: **Pareto frontiers across guidance scales $w \in [7.5, 25]$ on MS-COCO 5k (SDXL, $q = 0.30$).** Each marker is one w, increasing from left to right in (a) and (c). TCFG (Kwon et al., 2025) lowers the effective guidance scale, so we compare it along the frontier rather than at matched w: at matched CLIP its FID is on par with D^2-CFG, and it does not reach the high-alignment regime (CLIP > 27.15). TCFG is shown for $w \geq 10$.

Table 3: **Quantitative evaluation on Pick-a-Pic 5k with fixed $q = 0.30$ (SDXL).** We compare Standard CFG against TSR, DGSCFG, DGSAPG, DGSAPG ($\gamma_2 = 0.4$), and D^2-CFG ($\gamma_2 = 1.0, 0.6, 0.4$). PickScore and HPSv2 are reported on the $[0, 1]$ cosine scale. The **best** and second-best results per guidance scale are bolded and underlined, respectively.

Scale	Method	CLIP ↑	PickScore ↑	HPSv2 ↑	ImageReward ↑	Sat. ↓	Cont.
$w = 7.5$	Standard CFG	29.80	0.2210	0.2804	+0.8503	25.0%	0.1986
	TSR	29.80	0.2210	0.2803	+0.8499	25.0%	0.1986
	DGSCFG	29.80	**0.2212**	**0.2806**	+0.8576	25.4%	0.2010
	DGSAPG	29.75	0.2198	0.2766	+0.7906	**19.2%**	0.1564
	DGSAPG ($\gamma_2 = 0.4$)	29.80	0.2210	0.2800	+0.8500	23.6%	0.1883
	D^2-CFG ($\gamma_2 = 1.0$)	29.77	0.2200	0.2768	+0.7972	19.8%	0.1585
	D^2-CFG ($\gamma_2 = 0.6$)	**29.83**	0.2210	0.2793	+0.8552	22.9%	0.1822
	D^2-CFG ($\gamma_2 = 0.4$)	29.83	0.2212	0.2799	**+0.8633**	24.0%	0.1901
$w = 10.0$	Standard CFG	30.01	0.2215	0.2824	+0.9157	27.6%	0.2122
	TSR	30.00	0.2214	0.2824	+0.9159	27.6%	0.2122
	DGSCFG	30.02	0.2215	**0.2825**	+0.9217	28.0%	0.2142
	DGSAPG	30.02	0.2207	0.2794	+0.8782	**20.8%**	0.1707
	DGSAPG ($\gamma_2 = 0.4$)	30.02	0.2215	0.2820	+0.9172	25.8%	0.2020
	D^2-CFG ($\gamma_2 = 1.0$)	**30.06**	0.2208	0.2795	+0.8895	21.3%	0.1728
	D^2-CFG ($\gamma_2 = 0.6$)	30.06	0.2215	0.2815	+0.9250	25.0%	0.1961
	D^2-CFG ($\gamma_2 = 0.4$)	30.03	**0.2216**	0.2820	**+0.9294**	26.3%	0.2037
$w = 15.0$	Standard CFG	30.13	0.2212	0.2838	+0.9576	32.2%	0.2339
	TSR	30.13	0.2212	0.2838	+0.9559	32.2%	0.2339
	DGSCFG	30.15	0.2213	**0.2839**	+0.9635	32.5%	0.2354
	DGSAPG	30.21	0.2210	0.2824	+0.9571	**24.7%**	0.2001
	DGSAPG ($\gamma_2 = 0.4$)	30.18	0.2214	0.2837	+0.9673	30.1%	0.2254
	D^2-CFG ($\gamma_2 = 1.0$)	**30.27**	0.2212	0.2824	+0.9699	25.1%	0.2021
	D^2-CFG ($\gamma_2 = 0.6$)	30.21	0.2215	0.2835	**+0.9890**	29.1%	0.2205
	D^2-CFG ($\gamma_2 = 0.4$)	30.23	**0.2215**	0.2836	+0.9814	30.5%	0.2266
$w = 20.0$	Standard CFG	30.14	0.2204	0.2838	+0.9645	36.3%	0.2512
	TSR	30.13	0.2204	0.2838	+0.9635	36.3%	0.2512
	DGSCFG	30.14	0.2204	0.2838	+0.9681	36.6%	0.2525
	DGSAPG	30.20	0.2207	0.2835	+0.9755	**29.5%**	0.2276
	DGSAPG ($\gamma_2 = 0.4$)	30.16	0.2206	**0.2839**	+0.9740	34.2%	0.2447
	D^2-CFG ($\gamma_2 = 1.0$)	30.24	0.2207	0.2832	+0.9845	29.9%	0.2291
	D^2-CFG ($\gamma_2 = 0.6$)	**30.26**	**0.2207**	0.2837	**+0.9910**	33.3%	0.2415
	D^2-CFG ($\gamma_2 = 0.4$)	30.25	0.2207	0.2838	+0.9848	34.6%	0.2458
$w = 25.0$	Standard CFG	30.11	0.2194	**0.2831**	+0.9474	39.7%	0.2656
	TSR	30.11	0.2194	**0.2831**	+0.9480	39.7%	0.2656
	DGSCFG	30.11	0.2194	0.2831	+0.9514	40.0%	0.2666
	DGSAPG	30.04	0.2193	0.2826	+0.9382	**34.8%**	0.2499
	DGSAPG ($\gamma_2 = 0.4$)	30.11	0.2195	0.2831	+0.9555	38.2%	0.2612
	D^2-CFG ($\gamma_2 = 1.0$)	30.07	0.2193	0.2824	+0.9422	35.2%	0.2512
	D^2-CFG ($\gamma_2 = 0.6$)	30.17	0.2196	0.2831	**+0.9658**	37.5%	0.2591
	D^2-CFG ($\gamma_2 = 0.4$)	**30.19**	**0.2196**	0.2830	+0.9642	38.5%	0.2620

Comparison between Single (DGSAPG) and Double Projection (D^2-CFG). To assess whether simultaneous 2-plane projection provides a concrete advantage over sequentially stacking prior formulations, we compare D^2-CFG directly against DGSAPG, which projects out only the clean-space direction $\hat{\boldsymbol{x}}_0^c$ and rescales the residual by $\eta_t(q)$, leaving the score direction $\boldsymbol{\epsilon}_c$ unconstrained. As shown in Table 4, both methods achieve virtually identical saturation suppression across all guidance scales (ΔSat ≤ 0.5 pp), alongside closely matched text alignment (within 0.08 CLIP points). This empirically confirms that projection along $\hat{\boldsymbol{x}}_0^c$ uniquely isolates and eliminates dynamic-range burnout. However, D^2-CFG outperforms DGSAPG in distribution fidelity (FID) at every $w \geq 10$, and in mode coverage (Recall) at every w for $\gamma_2 = 1.0$, with the FID advantage widening monotonically from parity at $w = 7.5$ (ΔFID ≤ 0.01) to a margin of 0.57–0.80 at $w = 25.0$ (Table 4). This divergence is algebraically driven by the non-orthogonality of the two guidance axes ($\langle \hat{\boldsymbol{x}}_0^c, \boldsymbol{\epsilon}_c \rangle \neq 0$, with $|\cos| \approx 0.20$–0.28 on SDXL, Appendix H.1). Leaving $\boldsymbol{\epsilon}_c$ unprojected lets the APG-projected update $\boldsymbol{g}_{\text{APG}} := \Delta\boldsymbol{\epsilon} - \langle \Delta\boldsymbol{\epsilon}, \mathbf{e}_{\text{clean}} \rangle \mathbf{e}_{\text{clean}}$ keep a component along the score:

$$\langle \boldsymbol{g}_{\text{APG}}, \boldsymbol{\epsilon}_c \rangle \;=\; \langle \Delta\boldsymbol{\epsilon}, \boldsymbol{\epsilon}_c \rangle - \langle \Delta\boldsymbol{\epsilon}, \mathbf{e}_{\text{clean}} \rangle \langle \mathbf{e}_{\text{clean}}, \boldsymbol{\epsilon}_c \rangle \;\neq\; 0, \tag{15}$$

where $\mathbf{e}_{\text{clean}} = \hat{\boldsymbol{x}}_0^c / \|\hat{\boldsymbol{x}}_0^c\|$. By (8), this is an uncalibrated density forcing that scales linearly with w. Because $\langle \Delta\boldsymbol{\epsilon}, \boldsymbol{\epsilon}_c \rangle$ depends on the prompt, the resulting drift varies across prompts and cannot be removed by a constant recalibration of q.

Table 4: **D^2-CFG relative to DGSAPG at matched** γ_2 on MS-COCO 5k ($q = 0.30$, SDXL Base 1.0). Negative ΔFID and positive ΔRecall favor D^2-CFG.

		Guidance Scale w				
Quantity	γ_2	7.5	10.0	15.0	20.0	25.0
ΔFID ↓	1.0	-0.01	-0.10	-0.31	-0.42	$\mathbf{-0.80}$
	0.4	$+0.01$	-0.16	-0.25	-0.53	$\mathbf{-0.57}$
ΔRecall ↑	1.0	$+0.001$	$+0.004$	$+0.008$	$+0.006$	$\mathbf{+0.012}$
	0.4	$+0.012$	-0.002	$+0.009$	-0.003	$\mathbf{+0.011}$
ΔCLIP ↑	1.0	-0.02	$+0.02$	$+0.02$	$\mathbf{+0.05}$	$+0.03$
	0.4	-0.01	$+0.01$	$+0.02$	$+0.03$	$\mathbf{+0.08}$
ΔSat. ↓ (pp)	1.0	$+0.5$	$+0.4$	$+0.4$	$+0.3$	$+0.3$
	0.4	$+0.4$	$+0.4$	$+0.3$	$+0.3$	$+0.2$

Seed variation. Table 5 repeats the $w = 20$ comparison of Table 1 over three sampling seeds (all other tables report seed 0). At matched γ_2, D^2-CFG improves FID over DGSAPG by 0.40 and 0.53, more than twice the standard deviation of the difference; CLIP, Precision, Recall, and saturation differ by less than their seed spread.

Table 5: **Seed variation at** $w = 20$ **on MS-COCO 5k (SDXL,** $q = 0.30$**,** $\gamma_1 = 1$**).** Mean ± standard deviation over three sampling seeds. **Best** and second-best per column.

Method	**FID ↓**	**CLIP ↑**	**Prec ↑**	**Rec ↑**	**Sat. ↓**
Standard CFG	25.52 ± 0.28	27.30 ± 0.00	0.693 ± 0.010	0.583 ± 0.011	$32.46 \pm 0.11\%$
DGSAPG ($\gamma_2 = 1.0$)	25.10 ± 0.13	27.29 ± 0.03	0.695 ± 0.007	0.595 ± 0.006	$\mathbf{27.83 \pm 0.20}\%$
DGSAPG ($\gamma_2 = 0.4$)	25.36 ± 0.18	27.31 ± 0.01	0.691 ± 0.011	0.591 ± 0.004	$31.08 \pm 0.09\%$
D^2-CFG ($\gamma_2 = 1.0$, Ours)	$\mathbf{24.70 \pm 0.10}$	27.33 ± 0.04	0.701 ± 0.008	$\mathbf{0.604 \pm 0.006}$	$28.17 \pm 0.20\%$
D^2-CFG ($\gamma_2 = 0.4$, Ours)	24.83 ± 0.15	$\mathbf{27.34 \pm 0.02}$	$\mathbf{0.704 \pm 0.005}$	0.594 ± 0.005	$31.35 \pm 0.08\%$

C.2 SD3.5 Quantitative Evaluation

The results for MS-COCO are presented in Table 6, and for Pick-a-Pic in Table 8.

MS-COCO analysis. As shown in Table 6, at moderate guidance ($w \in [4.5, 10.0]$), baselines experience an initial rise in FID at $w = 7.5$ due to contrast and saturation accumulation, followed by a transient recovery at $w = 10.0$ driven by peak semantic recall (Recall rises to 0.561 for Standard CFG and 0.568 for DGSCFG). In this intermediate regime, D^2-CFG slightly lags DGSCFG in FID (29.60–30.12 vs. 28.42) due to evaluating under a fixed, stress-tested density budget ($q = 0.30$): by enforcing $\eta < 1.0$ and projecting out clean-space amplitude gain ($\gamma_2 > 0$), the model suppresses dynamic range (RMS contrast of 0.21–0.27 vs. 0.28 in baselines at $w = 10$), which Inception-v3 penalizes relative to natural COCO statistics. Conversely, at extreme guidance ($w = 15.0$), linear extrapolation causes catastrophic failure across all 1D baselines: Standard CFG, DGSCFG, and DGSAPG suffer severe distribution collapse (FID ≥ 35.75, saturation $> 56\%$, and precision < 0.59). In contrast, the spherical retraction of D^2-CFG bounds the norm growth of (14) and the projection removes the in-plane components, maintaining high precision (0.751–0.783 for $\gamma_2 \in \{0.4, 0.6\}$) and achieving the best overall distribution fidelity (FID $= 27.51$ for $\gamma_2 = 0.4$, and peak perceptual quality IQA $= 93.02$ for $\gamma_2 = 0.6$) while scaling CLIP score to 28.02.

Table 6: **Quantitative evaluation on MS-COCO 5k with fixed $q = 0.30$ on SD3.5 Medium.** We compare Standard CFG against DGSCFG, DGSAPG, DGSAPG ($\gamma_2 = 0.4$), and D^2-CFG ($\gamma_2 = 1.0, 0.6, 0.4$). The **best** and second-best results per guidance scale are bolded and underlined, respectively.

Scale	Method	FID ↓	CLIP ↑	Prec ↑	Rec ↑	Sat. ↓	Cont.	IQA ↑
$w = 4.5$	Standard CFG	29.62	**26.75**	0.788	0.545	39.46%	0.2244	90.05
	DGSCFG	28.69	26.59	0.800	0.565	39.60%	0.2256	89.89
	DGSAPG	29.07	26.59	0.803	0.566	39.48%	0.2250	89.99
	DGSAPG ($\gamma_2 = 0.4$)	28.85	26.59	0.803	0.565	39.70%	0.2253	89.91
	D^2-CFG ($\gamma_2 = 1.0$)	**28.37**	26.65	0.808	**0.574**	**34.39%**	0.2114	**90.25**
	D^2-CFG ($\gamma_2 = 0.6$)	28.58	26.59	**0.809**	0.571	36.92%	0.2177	90.17
	D^2-CFG ($\gamma_2 = 0.4$)	28.49	26.58	0.804	0.571	38.04%	0.2206	89.97
$w = 7.5$	Standard CFG	30.13	26.82	0.779	0.549	43.76%	0.2620	91.53
	DGSCFG	**29.48**	26.75	0.787	0.555	43.89%	0.2628	91.33
	DGSAPG	30.18	26.78	0.782	0.559	45.61%	0.2638	91.40
	DGSAPG ($\gamma_2 = 0.4$)	29.79	26.79	0.790	0.553	44.64%	0.2632	91.44
	D^2-CFG ($\gamma_2 = 1.0$)	30.06	**27.03**	**0.804**	0.551	**31.50%**	0.2195	92.00
	D^2-CFG ($\gamma_2 = 0.6$)	30.16	26.83	0.802	**0.561**	37.35%	0.2424	**92.06**
	D^2-CFG ($\gamma_2 = 0.4$)	30.24	26.77	**0.804**	0.552	39.60%	0.2513	91.84
$w = 10.0$	Standard CFG	29.01	26.93	0.731	0.561	49.02%	0.2767	90.55
	DGSCFG	**28.42**	26.89	0.738	**0.568**	49.03%	0.2773	90.38
	DGSAPG	28.86	26.89	0.744	0.559	52.57%	0.2793	90.35
	DGSAPG ($\gamma_2 = 0.4$)	28.58	26.91	0.744	0.555	50.36%	0.2789	90.51
	D^2-CFG ($\gamma_2 = 1.0$)	29.95	**27.48**	0.798	0.535	**28.64%**	0.2070	92.14
	D^2-CFG ($\gamma_2 = 0.6$)	30.12	26.91	**0.804**	0.538	37.63%	0.2540	**92.66**
	D^2-CFG ($\gamma_2 = 0.4$)	29.60	26.88	0.802	0.550	42.00%	0.2675	92.24
$w = 15.0$	Standard CFG	37.51	26.69	0.571	0.574	56.44%	0.2819	83.53
	DGSCFG	35.75	26.76	0.587	**0.585**	56.59%	0.2830	82.95
	DGSAPG	37.73	26.72	0.578	0.553	60.21%	0.2813	83.62
	DGSAPG ($\gamma_2 = 0.4$)	36.46	26.80	0.585	0.554	58.09%	0.2838	84.04
	D^2-CFG ($\gamma_2 = 1.0$)	33.32	**28.02**	0.717	0.503	**24.78%**	0.1556	81.59
	D^2-CFG ($\gamma_2 = 0.6$)	28.26	27.11	**0.783**	0.549	40.78%	0.2694	**93.02**
	D^2-CFG ($\gamma_2 = 0.4$)	**27.51**	27.05	0.751	0.556	50.96%	0.2849	90.89

Seed variation. Table 7 repeats the $w = 15$ comparison of Table 6 over three sampling seeds. At matched $\gamma_2 = 1.0$, D^2-CFG improves FID over DGSAPG by 4.0 points and reduces saturation from 60% to 25%; at $\gamma_2 = 0.6$ it attains the best FID and Precision in the table. The seed spread is below 0.4 FID for every method, an order of magnitude smaller than these gaps.

Table 7: **Seed variation at $w = 15$ on MS-COCO 5k (SD3.5 Medium, $q = 0.30$, $\gamma_1 = 1$).** Mean ± standard deviation over three sampling seeds. **Best** and second-best per column.

Method	FID ↓	CLIP ↑	Prec ↑	Rec ↑	Sat. ↓
Standard CFG	37.28 ± 0.27	26.73 ± 0.03	0.573 ± 0.010	$\mathbf{0.567 \pm 0.007}$	$56.26 \pm 0.17\%$
DGSAPG ($\gamma_2 = 1.0$)	37.38 ± 0.35	26.73 ± 0.01	0.575 ± 0.004	0.552 ± 0.007	$59.98 \pm 0.20\%$
DGSAPG ($\gamma_2 = 0.4$)	36.02 ± 0.39	26.82 ± 0.02	0.584 ± 0.011	0.556 ± 0.009	$57.87 \pm 0.21\%$
D^2-CFG ($\gamma_2 = 1.0$, Ours)	33.35 ± 0.09	$\mathbf{28.03 \pm 0.04}$	0.718 ± 0.011	0.492 ± 0.014	$\mathbf{24.76 \pm 0.04}\%$
D^2-CFG ($\gamma_2 = 0.6$, Ours)	$\mathbf{28.40 \pm 0.12}$	27.11 ± 0.02	$\mathbf{0.785 \pm 0.005}$	0.544 ± 0.008	$40.78 \pm 0.09\%$

Pick-a-Pic analysis. As w increases on SD 3.5 Medium, Standard CFG saturates and clips pixels, and the preference scores drop. DGSCFG alone is insufficient because it leaves the intra-step directional composition of the velocity update unchanged. With spherical retraction and the clean-space projection, D^2-CFG sustains strong text adherence at high guidance without sacrificing preference scores: while $\gamma_2 = 1.0$ eliminates burnout entirely, the balanced regime ($\gamma_2 = 0.6$) establishes a stable saturation plateau and preserves natural contrast, yielding the global Pareto frontier across PickScore, HPSv2, and ImageReward.

Table 8: **Quantitative evaluation on Pick-a-Pic 5k with fixed $q = 0.30$ on SD3.5 Medium.** We compare Standard CFG against DGSCFG, TSR, CFG-Zero*, DGSAPG ($\gamma_2 = 1.0, 0.6, 0.4$), and D^2-CFG ($\gamma_2 = 1.0, 0.6, 0.4$, with $\gamma_1 = 1$ fixed). The **best** and second-best results per guidance scale are bolded and underlined, respectively.

Scale	Method	CLIP ↑	PickScore ↑	HPSv2 ↑	ImageReward ↑	Sat. ↓	Cont.
$w = 4.5$	Standard CFG	28.44	0.2217	0.2825	+0.9986	45.0%	0.2170
	DGSCFG	28.37	0.2220	0.2823	+1.0140	45.3%	0.2183
	TSR	28.44	0.2217	0.2825	+0.9985	45.0%	0.2170
	CFG-Zero*	**28.45**	0.2221	0.2827	+1.0152	40.0%	0.2135
	DGSAPG	28.39	**0.2226**	0.2829	+1.0413	39.3%	0.2110
	DGSAPG ($\gamma_2 = 0.6$)	28.39	0.2222	0.2826	+1.0264	44.5%	0.2178
	DGSAPG ($\gamma_2 = 0.4$)	28.39	0.2222	0.2825	+1.0236	45.1%	0.2182
	D^2-CFG ($\gamma_2 = 1.0$)	28.38	**0.2226**	0.2828	**+1.0414**	**39.0%**	0.2108
	D^2-CFG ($\gamma_2 = 0.6$)	28.40	**0.2226**	**0.2830**	+1.0397	41.7%	0.2144
	D^2-CFG ($\gamma_2 = 0.4$)	28.38	0.2225	0.2828	+1.0350	43.0%	0.2159
$w = 7.5$	Standard CFG	28.46	0.2205	0.2831	+1.0493	49.8%	0.2515
	DGSCFG	28.43	0.2208	0.2829	+1.0581	50.0%	0.2523
	TSR	28.47	0.2205	0.2831	+1.0499	49.7%	0.2515
	CFG-Zero*	28.56	0.2220	0.2853	+1.1093	47.7%	0.2450
	DGSAPG	28.47	0.2213	0.2838	+1.0878	50.8%	0.2529
	DGSAPG ($\gamma_2 = 0.6$)	28.52	0.2212	0.2836	+1.0790	48.1%	0.2484
	DGSAPG ($\gamma_2 = 0.4$)	28.45	0.2211	0.2835	+1.0766	50.4%	0.2524
	D^2-CFG ($\gamma_2 = 1.0$)	28.59	0.2222	0.2847	+1.1333	**36.4%**	0.2226
	D^2-CFG ($\gamma_2 = 0.6$)	**28.64**	**0.2229**	**0.2855**	**+1.1508**	42.9%	0.2403
	D^2-CFG ($\gamma_2 = 0.4$)	28.62	0.2226	0.2851	+1.1341	45.8%	0.2459
$w = 10.0$	Standard CFG	28.12	0.2174	0.2798	+0.9556	53.4%	0.2680
	DGSCFG	28.14	0.2177	0.2798	+0.9643	53.5%	0.2687
	TSR	28.20	0.2179	0.2805	+0.9809	52.8%	0.2683
	CFG-Zero*	28.53	0.2205	0.2843	+1.1009	54.0%	0.2620
	DGSAPG	28.30	0.2184	0.2809	+1.0263	51.2%	0.2584
	DGSAPG ($\gamma_2 = 0.6$)	28.38	0.2183	0.2808	+1.0175	51.8%	0.2653
	DGSAPG ($\gamma_2 = 0.4$)	28.39	0.2194	0.2821	+1.0561	51.3%	0.2665
	D^2-CFG ($\gamma_2 = 1.0$)	28.57	0.2202	0.2830	+1.1311	**33.5%**	0.2145
	D^2-CFG ($\gamma_2 = 0.6$)	**28.67**	**0.2223**	**0.2856**	**+1.1775**	43.5%	0.2535
	D^2-CFG ($\gamma_2 = 0.4$)	28.61	0.2215	0.2847	+1.1460	47.9%	0.2629
$w = 15.0$	Standard CFG	26.97	0.2094	0.2698	+0.5067	56.4%	0.2793
	DGSCFG	26.97	0.2095	0.2697	+0.5094	56.5%	0.2796
	TSR	27.85	0.2141	0.2762	+0.8228	54.6%	0.2859
	CFG-Zero*	28.25	0.2145	0.2760	+0.9610	55.8%	0.2750
	DGSAPG	27.55	0.2109	0.2725	+0.7038	49.9%	0.2480
	DGSAPG ($\gamma_2 = 0.6$)	27.28	0.2107	0.2721	+0.6583	59.7%	0.2857
	DGSAPG ($\gamma_2 = 0.4$)	27.26	0.2105	0.2716	+0.6323	58.6%	0.2843
	D^2-CFG ($\gamma_2 = 1.0$)	28.04	0.2139	0.2752	+0.9306	**28.0%**	0.1722
	D^2-CFG ($\gamma_2 = 0.6$)	**28.68**	**0.2203**	**0.2840**	**+1.1717**	44.4%	0.2704
	D^2-CFG ($\gamma_2 = 0.4$)	28.36	0.2181	0.2814	+1.0762	53.0%	0.2845

C.3 Fractal 2D Experiment

We construct a 2D synthetic dataset using a multi-modal *fractal Gaussian mixture* ($K = 1016$ components) organized along a shared central trunk that bifurcates into distinct class-conditioned branches (A and B). The score network is parameterized as a 4-layer MLP (256 hidden units, SiLU activations, sinusoidal time embeddings) trained for $50{,}000$ steps with Adam (lr $= 10^{-3}$, batch size 512) under standard VP-SDE diffusion. Samples are generated by integrating the probability-flow ODE over $N = 600$ steps, evaluating exact ground-truth likelihood analytically via the mixture PDF. Autoguidance uses the weak model and setup of Karras et al. (2024) at $w = 4$. Because $\mathcal{V}_2 = \mathrm{span}\{\boldsymbol{\epsilon}_c, \hat{\boldsymbol{x}}_0^c\} = \mathbb{R}^2$ in two dimensions, the joint projection is degenerate and both 2D benchmarks use the score-axis projection alone ($\gamma_1 = 1$, $\gamma_2 = 0$), which removes the density forcing R while keeping the component of $\Delta\boldsymbol{\epsilon}$ orthogonal to $\boldsymbol{\epsilon}_c$.

Results. Figure 11 directly validates the interference relation in (9), where we sweep $q \in \{0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.8\}$ across guidance scales $w \in [2, 6]$. The gap between DGSCFG and D^2-CFG widens as w increases, precisely as predicted by our analysis: larger guidance scales amplify the interference term $R(\boldsymbol{x}_t, t; w)$. Furthermore, when $q = 0.5$ ($\eta_t = 1$), DGSCFG reduces to standard CFG, demonstrating that it directly inherits CFG's off-manifold degradation. In contrast, D^2-CFG retains its orthogonal projection ($\Delta\boldsymbol{\epsilon}_\perp$), maintaining high likelihood and manifold fidelity even in the unboosted regime.

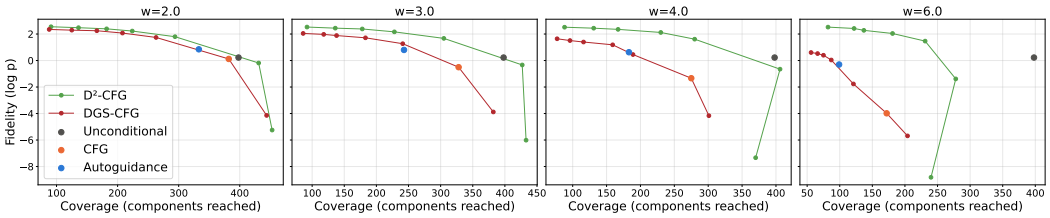


Figure 11: **Fidelity ($\mathbb{E}[\log p_{\text{gt}}]$) vs. Mode Coverage across guidance scales $w \in \{2, 3, 4, 6\}$.** At every guidance scale, the fidelity–coverage frontier of D^2-CFG lies above that of DGSCFG (red curve), standard CFG (orange dot), and autoguidance (blue dot). At $w = 6$, D^2-CFG retains $\log p \approx 1.5$ at about 230 components, whereas DGSCFG falls below 0 beyond about 100.

C.4 Toy Example: 6-Mode Gaussian Mixture

We also validate level-set tracking on a 2D Gaussian Mixture Model (GMM) with $K = 6$ modes arranged on a circle of radius 4 with unequal weights and covariances, where the conditioning targets the top mode (green boundary circle in Fig. 12). Trajectories are integrated via the probability-flow ODE from $\sigma_{\max} = 15.0$ to $\sigma_{\min} = 0.01$ using the exact analytical score and Laplacian across guidance scales $w \in \{1, 4\}$ and quantiles $q \in \{0.1, 0.5, 0.9\}$.

- **Moderate guidance ($w = 1$, Fig. 12a):** At moderate guidance, both DGSCFG and D^2-CFG smoothly contract samples into concentric iso-density rings from the low-density tail ($q = 0.1$) to the high-density mode core ($q = 0.9$).
- **High-guidance regime ($w = 4$, Fig. 12b):** Standard CFG pushes the particle cloud upward off the target mode into the low-density void, and DGSCFG inherits this uncalibrated push, forming an asymmetric smeared bulge. In contrast, D^2-CFG eliminates parallel score forcing ($R \equiv 0$), concentrating samples on the target level set.



(a) Moderate guidance ($w = 1$) (b) High-guidance regime ($w = 4$)

Figure 12: **Level-set validation on the 6-mode GMM. (a)** At $w = 1$, sweeping $q \in \{0.1, 0.5, 0.9\}$ smoothly modulates trajectories across concentric iso-density rings. **(b)** At $w = 4$, vanilla CFG displaces samples off the target mode and DGSCFG smears across density, whereas D^2-CFG eliminates parallel forcing ($R \equiv 0$) and concentrates samples on the requested level set for each q. In each grid, the bottom-right panel is D^2-CFG at $q = 0.9$.

D Continuous Likelihood Tracking and the Density Dial

Given the probability-flow ODE (4), the instantaneous change of variables theorem (Chen et al., 2018) describes the time evolution of the sample log-density along the trajectory under the marginal p_t:

$$\frac{d}{dt} \log p_t(\boldsymbol{x}_t) \;=\; -\operatorname{div} \boldsymbol{u}_t(\boldsymbol{x}_t). \tag{16}$$

When the trajectory is steered by a perturbed velocity field $d\boldsymbol{x}_t = \tilde{\boldsymbol{u}}_t(\boldsymbol{x}_t)\, dt$, the evolution of the sample log-density under the conditional marginal $p_t(\cdot \mid c)$ generalizes to (Karczewski et al., 2025a; Skreta et al., 2025):

$$\frac{d}{dt} \log p_t(\boldsymbol{x}_t \mid c) \;=\; \underbrace{-\operatorname{div} \boldsymbol{u}_t(\boldsymbol{x}_t)}_{\text{unperturbed rate}} + \underbrace{\big\langle \nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t \mid c),\; \tilde{\boldsymbol{u}}_t(\boldsymbol{x}_t) - \boldsymbol{u}_t(\boldsymbol{x}_t) \big\rangle}_{\text{guidance-induced density forcing } R}. \tag{17}$$

Under the probability-flow ODE drift $\boldsymbol{u}_t(\boldsymbol{x}_t) = -\frac{1}{2}\beta(t)\big(\boldsymbol{x}_t - \frac{1}{\sigma_t}\boldsymbol{\epsilon}_c\big)$ in (4), replacing $\boldsymbol{\epsilon}_c$ with the CFG noise prediction $\boldsymbol{\epsilon}_{\text{cfg}} = \boldsymbol{\epsilon}_c + w\Delta\boldsymbol{\epsilon}$ directly yields the velocity perturbation:

$$\tilde{\boldsymbol{u}}_t(\boldsymbol{x}_t) - \boldsymbol{u}_t(\boldsymbol{x}_t) \;=\; -\frac{1}{2}\beta(t)\left(\boldsymbol{x}_t - \frac{\boldsymbol{\epsilon}_c + w\Delta\boldsymbol{\epsilon}}{\sigma_t}\right) + \frac{1}{2}\beta(t)\left(\boldsymbol{x}_t - \frac{\boldsymbol{\epsilon}_c}{\sigma_t}\right) \;=\; \frac{\beta(t)}{2\sigma_t} w\, \Delta\boldsymbol{\epsilon}. \tag{18}$$

Substituting this perturbation alongside $\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t \mid c) \approx -\boldsymbol{\epsilon}_c/\sigma_t$ into (17) recovers Eq. (8) in the main text, confirming that any guidance update orthogonal to $\boldsymbol{\epsilon}_c$ exerts zero direct forcing on sample likelihood.

D.1 Asymptotic Normality of the Score Curvature

In Karczewski et al. (2025a), Eq. (17) was leveraged constructively to build a minimum-norm velocity perturbation $\min_{\tilde{\boldsymbol{u}}} \frac{1}{2}\|\tilde{\boldsymbol{u}} - \boldsymbol{u}_t\|^2$ that enforces a prescribed log-density rate. The optimal perturbation is strictly parallel to the score, governed by the Laplacian and score norm identity:

$$\Delta \log p_t(\boldsymbol{x}) + \|\nabla \log p_t(\boldsymbol{x})\|^2 \;=\; \frac{\Delta p_t(\boldsymbol{x})}{p_t(\boldsymbol{x})}. \tag{19}$$

To characterize the typical values of this quantity in high dimensions, define the normalized score curvature statistic $h(\boldsymbol{x})$:

$$h(\boldsymbol{x}) \;:=\; \frac{\sigma_t^2\big(\Delta \log p_t(\boldsymbol{x}) + \|\nabla \log p_t(\boldsymbol{x})\|^2\big)}{\sqrt{2D}}, \qquad D \coloneqq \dim(\boldsymbol{x}). \tag{20}$$

Karczewski et al. (2025a) prove that $h(\boldsymbol{x}) \xrightarrow{d} \mathcal{N}(0,1)$ as $D \to \infty$ for isotropic Gaussians, for non-isotropic Gaussians under a non-dominance condition on the precision spectrum (De Jong, 1987), and for empirical Gaussian mixtures; we refer to their paper for the three derivations. Mapping a target density quantile $q \in (0,1)$ through the inverse CDF $\Phi^{-1}(q)$ of $\mathcal{N}(0,1)$, i.e., selecting the q-quantile of $h(\boldsymbol{x})$, yields the DGS score-rescaling factor:

$$\eta_t(\boldsymbol{x}_t) \;=\; 1 + \frac{\sqrt{2D}\,\Phi^{-1}(q)}{\|\boldsymbol{\epsilon}_c\|^2} \;\approx\; 1 + \sqrt{\frac{2}{D}}\,\Phi^{-1}(q), \tag{21}$$

where the right-hand approximation reflects the high-dimensional concentration $\|\boldsymbol{\epsilon}_c\|^2 \approx D$.

D.2 The Effective Dimension $\tilde{D}$ as a Bandwidth Knob

In SDXL ($D = 4 \times 128 \times 128 = 65{,}536$), the nominal bandwidth $\sqrt{2/D} \approx 0.55\%$ provides sufficient dynamic range for $q \in (0,1)$ to regulate typicality across structural sampling steps ($\mathrm{logSNR}(t) \leq 1.0$). However, in higher-dimensional latent spaces such as SD3.5 ($D = 16 \times 128 \times 128 = 262{,}144$), the $\mathcal{O}(D^{-1/2})$ concentration of measure contracts the modulation factor to $\sqrt{2/D} \approx 0.28\%$, leaving q almost no effect over 28 sampling steps; we therefore treat $\tilde{D}$ as an explicit bandwidth hyperparameter.

To preserve an active density dial as D grows, we introduce an effective dimension $\tilde{D} \leq D$ into (21):

$$\eta_t(q) \;=\; 1 + \frac{D\sqrt{2/\tilde{D}}\,\Phi^{-1}(q)}{\|\boldsymbol{\epsilon}_c\|^2} \;\approx\; 1 + \sqrt{\frac{2}{\tilde{D}}}\,\Phi^{-1}(q). \tag{22}$$

When $\tilde{D} = D$ (used for SDXL), Eq. (22) reduces identically to (21). For SD3.5, setting $\tilde{D} = 1{,}024$ provides a well-conditioned modulation bandwidth ($\sqrt{2/\tilde{D}} \approx 4.4\%$), ensuring that q actively regulates trajectory density. $\tilde{D}$ thus compensates for the growing dimensionality of latent diffusion models: it holds the per-step rescaling at a numerically effective level (about 7% at the extreme quantiles on SD3.5) instead of letting it vanish as D grows. With $\tilde{D} < D$, q no longer equals the quantile of $h(\boldsymbol{x})$ in (20); it indexes the same one-parameter family of rescalings with a larger gain, so $\tilde{D}$ is a bandwidth knob rather than a calibration.

E Denoiser Perspective of D^2-CFG

Applying Tweedie's identity to (13) reveals the complementary clean-space decomposition directly in terms of the denoiser predictions $\hat{\boldsymbol{x}}_0^c$ and $\hat{\boldsymbol{x}}_0^{\varnothing}$.

Defining the clean denoising displacement $\boldsymbol{d}_t \coloneqq \hat{\boldsymbol{x}}_0^c - \frac{\boldsymbol{x}_t}{\alpha_t}$ and the orthogonal semantic update $\Delta_\perp \coloneqq \Pi_{\mathcal{V}_2^\perp}\big(\hat{\boldsymbol{x}}_0^c - \hat{\boldsymbol{x}}_0^{\varnothing}\big)$ with $\mathcal{V}_2 = \mathrm{span}\{\boldsymbol{x}_t, \hat{\boldsymbol{x}}_0^c\} = \mathrm{span}\{\boldsymbol{\epsilon}_c, \hat{\boldsymbol{x}}_0^c\}$, grouping terms yields:

$$\hat{\boldsymbol{x}}_0^{\mathrm{D}^2} \;=\; \underbrace{\frac{\boldsymbol{x}_t}{\alpha_t}}_{\text{current state}} \;+\; \underbrace{\eta_t(q)\,\boldsymbol{d}_t}_{\text{controlled denoising step}} \;+\; \underbrace{w\,\Delta_\perp}_{\text{orthogonal semantic steer}}. \tag{23}$$

Eq. (23) makes the disentanglement explicit: (i) the noisy reference $\frac{\boldsymbol{x}_t}{\alpha_t}$ is stepped along the denoising vector $\boldsymbol{d}_t$ scaled purely by the typicality factor $\eta_t(q)$, isolating trajectory likelihood from semantic guidance; (ii) when $\eta_t(q) = 1$, the trajectory exactly tracks the unperturbed conditional mean without score-parallel drift; (iii) the guidance scale w acts exclusively through the orthogonal vector $\Delta_\perp \in \mathcal{V}_2^\perp$, steering semantic structure with zero projection onto either the current state or the clean conditional manifold.

E.1 Linear Diffusion Model

To understand the mechanism in an analytically solvable setting, we examine D^2-CFG through the lens of linear Gaussian diffusion models (Li et al., 2026).

In particular, we consider a class-conditional setting where we model $p(\boldsymbol{x}_0 \mid c) = \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$ as the class-conditional Gaussian and $p(\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u)$ as a Gaussian approximation of the unconditional mixture. By the Law of Total Variance, the unconditional covariance satisfies $\boldsymbol{\Sigma}_u = \mathbb{E}_c[\boldsymbol{\Sigma}_c] + \mathrm{Cov}_c(\boldsymbol{\mu}_c) \succeq \mathbb{E}_c[\boldsymbol{\Sigma}_c]$. Under the forward process $\boldsymbol{x}_t = \alpha_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}$, the noisy state conditioned on class c is $\boldsymbol{x}_t \mid c \sim \mathcal{N}(\alpha_t \boldsymbol{\mu}_c, \boldsymbol{\Omega}_t^c)$ with $\boldsymbol{\Omega}_t^c \coloneqq \alpha_t^2 \boldsymbol{\Sigma}_c + \sigma_t^2 \mathbf{I}$.

The main-text claim is the following; the rest of this section derives it.

Proposition 1 (Linear Diffusion Model)**.** *Let $p(\boldsymbol{x}_0 \mid c) = \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$ and $p(\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u)$ (a single-Gaussian approximation of the unconditional mixture, with $\boldsymbol{\Sigma}_u \succ \boldsymbol{\Sigma}_c$), with $\boldsymbol{x}_t \mid c \sim \mathcal{N}(\alpha_t \boldsymbol{\mu}_c, \boldsymbol{\Omega}_t^c)$, $\boldsymbol{\Omega}_t^c \coloneqq \alpha_t^2 \boldsymbol{\Sigma}_c + \sigma_t^2 \mathbf{I}$, and $\mathbf{W}_t^c \coloneqq \alpha_t \boldsymbol{\Sigma}_c (\boldsymbol{\Omega}_t^c)^{-1}$. Under the Bayes-optimal affine denoisers, standard CFG satisfies:*

$$\begin{aligned} \mathbb{E}_{\boldsymbol{x}_t|c}\big[\langle \Delta\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t),\, \hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t)\rangle\big] &= \mathrm{Tr}\big(\Delta\mathbf{W}_t \boldsymbol{\Omega}_t^c (\mathbf{W}_t^c)^\top\big) + \sigma_t^2 \boldsymbol{\mu}_c^\top (\boldsymbol{\Omega}_t^u)^{-1} (\boldsymbol{\mu}_c - \boldsymbol{\mu}_u), \\ \mathbb{E}_{\boldsymbol{x}_t|c}\big[\langle \Delta\boldsymbol{\epsilon}(\boldsymbol{x}_t),\, \boldsymbol{\epsilon}_c(\boldsymbol{x}_t)\rangle\big] &= \sigma_t^2 \mathrm{Tr}\big((\boldsymbol{\Omega}_t^c)^{-1} - (\boldsymbol{\Omega}_t^u)^{-1}\big) > 0. \end{aligned} \tag{24}$$

The Bayes-optimal denoisers and score predictions are affine mappings:

$$\hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t) = \mathbf{W}_t^c \boldsymbol{x}_t + \boldsymbol{b}_t^c, \qquad \hat{\boldsymbol{x}}_0^u(\boldsymbol{x}_t) = \mathbf{W}_t^u \boldsymbol{x}_t + \boldsymbol{b}_t^u, \tag{25}$$

$$\boldsymbol{\epsilon}_c(\boldsymbol{x}_t) = \mathbf{K}_t^c \boldsymbol{x}_t - \tilde{\boldsymbol{b}}_t^c, \qquad \boldsymbol{\epsilon}_u(\boldsymbol{x}_t) = \mathbf{K}_t^u \boldsymbol{x}_t - \tilde{\boldsymbol{b}}_t^u, \tag{26}$$

where the linear operators and affine bias vectors are defined in closed form as:

$$\begin{aligned} \mathbf{W}_t^c &= \alpha_t \boldsymbol{\Sigma}_c \big(\boldsymbol{\Omega}_t^c\big)^{-1}, & \boldsymbol{b}_t^c &= \big(\mathbf{I} - \alpha_t \mathbf{W}_t^c\big)\boldsymbol{\mu}_c = \sigma_t^2 \big(\boldsymbol{\Omega}_t^c\big)^{-1} \boldsymbol{\mu}_c, \\ \mathbf{K}_t^c &= \sigma_t \big(\boldsymbol{\Omega}_t^c\big)^{-1}, & \tilde{\boldsymbol{b}}_t^c &= \frac{\alpha_t}{\sigma_t} \boldsymbol{b}_t^c = \alpha_t \sigma_t \big(\boldsymbol{\Omega}_t^c\big)^{-1} \boldsymbol{\mu}_c, \end{aligned} \tag{27}$$

with identical expressions holding for $\mathbf{W}_t^u, \boldsymbol{b}_t^u, \mathbf{K}_t^u, \tilde{\boldsymbol{b}}_t^u$ upon replacing $(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c, \boldsymbol{\Omega}_t^c)$ with $(\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u, \boldsymbol{\Omega}_t^u)$.

Why Standard CFG Conflates Saturation and Density. Standard CFG applies the raw guidance perturbation:

$$\Delta\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) = \Delta\mathbf{W}_t \boldsymbol{x}_t + \Delta\boldsymbol{b}_t, \qquad \Delta\boldsymbol{\epsilon}(\boldsymbol{x}_t) = \Delta\mathbf{K}_t \boldsymbol{x}_t - \Delta\tilde{\boldsymbol{b}}_t, \tag{28}$$

where $\Delta\mathbf{W}_t \coloneqq \mathbf{W}_t^c - \mathbf{W}_t^u$, $\Delta\boldsymbol{b}_t \coloneqq \boldsymbol{b}_t^c - \boldsymbol{b}_t^u$, $\Delta\mathbf{K}_t \coloneqq \mathbf{K}_t^c - \mathbf{K}_t^u$, and $\Delta\tilde{\boldsymbol{b}}_t \coloneqq \tilde{\boldsymbol{b}}_t^c - \tilde{\boldsymbol{b}}_t^u$. This guidance simultaneously corrupts two distinct geometric axes:

1. **Saturation Axis (Mean-Shift Content Amplification):** The instantaneous content gain is measured by the parallel projection onto the clean state:

$$\rho_t(\boldsymbol{x}_t) \coloneqq \frac{\langle \Delta\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t),\, \hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t)\rangle}{\|\hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t)\|^2} = \frac{\langle \Delta\mathbf{W}_t \boldsymbol{x}_t + \Delta\boldsymbol{b}_t,\, \mathbf{W}_t^c \boldsymbol{x}_t + \boldsymbol{b}_t^c\rangle}{\|\mathbf{W}_t^c \boldsymbol{x}_t + \boldsymbol{b}_t^c\|^2}. \tag{29}$$

   Taking the expectation over $\boldsymbol{x}_t \mid c \sim \mathcal{N}(\alpha_t \boldsymbol{\mu}_c, \boldsymbol{\Omega}_t^c)$, the numerator decomposes into:

$$\mathbb{E}_{\boldsymbol{x}_t|c}\big[\langle \Delta\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t),\, \hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t)\rangle\big] = \mathrm{Tr}\big(\Delta\mathbf{W}_t \boldsymbol{\Omega}_t^c (\mathbf{W}_t^c)^\top\big) + \sigma_t^2 \boldsymbol{\mu}_c^\top (\boldsymbol{\Omega}_t^u)^{-1} (\boldsymbol{\mu}_c - \boldsymbol{\mu}_u). \tag{30}$$

   Although $\boldsymbol{\Sigma}_u \succeq \boldsymbol{\Sigma}_c$ yields $\Delta\mathbf{W}_t \preceq 0$ along shared covariance modes, the class-mean term is positive (centering the data at $\boldsymbol{\mu}_u = \mathbf{0}$ gives $\sigma_t^2 \boldsymbol{\mu}_c^\top (\boldsymbol{\Omega}_t^u)^{-1} \boldsymbol{\mu}_c > 0$) and dominates in high dimensions where classes are well separated (Li et al., 2026). This forces a positive expected parallel component, i.e., a positive content gain $\rho_t > 0$ for typical samples. Scaling $w > 0$ directly inflates the clean signal amplitude $(1 + w\rho_t)\hat{\boldsymbol{x}}_0^c$ (Section 3.1), causing dynamic range burnout and oversaturation.

2. **Density Axis (Second-Order Precision/Covariance Contraction):** In noise/score space, writing $\boldsymbol{z} \coloneqq \boldsymbol{x}_t - \alpha_t \boldsymbol{\mu}_c \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_t^c)$ gives $\boldsymbol{\epsilon}_c(\boldsymbol{x}_t) = \sigma_t (\boldsymbol{\Omega}_t^c)^{-1} \boldsymbol{z}$ and decomposes the guidance vector into a zero-mean precision-gap term and a deterministic mean-shift term:

$$\Delta\boldsymbol{\epsilon}(\boldsymbol{x}_t) = \sigma_t \big((\boldsymbol{\Omega}_t^c)^{-1} - (\boldsymbol{\Omega}_t^u)^{-1}\big)\boldsymbol{z} - \alpha_t \sigma_t (\boldsymbol{\Omega}_t^u)^{-1} (\boldsymbol{\mu}_c - \boldsymbol{\mu}_u). \tag{31}$$

Because the conditional score predictor is zero-mean ($\mathbb{E}_{\boldsymbol{x}_t|c}[\boldsymbol{\epsilon}_c(\boldsymbol{x}_t)] = \mathbf{0}$), the deterministic class-mean shift ($\boldsymbol{\mu}_c - \boldsymbol{\mu}_u$) vanishes in expectation against $\boldsymbol{\epsilon}_c(\boldsymbol{x}_t)$, and the expected inner product is governed purely by the second-order precision gap:

$$\mathbb{E}_{\boldsymbol{x}_t|c}\big[\langle \Delta\boldsymbol{\epsilon}(\boldsymbol{x}_t),\, \boldsymbol{\epsilon}_c(\boldsymbol{x}_t)\rangle\big] \;=\; \mathrm{Tr}\big(\Delta\mathbf{K}_t\boldsymbol{\Omega}_t^c(\mathbf{K}_t^c)^\top\big) \;=\; \sigma_t^2\mathrm{Tr}\big((\boldsymbol{\Omega}_t^c)^{-1} - (\boldsymbol{\Omega}_t^u)^{-1}\big) \;>\; 0. \tag{32}$$

Because a single class has smaller variance than the unconditional mixture ($\boldsymbol{\Sigma}_c \prec \boldsymbol{\Sigma}_u \implies (\boldsymbol{\Omega}_t^c)^{-1} \succ (\boldsymbol{\Omega}_t^u)^{-1}$), scaling $\Delta\boldsymbol{\epsilon}$ by w artificially contracts intra-class variance toward $\alpha_t\boldsymbol{\mu}_c$, injecting the uncalibrated density forcing $R(\boldsymbol{x}_t, t; w) = -\frac{\beta(t)w}{2\sigma_t^2}\langle\Delta\boldsymbol{\epsilon}, \boldsymbol{\epsilon}_c\rangle \neq 0$ (Section 3.2).

How D^2-CFG Resolves Both Pathologies. Because $\boldsymbol{x}_t = \alpha_t\hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t) + \sigma_t\boldsymbol{\epsilon}_c(\boldsymbol{x}_t)$, the 2-plane subspace is defined directly from the complete affine estimators: $\mathcal{V}_2(\boldsymbol{x}_t) = \mathrm{span}\{\hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t), \boldsymbol{\epsilon}_c(\boldsymbol{x}_t)\}$. By projecting the contrastive guidance update onto the orthogonal complement:

$$\Delta_\perp(\boldsymbol{x}_t) \;=\; \mathbf{P}_{\mathcal{V}_2^\perp}(\boldsymbol{x}_t)\,\Delta\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) \;=\; \mathbf{P}_{\mathcal{V}_2^\perp}(\boldsymbol{x}_t)\big(\Delta\mathbf{W}_t\boldsymbol{x}_t + \Delta\boldsymbol{b}_t\big), \tag{33}$$

D^2-CFG eliminates both pathologies identically:

1. **Elimination of Saturation:** Because $\hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t) \in \mathcal{V}_2(\boldsymbol{x}_t)$, we have $\mathbf{P}_{\mathcal{V}_2^\perp}(\boldsymbol{x}_t)\hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t) \equiv \mathbf{0}$. The effective content gain vanishes for every sample:
$$\langle\Delta_\perp(\boldsymbol{x}_t),\, \hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t)\rangle \;=\; \langle\Delta\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t),\, \underbrace{\mathbf{P}_{\mathcal{V}_2^\perp}(\boldsymbol{x}_t)\hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t)}_{=\mathbf{0}}\rangle \;\equiv\; 0. \tag{34}$$
This enforces $\rho_t = 0$ identically, preventing w from inflating clean signal amplitude regardless of the sign of $(\boldsymbol{\Sigma}_c - \boldsymbol{\Sigma}_u)$ or the magnitude of $\Delta\boldsymbol{b}_t$.
2. **Orthogonal Density Regulation:** Similarly, because $\boldsymbol{\epsilon}_c(\boldsymbol{x}_t) \in \mathcal{V}_2(\boldsymbol{x}_t)$, parasitic score alignment is identically eliminated:
$$\langle\Delta\boldsymbol{\epsilon}_\perp(\boldsymbol{x}_t),\, \boldsymbol{\epsilon}_c(\boldsymbol{x}_t)\rangle \;=\; \langle\Delta\boldsymbol{\epsilon}(\boldsymbol{x}_t),\, \underbrace{\mathbf{P}_{\mathcal{V}_2^\perp}(\boldsymbol{x}_t)\boldsymbol{\epsilon}_c(\boldsymbol{x}_t)}_{=\mathbf{0}}\rangle \;\equiv\; 0. \tag{35}$$
This enforces $R(\boldsymbol{x}_t, t; w) \equiv 0$, eliminating unintended density drift.

Numerical result. Lastly, Fig. 13 shows samples from the linear diffusion model of this section fitted to ImageNet 64×64 (Russakovsky et al., 2015), comparing CFG, DGSCFG, and D^2-CFG across q. In a Gaussian model the log-density is a function of the Mahalanobis radius, so the density dial has an exact meaning: high q moves samples inward toward the class mean, which for a Gaussian is a blur, and low q pushes them onto an outer shell, which adds high-frequency texture. D^2-CFG traverses this range as q varies, whereas DGSCFG barely moves because the CFG interference term of (9) overrides the dial.

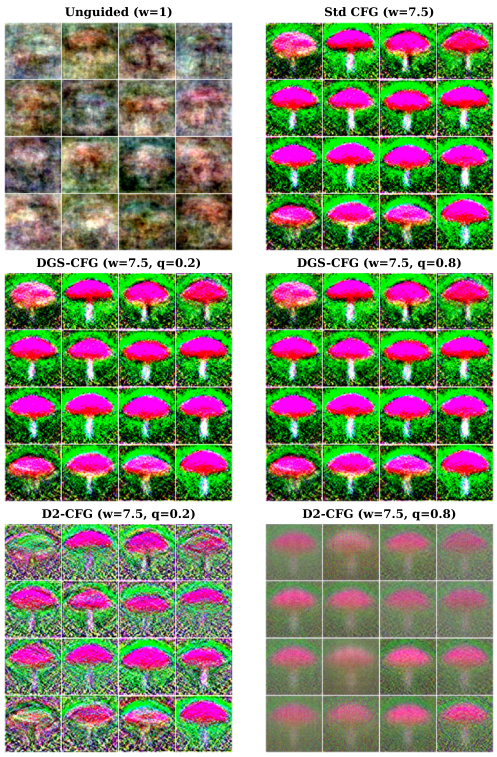


Figure 13: **Linear diffusion model fitted to ImageNet** 64×64 **(Appendix E.1).** In a Gaussian model, log-density is a function of the Mahalanobis radius, so q has an exact meaning: high q pulls samples toward the class mean (a blur), low q pushes them onto an outer shell (high-frequency texture). At high guidance, D^2-CFG traverses this range with q, whereas DGSCFG produces nearly identical samples across quantiles because the CFG interference term overrides the dial.

F Flow Matching Formulation of D^2-CFG

In this section, we provide the continuous-time formulation and algorithmic implementation of D^2-CFG for Flow Matching (Lipman et al., 2023; Albergo et al., 2025), as deployed in models such as Stable Diffusion 3.5. Following the flow-matching convention, $\boldsymbol{x}_0$ denotes the noise endpoint and $\boldsymbol{x}_1$ the data endpoint in this section, the reverse of the diffusion notation in the main text.

F.1 Background and Dual Endpoint Decomposition

Flow matching and stochastic interpolants define a family of continuous-time processes that interpolate between a source noise prior $p_0 = \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and a target data distribution $p_1 = p_{\text{data}}$:

$$\boldsymbol{x}_t \;=\; \alpha_t\, \boldsymbol{x}_0 \;+\; \beta_t\, \boldsymbol{x}_1, \qquad t \in [0,1], \tag{36}$$

with boundary conditions $\alpha_0 = 1, \alpha_1 = 0$ and $\beta_0 = 0, \beta_1 = 1$. Under the canonical linear interpolant $\alpha_t = 1 - t$ and $\beta_t = t$:

$$\boldsymbol{x}_t \;=\; (1-t)\, \boldsymbol{x}_0 \;+\; t\, \boldsymbol{x}_1, \tag{37}$$

a neural network $\boldsymbol{v}_\theta(\boldsymbol{x}_t, \boldsymbol{c}, t)$ is trained to predict the conditional velocity field $\boldsymbol{v}_c(\boldsymbol{x}_t, t) \approx \mathbb{E}[\boldsymbol{x}_1 - \boldsymbol{x}_0 \mid \boldsymbol{x}_t, \boldsymbol{c}]$, defining the deterministic probability-flow ODE $\dot{\boldsymbol{x}}_t = \boldsymbol{v}_c(\boldsymbol{x}_t, t)$. Inverting (37) yields exact instantaneous estimators for the clean data endpoint $\hat{\boldsymbol{x}}_1^c$ and source noise endpoint $\hat{\boldsymbol{x}}_0^c$:

$$\hat{\boldsymbol{x}}_1^c(\boldsymbol{x}_t, t) \;=\; \boldsymbol{x}_t \;+\; (1-t)\, \boldsymbol{v}_c(\boldsymbol{x}_t, t), \qquad \hat{\boldsymbol{x}}_0^c(\boldsymbol{x}_t, t) \;=\; \boldsymbol{x}_t \;-\; t\, \boldsymbol{v}_c(\boldsymbol{x}_t, t). \tag{38}$$

Given the unconditional velocity $\boldsymbol{v}_u(\boldsymbol{x}_t, t) = \boldsymbol{v}_\theta(\boldsymbol{x}_t, \emptyset, t)$, the guidance difference $\Delta\boldsymbol{v} = \boldsymbol{v}_c - \boldsymbol{v}_u$ linearly perturbs both endpoints via $\Delta\hat{\boldsymbol{x}}_1 = (1-t)\Delta\boldsymbol{v}$ and $\Delta\hat{\boldsymbol{x}}_0 = -t\Delta\boldsymbol{v}$.

F.2 JOINT 2-PLANE ORTHONORMAL PROJECTION

We construct an orthonormal basis $\{\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2\}$ for the joint 2-plane $\mathcal{V}_2 = \mathrm{span}\{\hat{\boldsymbol{x}}_0^c, \hat{\boldsymbol{x}}_1^c\}$ via Gram–Schmidt, in the same order as in Section 3.3: the noise endpoint $\hat{\boldsymbol{x}}_0^c$ (the analogue of $\boldsymbol{\epsilon}_c$) is orthogonalized first, then the data endpoint $\hat{\boldsymbol{x}}_1^c$ (the analogue of the clean estimate in the main text), with $\hat{\boldsymbol{e}}_{\text{data}} = \hat{\boldsymbol{x}}_1^c / \|\hat{\boldsymbol{x}}_1^c\|$:

$$\hat{\boldsymbol{e}}_1 \;=\; \frac{\hat{\boldsymbol{x}}_0^c}{\|\hat{\boldsymbol{x}}_0^c\|}, \qquad \tilde{\boldsymbol{e}}_2 \;=\; \hat{\boldsymbol{e}}_{\text{data}} \;-\; \langle \hat{\boldsymbol{e}}_{\text{data}},\, \hat{\boldsymbol{e}}_1 \rangle\, \hat{\boldsymbol{e}}_1, \qquad \hat{\boldsymbol{e}}_2 \;=\; \frac{\tilde{\boldsymbol{e}}_2}{\|\tilde{\boldsymbol{e}}_2\|}. \tag{39}$$

Projecting the raw velocity guidance difference $\Delta\boldsymbol{v}$ onto $\mathcal{V}_2^{\perp}$ (with $\gamma_1 = 1$ fixed) yields:

$$\Delta\boldsymbol{v}_{\perp} \;=\; \Delta\boldsymbol{v} \;-\; \langle \Delta\boldsymbol{v},\, \hat{\boldsymbol{e}}_1 \rangle\, \hat{\boldsymbol{e}}_1 \;-\; \gamma_2\, \langle \Delta\boldsymbol{v},\, \hat{\boldsymbol{e}}_2 \rangle\, \hat{\boldsymbol{e}}_2, \tag{40}$$

where eliminating the component along $\hat{\boldsymbol{e}}_1$ ($\gamma_1 = 1$) removes noise-space density interference, while $\gamma_2 \in [0, 1]$ governs clean-space amplitude suppression along $\hat{\boldsymbol{e}}_2$.

F.3 DUAL ENDPOINT GUIDANCE AND VELOCITY SYNTHESIS

Standard CFG and DGSCFG conflate $\hat{\boldsymbol{x}}_1$ and $\hat{\boldsymbol{x}}_0$ through the coupled velocity combination $\boldsymbol{v}_{\text{CFG}} = \boldsymbol{v}_c + w\Delta\boldsymbol{v}$, so any global scalar rescaling indiscriminately dampens the conditional field $\boldsymbol{v}_c$. Instead, perturbing the velocity orthogonally via $w\Delta\boldsymbol{v}_{\perp} \in \mathcal{V}_2^{\perp}$ shifts each endpoint independently: $\hat{\boldsymbol{x}}_1^{\text{raw}} = \hat{\boldsymbol{x}}_1^c + (1-t)w\Delta\boldsymbol{v}_{\perp}$ and $\hat{\boldsymbol{x}}_0^{\text{raw}} = \hat{\boldsymbol{x}}_0^c - tw\Delta\boldsymbol{v}_{\perp}$. Because $\Delta\boldsymbol{v}_{\perp}$ is orthogonal to $\hat{\boldsymbol{x}}_0^c$ by construction, and to $\hat{\boldsymbol{x}}_1^c$ when $\gamma_2 = 1$, each endpoint norm expands by Pythagoras, the flow-matching analogue of (14) with $(\sigma_t/\alpha_t)^2$ replaced by $(1-t)^2$ and t^2 (for $\gamma_2 < 1$ the first identity holds up to a residual cross term $2(1-t)w\langle \hat{\boldsymbol{x}}_1^c, \Delta\boldsymbol{v}_{\perp} \rangle$ that vanishes as $\gamma_2 \to 1$):

$$\|\hat{\boldsymbol{x}}_1^{\text{raw}}\|^2 \;=\; \|\hat{\boldsymbol{x}}_1^c\|^2 + (1-t)^2 w^2 \|\Delta\boldsymbol{v}_{\perp}\|^2, \qquad \|\hat{\boldsymbol{x}}_0^{\text{raw}}\|^2 \;=\; \|\hat{\boldsymbol{x}}_0^c\|^2 + t^2 w^2 \|\Delta\boldsymbol{v}_{\perp}\|^2. \tag{41}$$

D^2-CFG regulates both endpoints independently before synthesizing the guided velocity:

1. **Clean data energy normalization (Spherical Retraction):** Retracting $\hat{\boldsymbol{x}}_1^{\text{raw}}$ onto the base energy sphere $\|\hat{\boldsymbol{x}}_1^c\|$ eliminates the $(1-t)^2 w^2 \|\Delta\boldsymbol{v}_{\perp}\|^2$ norm inflation and prevents VAE color clipping:

$$\hat{\boldsymbol{x}}_1^{\text{guided}} \;=\; \hat{\boldsymbol{x}}_1^{\text{raw}} \cdot \frac{\|\hat{\boldsymbol{x}}_1^c\|}{\|\hat{\boldsymbol{x}}_1^{\text{raw}}\|}. \tag{42}$$

   This is the metric projection retraction onto the sphere of radius $\|\hat{\boldsymbol{x}}_1^c\|$ (Absil et al., 2008): for $\gamma_2 = 1$ the update is tangent to the sphere at $\hat{\boldsymbol{x}}_1^c$, and (48) is its second-order expansion.

2. **Source noise density regulation:** Scaling $\hat{\boldsymbol{x}}_0^{\text{raw}}$ to the target density radius $\eta_t(q)\|\hat{\boldsymbol{x}}_0^c\|$ (with $\eta_t(q)$ from (22) active during the structural phase $\mathrm{logSNR}(t) \le 1$ and $\eta_t(q) = 1$ otherwise, as in the diffusion case) regulates trajectory typicality to quantile $q \in (0, 1)$:

$$\hat{\boldsymbol{x}}_0^{\text{guided}} \;=\; \hat{\boldsymbol{x}}_0^{\text{raw}} \cdot \frac{\eta_t(q)\, \|\hat{\boldsymbol{x}}_0^c\|}{\|\hat{\boldsymbol{x}}_0^{\text{raw}}\|}. \tag{43}$$

The guided velocity integrated by the ODE solver is then synthesized directly from the regulated endpoints:

$$\boxed{\boldsymbol{v}_{\text{D}^2}(\boldsymbol{x}_t, t) \;=\; \hat{\boldsymbol{x}}_1^{\text{guided}}(\boldsymbol{x}_t, t) \;-\; \hat{\boldsymbol{x}}_0^{\text{guided}}(\boldsymbol{x}_t, t)} \tag{44}$$

Algorithm 2 summarizes the resulting sampler; it mirrors Algorithm 1 with the two endpoint retractions added. Unlike CFG-Rescale (Lin et al., 2024), which matches the standard deviation of the guided noise prediction to the conditional one and blends the result with the unrescaled prediction, the retraction resets the norms of both endpoints to targets that include $\eta_t(q)$, removes exactly the quadratic term in (41), and uses no blending.

Which ingredient does what on SD3.5. Table 9 separates the retraction from the projection on Pick-a-Pic ($q = 0.30$). Without retraction, every variant burns out at $w = 15$ regardless of γ_1 and γ_2, as predicted by the quadratic term in (14), which does not depend on the basis. The retraction alone lifts standard CFG to ImageReward $= 1.12$; adding the density factor, the clean-space projection ($\gamma_2 = 0.6$), and the score-space projection ($\gamma_1 = 1$) then improves ImageReward at every step, and the full method is best on all five metrics at both scales with baselines ($w = 10$ and 15).

Table 9: **Retraction and projection ablation on SD3.5 (Pick-a-Pic 5k, $q = 0.30$).** "Flat" denotes Euler steps without spherical retraction; baseline rows are reported at $w = 10$ and 15. Burnout is the fraction of clipped pixels. Without retraction, all variants saturate at high guidance; with it, each added ingredient improves ImageReward, and the full method is best on all five metrics. **Best** per guidance scale.

Scale	Variant	γ_2	γ_1	Retr.	ImageReward ↑	CLIP ↑	IQA ↑	Sat. % ↓	Burnout % ↓
$w = 4.5$	D^2-CFG (Flat)	0.6	1	–	1.01	28.39	89.01	44.9	0.37
	D^2-CFG (Ours)	0.6	1	✓	**1.04**	**28.40**	**89.07**	**41.7**	**0.17**
$w = 7.5$	D^2-CFG (Flat)	0.6	1	–	1.05	28.36	89.92	50.4	2.40
	D^2-CFG (Ours)	0.6	1	✓	**1.15**	**28.64**	**91.00**	**42.9**	**0.52**
$w = 10.0$	Standard CFG (Flat)	0	0	–	0.96	28.12	88.41	53.4	8.45
	DGSAPG (Flat)	0.6	0	–	1.02	28.38	89.52	51.8	5.73
	D^2-CFG (Flat)	0.6	1	–	1.01	28.21	88.99	55.1	7.51
	Standard CFG + Retraction	0	0	✓	1.13	28.66	91.11	44.1	1.14
	DGSCFG + Retraction	0	0	✓	1.14	28.65	91.14	44.3	1.24
	DGSAPG + Retraction	0.6	0	✓	1.16	28.65	91.16	43.7	1.02
	D^2-CFG (Ours)	0.6	1	✓	**1.18**	**28.67**	**91.30**	**43.5**	**0.95**
$w = 15.0$	Standard CFG (Flat)	0	0	–	0.51	26.97	80.61	56.4	24.29
	DGSAPG (Flat)	0.6	0	–	0.66	27.28	82.25	59.7	23.72
	D^2-CFG (Flat)	0.6	1	–	0.66	27.19	81.97	59.6	23.10
	Standard CFG + Retraction	0	0	✓	1.12	28.47	90.84	45.9	2.95
	DGSCFG + Retraction	0	0	✓	1.13	28.52	90.88	46.2	3.14
	DGSAPG + Retraction	0.6	0	✓	1.16	28.65	90.65	45.7	2.53
	D^2-CFG (Ours)	0.6	1	✓	**1.17**	**28.68**	**91.38**	**44.4**	**1.78**

Algorithm 2 D^2-CFG Sampling for Flow Matching (Euler ODE; t runs from noise $t = 0$ to data $t = 1$)

Require: Conditional and unconditional velocity networks $\boldsymbol{v}_\theta(\cdot, t, c)$, $\boldsymbol{v}_\theta(\cdot, t, \varnothing)$; time grid $0 = t_0 < \cdots < t_K = 1$; guidance scale $w \geq 0$; density quantile $q \in (0, 1)$; effective dimension $\tilde{D}$ (1,024 for SD3.5); clean-space projection weight $\gamma_2 \in [0, 1]$ (with $\gamma_1 = 1$ fixed).

Initialize $\boldsymbol{x}_{t_0} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.
for $k = 0$ **to** $K - 1$ **do**
  $\boldsymbol{v}_c \leftarrow \boldsymbol{v}_\theta(\boldsymbol{x}_{t_k}, t_k, c), \quad \boldsymbol{v}_u \leftarrow \boldsymbol{v}_\theta(\boldsymbol{x}_{t_k}, t_k, \varnothing), \quad \Delta\boldsymbol{v} \leftarrow \boldsymbol{v}_c - \boldsymbol{v}_u$
  Endpoints: $\hat{\boldsymbol{x}}_1^c \leftarrow \boldsymbol{x}_{t_k} + (1 - t_k)\boldsymbol{v}_c, \quad \hat{\boldsymbol{x}}_0^c \leftarrow \boldsymbol{x}_{t_k} - t_k \boldsymbol{v}_c$
  Gram–Schmidt (noise endpoint first): $\hat{\boldsymbol{e}}_1 \leftarrow \frac{\hat{\boldsymbol{x}}_0^c}{\|\hat{\boldsymbol{x}}_0^c\|}, \quad \tilde{\boldsymbol{e}}_2 \leftarrow \hat{\boldsymbol{x}}_1^c - \langle \hat{\boldsymbol{x}}_1^c, \hat{\boldsymbol{e}}_1 \rangle \hat{\boldsymbol{e}}_1, \quad \hat{\boldsymbol{e}}_2 \leftarrow \frac{\tilde{\boldsymbol{e}}_2}{\|\tilde{\boldsymbol{e}}_2\|}$
  Project: $\Delta\boldsymbol{v}_\perp \leftarrow \Delta\boldsymbol{v} - \langle \Delta\boldsymbol{v}, \hat{\boldsymbol{e}}_1 \rangle \hat{\boldsymbol{e}}_1 - \gamma_2 \langle \Delta\boldsymbol{v}, \hat{\boldsymbol{e}}_2 \rangle \hat{\boldsymbol{e}}_2$
  Guided endpoints: $\hat{\boldsymbol{x}}_1^{\text{raw}} \leftarrow \hat{\boldsymbol{x}}_1^c + (1 - t_k) w \Delta\boldsymbol{v}_\perp, \quad \hat{\boldsymbol{x}}_0^{\text{raw}} \leftarrow \hat{\boldsymbol{x}}_0^c - t_k w \Delta\boldsymbol{v}_\perp$
  Density factor: $\eta \leftarrow 1 + \frac{D\sqrt{2/\tilde{D}}\,\Phi^{-1}(q)}{\|\hat{\boldsymbol{x}}_0^c\|^2}$ if $\mathrm{logSNR}(t_k) \leq 1$, else $\eta \leftarrow 1$
  Retract: $\hat{\boldsymbol{x}}_1^{\text{g}} \leftarrow \hat{\boldsymbol{x}}_1^{\text{raw}} \frac{\|\hat{\boldsymbol{x}}_1^c\|}{\|\hat{\boldsymbol{x}}_1^{\text{raw}}\|}, \quad \hat{\boldsymbol{x}}_0^{\text{g}} \leftarrow \hat{\boldsymbol{x}}_0^{\text{raw}} \frac{\eta\,\|\hat{\boldsymbol{x}}_0^c\|}{\|\hat{\boldsymbol{x}}_0^{\text{raw}}\|}$
  Step: $\boldsymbol{v}_{\text{D}^2} \leftarrow \hat{\boldsymbol{x}}_1^{\text{g}} - \hat{\boldsymbol{x}}_0^{\text{g}}, \quad \boldsymbol{x}_{t_{k+1}} \leftarrow \boldsymbol{x}_{t_k} + (t_{k+1} - t_k)\, \boldsymbol{v}_{\text{D}^2}$
end for
return $\boldsymbol{x}_{t_K}$.

F.4 Why Flow Matching Benefits More from Endpoint Retraction than Diffusion

Section 3.3 explains the difference through the prefactors of (14) and (41); two points complete the argument. First, in diffusion the prefactor $(\sigma_t/\alpha_t)^2$ does not vanish near noise ($\alpha_t \to 0$), but there $\frac{\sigma_t}{\alpha_t}\Delta\boldsymbol{\epsilon}_\perp$ is the $\mathcal{O}(1)$ difference between the conditional and unconditional clean estimates, so the quadratic term stays bounded and is confined to the early high-noise steps; moreover, $\eta_t(q)$ is gated to $\mathrm{logSNR}(t) \leq 1$ and inactive in the terminal regime, so flat 2-plane projection suffices. Second, in flow matching the coefficients $(1 - t)$ and t sum to one, so $\|\Delta\boldsymbol{v}_\perp\| = \mathcal{O}(1)$ enters the endpoint norms at every step and the growth accumulates across the ODE steps; the retraction removes it exactly. The noise-prediction identity in (14) has no vanishing prefactor, but a DDIM step uses the guided noise prediction only through $\hat{\boldsymbol{x}}_0$ and the renoising term $\sigma_{t-1}\boldsymbol{\epsilon}$, whose weight vanishes as $t \to 0$, so that growth does not accumulate in the state. For $\gamma_2 < 1$ the clean-estimate identity holds

up to a cross term ($-2\frac{\sigma_t}{\alpha_t} w\langle \hat{\boldsymbol{x}}_0^c, \Delta\boldsymbol{\epsilon}_\perp\rangle$ in diffusion) that vanishes as $\gamma_2 \to 1$; the noise identity is exact for any γ_2 because $\gamma_1 = 1$.

G Why the 2-Plane: Constraints, Uniqueness, and Relaxation

This section expands on the construction behind Section 3.3. It shows three things: the projection onto $\mathcal{V}_2^\perp$ is the unique solution of (10); the plane $\mathcal{V}_2$ does not depend on the basis one writes down, because it is the plane of the unguided sampler step; and relaxing $\gamma_2 < 1$ sets the amplitude gain to a prescribed value instead of zero.

Proposition 2 (Minimum-norm solution and its relaxation)**.** *Let $\{\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2\}$ be the Gram–Schmidt basis (11) of $\mathcal{V}_2 = \mathrm{span}\{\boldsymbol{\epsilon}_c, \hat{\boldsymbol{x}}_0^c\}$ with $\boldsymbol{\epsilon}_c$ ordered first, and write $\hat{\boldsymbol{x}}_0^c = p\,\hat{\boldsymbol{e}}_1 + r\,\hat{\boldsymbol{e}}_2$ with $r \neq 0$. For any $\gamma_2 \in [0,1]$, the vector $\Delta\boldsymbol{\epsilon}_\perp^{joint}$ in (12) with $\gamma_1 = 1$ is the unique solution of*

$$\min_{\boldsymbol{g}} \ \tfrac{1}{2}\|\boldsymbol{g} - \Delta\boldsymbol{\epsilon}\|^2 \quad \textit{s.t.} \quad \langle \boldsymbol{g}, \boldsymbol{\epsilon}_c\rangle = 0, \qquad \langle \boldsymbol{g}, \hat{\boldsymbol{x}}_0^c\rangle = (1-\gamma_2)\, r\, \langle \Delta\boldsymbol{\epsilon}, \hat{\boldsymbol{e}}_2\rangle. \tag{45}$$

In particular, $\gamma_2 = 1$ recovers the doubly-constrained problem (10), whose solution is the orthogonal projection $\boldsymbol{g}^\star = \boldsymbol{P}_{\mathcal{V}_2^\perp}\Delta\boldsymbol{\epsilon}$.

Proof. The Lagrangian is $\frac{1}{2}\|\boldsymbol{g} - \Delta\boldsymbol{\epsilon}\|^2 + \lambda_1\langle \boldsymbol{g}, \boldsymbol{\epsilon}_c\rangle + \lambda_2\langle \boldsymbol{g}, \hat{\boldsymbol{x}}_0^c\rangle$; setting its gradient in $\boldsymbol{g}$ to zero gives $\boldsymbol{g} = \Delta\boldsymbol{\epsilon} - \lambda_1\boldsymbol{\epsilon}_c - \lambda_2\hat{\boldsymbol{x}}_0^c$, so $\boldsymbol{g} \in \Delta\boldsymbol{\epsilon} + \mathcal{V}_2$ and the components of $\boldsymbol{g}$ outside $\mathcal{V}_2$ coincide with those of $\Delta\boldsymbol{\epsilon}$. It remains to match the two in-plane coordinates. Since $\boldsymbol{\epsilon}_c = \|\boldsymbol{\epsilon}_c\|\hat{\boldsymbol{e}}_1$, the first constraint forces $\langle \boldsymbol{g}, \hat{\boldsymbol{e}}_1\rangle = 0$. Then $\langle \boldsymbol{g}, \hat{\boldsymbol{x}}_0^c\rangle = r\,\langle \boldsymbol{g}, \hat{\boldsymbol{e}}_2\rangle$, and the second constraint forces $\langle \boldsymbol{g}, \hat{\boldsymbol{e}}_2\rangle = (1-\gamma_2)\langle \Delta\boldsymbol{\epsilon}, \hat{\boldsymbol{e}}_2\rangle$. Both in-plane coordinates are uniquely determined, which yields (12) with $\gamma_1 = 1$; uniqueness follows from strict convexity of the objective under affine constraints. □

The plane does not depend on the basis. Because $\boldsymbol{x}_t = \alpha_t\hat{\boldsymbol{x}}_0^c + \sigma_t\boldsymbol{\epsilon}_c$ with $\alpha_t, \sigma_t > 0$, the pairs $\{\boldsymbol{x}_t, \boldsymbol{\epsilon}_c\}$ and $\{\boldsymbol{x}_t, \hat{\boldsymbol{x}}_0^c\}$ are images of $\{\hat{\boldsymbol{x}}_0^c, \boldsymbol{\epsilon}_c\}$ under invertible 2×2 linear maps (with determinants $\pm\alpha_t$ and $\pm\sigma_t$) and hence span the same plane whenever $\boldsymbol{\epsilon}_c$ and $\hat{\boldsymbol{x}}_0^c$ are linearly independent. The same holds for the score $-\boldsymbol{\epsilon}_c/\sigma_t$ and, for schedules with strictly monotone SNR, for the velocity $\dot{\alpha}_t\hat{\boldsymbol{x}}_0^c + \dot{\sigma}_t\boldsymbol{\epsilon}_c$. The projection $\boldsymbol{P}_{\mathcal{V}_2^\perp}$ is therefore independent of which basis one writes down; we use $\{\boldsymbol{\epsilon}_c, \hat{\boldsymbol{x}}_0^c\}$ because R and ρ_t are precisely the components of the update along these two vectors. Planes built from $\boldsymbol{\epsilon}_u$ instead do not nullify both R and ρ_t, and since $\Delta\boldsymbol{\epsilon} \in \mathrm{span}\{\boldsymbol{\epsilon}_c, \boldsymbol{\epsilon}_u\}$, projecting $\Delta\boldsymbol{\epsilon}$ out of that plane would annihilate guidance entirely.

Proposition 3 (In-plane guidance is a reparameterization of the unguided sampler)**.** *Consider a DDIM / probability-flow step driven by a guided prediction $\boldsymbol{\epsilon}_c + \boldsymbol{g}$,*

$$\boldsymbol{x}_{t-1} \ = \ \frac{\alpha_{t-1}}{\alpha_t}\,\boldsymbol{x}_t \ - \ (c_1 - c_2)\big(\boldsymbol{\epsilon}_c + \boldsymbol{g}\big), \qquad c_1 \coloneqq \tfrac{\alpha_{t-1}}{\alpha_t}\sigma_t, \quad c_2 \coloneqq \sigma_{t-1}. \tag{46}$$

(i) If $\boldsymbol{g} \in \mathcal{V}_2$, i.e., $\boldsymbol{g} = a\,\boldsymbol{x}_t + b\,\boldsymbol{\epsilon}_c$, the step equals an unguided step with modified coefficients,

$$\boldsymbol{x}_{t-1} \ = \ \Big[\tfrac{\alpha_{t-1}}{\alpha_t} - (c_1 - c_2)\,a\Big]\boldsymbol{x}_t \ - \ (c_1 - c_2)(1+b)\,\boldsymbol{\epsilon}_c, \tag{47}$$

so the two in-plane degrees of freedom only rescale the two coefficients of the unguided step (state amplitude and effective score). Because $\boldsymbol{\epsilon}_c$ and $\hat{\boldsymbol{x}}_0^c$ are non-collinear, the map $(a,b) \mapsto (\langle \boldsymbol{g}, \hat{\boldsymbol{x}}_0^c\rangle, \langle \boldsymbol{g}, \boldsymbol{\epsilon}_c\rangle)$ is an invertible linear map, so these two coefficients are in one-to-one correspondence with the pair (ρ_t, R). (ii) Conversely, every step of the unguided sampler lies in $\mathcal{V}_2$ for any schedule.

Proof. For (i), substitute $\boldsymbol{g} = a\boldsymbol{x}_t + b\boldsymbol{\epsilon}_c$ and collect the $\boldsymbol{x}_t$ and $\boldsymbol{\epsilon}_c$ terms. For (ii), the unguided update $\boldsymbol{x}_{t-1} = \alpha_{t-1}\hat{\boldsymbol{x}}_0^c + \sigma_{t-1}\boldsymbol{\epsilon}_c$ with $\hat{\boldsymbol{x}}_0^c = (\boldsymbol{x}_t - \sigma_t\boldsymbol{\epsilon}_c)/\alpha_t$ is a linear combination of $\boldsymbol{x}_t$ and $\boldsymbol{\epsilon}_c$. □

Proposition 3 is what makes the two constraints in (10) exhaustive: $\mathcal{V}_2$ is exactly the set of directions the unguided sampler can already produce at $\boldsymbol{x}_t$, and since $\dim\mathcal{V}_2 = 2$ there are exactly two in-plane effects, the amplitude and density distortions of Proposition 1. D^2-CFG keeps the component of $\Delta\boldsymbol{\epsilon}$ that is new relative to the unguided dynamics ($\mathcal{V}_2^\perp$) and removes the component that is an

uncalibrated, w-coupled rescaling of the sampler itself. Note that the DGS factor $\eta_t(q)$ acts on the $\boldsymbol{\epsilon}_c$ coordinate of the in-plane component and APG's projection on the $\hat{\boldsymbol{x}}_0^c$ coordinate: D^2-CFG does not delete the two in-plane degrees of freedom but re-assigns each to an explicit dial ($\eta_t(q)$ and γ_2) instead of leaving them under w's control.

The retraction is an in-plane correction. The same proposition explains the spherical retraction of Section 3.3. Write $\hat{\boldsymbol{x}}_0^{\text{raw}} = \hat{\boldsymbol{x}}_0^c + \kappa\, \boldsymbol{g}$ with $\boldsymbol{g} \perp \hat{\boldsymbol{x}}_0^c$ and $\kappa = \sigma_t/\alpha_t$ (diffusion) or $\kappa = 1 - t$ for the data endpoint in the flow-matching convention of Appendix F. Rescaling to the original norm gives, to second order in $\kappa\|\boldsymbol{g}\|/\|\hat{\boldsymbol{x}}_0^c\|$,

$$\hat{\boldsymbol{x}}_0^{\text{raw}} \frac{\|\hat{\boldsymbol{x}}_0^c\|}{\|\hat{\boldsymbol{x}}_0^{\text{raw}}\|} - \hat{\boldsymbol{x}}_0^{\text{raw}} \;=\; -\frac{\kappa^2\|\boldsymbol{g}\|^2}{2\,\|\hat{\boldsymbol{x}}_0^c\|^2}\,\hat{\boldsymbol{x}}_0^c \;+\; \mathcal{O}\big(\kappa^3\big), \tag{48}$$

a correction along $\hat{\boldsymbol{x}}_0^c$, and the analogous rescaling of the noise estimate gives a correction along $\boldsymbol{\epsilon}_c$. Both lie in $\mathcal{V}_2$, so by Proposition 3 the retraction only adjusts the two coefficients of the sampler step; it sets the in-plane coordinates so that the norm constraints of (14) hold exactly rather than to first order. The proposition itself carries over to the flow-matching Euler step, since $\boldsymbol{x}_{t\pm\Delta t} = \boldsymbol{x}_t \pm \Delta t\, \boldsymbol{v}_c \in \operatorname{span}\{\boldsymbol{x}_t, \boldsymbol{v}_c\} = \mathcal{V}_2$.

Uniqueness of the relaxation. Gram–Schmidt is order-dependent in general, and the order matters here: only with $\boldsymbol{\epsilon}_c$ orthogonalized first does removing a single basis coordinate enforce $\langle \boldsymbol{g}, \boldsymbol{\epsilon}_c \rangle = 0$. Given that hard constraint ($\gamma_1 = 1$), however, the residual in-plane freedom is the line $\mathcal{V}_2 \cap \boldsymbol{\epsilon}_c^{\perp} = \operatorname{span}\{\hat{\boldsymbol{e}}_2\}$, which is one-dimensional because $\dim \mathcal{V}_2 = 2$ and $\boldsymbol{\epsilon}_c \in \mathcal{V}_2$. Hence $\hat{\boldsymbol{e}}_2$ is determined up to sign by the constraint alone, the family indexed by $\gamma_2 \in [0, 1]$ is unique, and $\rho_t = 0$ if and only if $\gamma_2 = 1$.

H Additional Ablations and Parameter Sweeps

H.1 Analysis of the Joint Projection

To verify that the 2-plane Gram–Schmidt basis does not suffer from collinear collapse ($|\cos\theta| \to 1.0$), we evaluate the collinearity $|\cos(\hat{\mathbf{x}}_0^c, \boldsymbol{\epsilon}_c)|$ across 50 randomly sampled MS-COCO prompts throughout the 30-step DDIM trajectory (Fig. 14). Across all guidance configurations, alignment begins near exact orthogonality at high noise ($t \approx 1000$), peaks at $|\cos\theta| \approx 0.28 \pm 0.05$ ($\theta \geq 70°$) during the intermediate semantic layout phase ($t \approx 300$–500), and relaxes as $t \to 0$. Hence $\mathcal{V}_2 = \operatorname{span}\{\hat{\mathbf{x}}_0^c, \boldsymbol{\epsilon}_c\}$ is uniformly well-conditioned (the Gram–Schmidt step never approaches degeneracy), while $\hat{\mathbf{x}}_0^c$ and $\boldsymbol{\epsilon}_c$ are genuinely non-orthogonal during synthesis.

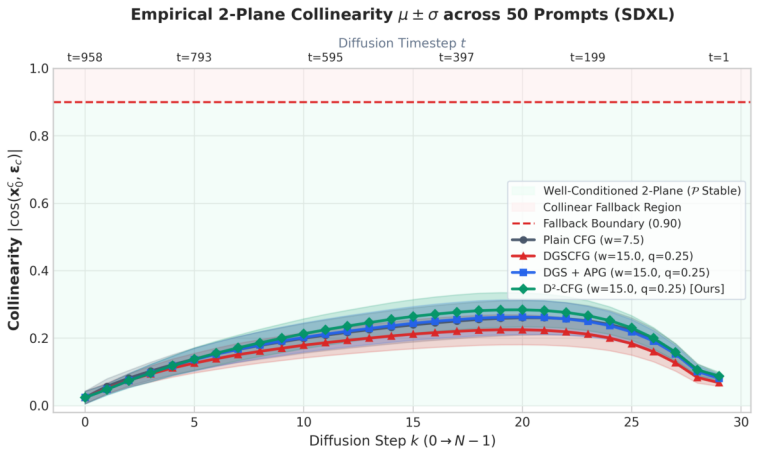
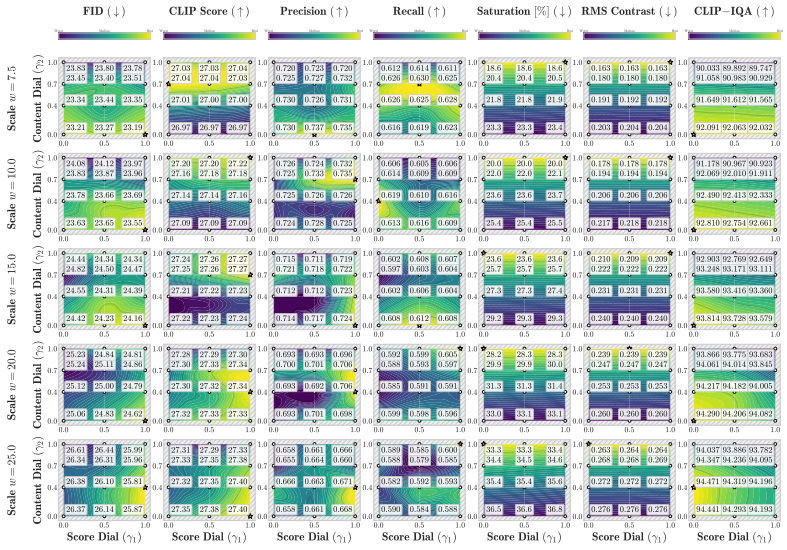


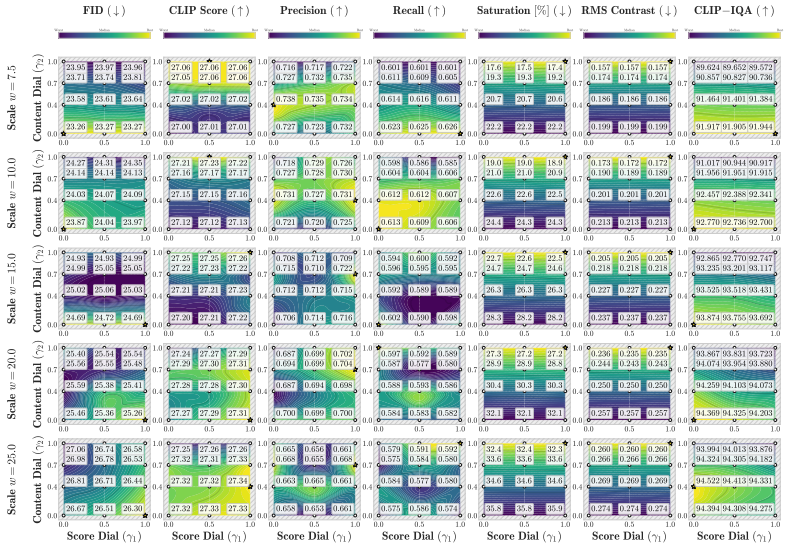
Figure 14: **Empirical 2-Plane Conditioning across 50 SDXL Prompts.** Mean collinearity $|\cos(\hat{\mathbf{x}}_0^c, \boldsymbol{\epsilon}_c)| \pm 1\sigma$ across 50 MS-COCO prompts (30 DDIM steps, $q = 0.25$). Alignment peaks at $|\cos\theta| \approx 0.28$ during structural timesteps ($t \approx 300$–500) and stays far from collinearity (angle $\geq 70°$), confirming that $\mathcal{V}_2 = \operatorname{span}\{\hat{\mathbf{x}}_0^c, \boldsymbol{\epsilon}_c\}$ remains non-degenerate throughout generation. The dashed line at 0.9 marks a diagnostic threshold that is never reached; the method uses no fallback.

H.2 SWEEPING THE PARAMETERS γ_1, γ_2

- **Quantitative behavior on SDXL (Fig. 15):** Enforcing score orthogonality ($\gamma_1 \to 1.0$) isolates semantic steering from density drift, improving FID most clearly at $w \geq 20$ with little change at $w \leq 15$, while preserving stable precision and recall. Meanwhile, γ_2 monotonically controls image saturation and RMS contrast—reducing oversaturation by $> 10\%$ across all scales—with intermediate settings ($\gamma_2 \in [0.4, 0.7]$) maximizing sample recall and CLIP alignment.
- **Qualitative behavior on SD3.5 (Figs. 16–18):** Setting $\gamma_1 = 1.0$ prevents the severe saturation and blowout observed at $\gamma_1 = 0.0$ under extreme guidance ($w = 20.0$), preserving fine micro-textures while ensuring quantile q remains an effective dial. Meanwhile, $\gamma_2 \in [0.4, 0.7]$ strikes an optimal balance between avoiding the chromatic fading of complete projection ($\gamma_2 = 1.0$) and preventing the crushed shadows of unprojected guidance ($\gamma_2 = 0.0$).

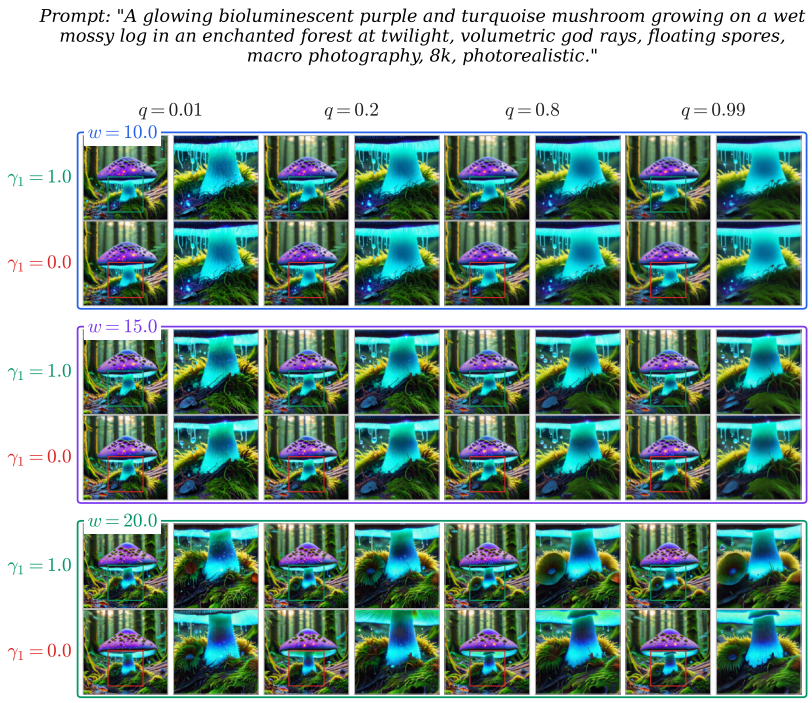


(a) Canonical density regime ($q = 0.30$).



(b) Higher-density regime ($q = 0.70$).

Figure 15: **Parameter sweep of (γ_1, γ_2) across guidance scales in SDXL.** Across both $q = 0.30$ and $q = 0.70$, score orthogonality ($\gamma_1 \to 1.0$) improves fidelity (FID) and text alignment (CLIP) at high guidance, most clearly at $w \geq 20$, justifying fixing $\gamma_1 = 1$. Meanwhile, γ_2 independently governs saturation and contrast (horizontal contours), with intermediate values ($\gamma_2 \approx 0.4$–0.7, stars) optimizing Recall and IQA.



(a) MS-COCO prompt.



(b) Pick-a-Pic prompt.

Figure 16: **Ablation of score projection γ_1 across guidance scales in SD3.5.** Comparing strict orthogonal projection ($\gamma_1 = 1.0$, Ours) against unconstrained score leakage ($\gamma_1 = 0.0$, an APG-like clean-space projection) across $w \in \{10, 15, 20\}$ and quantiles $q \in [0.01, 0.99]$. Fixing $\gamma_1 = 1.0$ preserves density control even at $w = 20$.

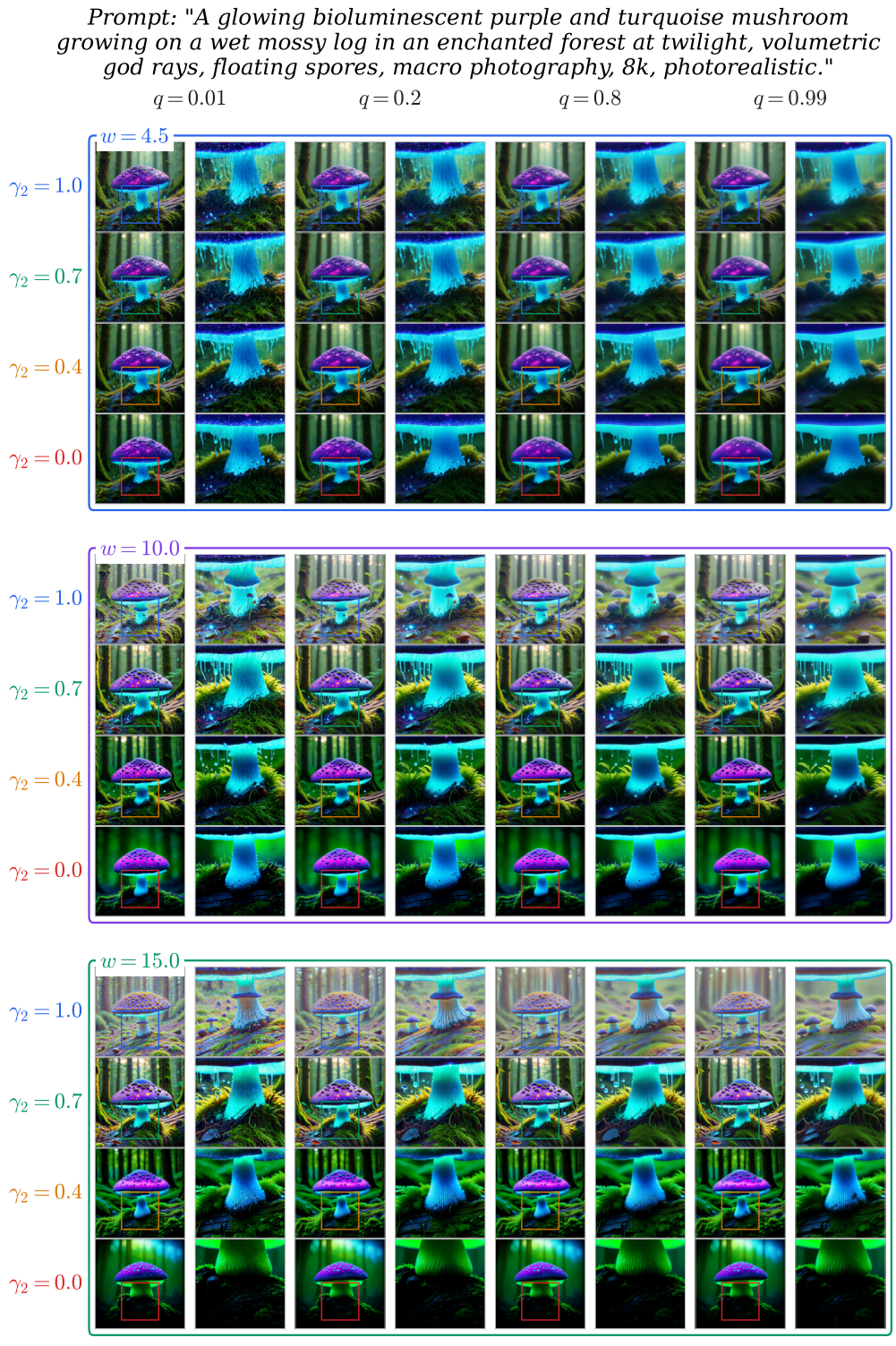


Figure 17: **Ablation of clean-space projection γ_2 across guidance scales in SD3.5 (Prompt 1).** Setting $\gamma_2 = 0.7$ prevents chromatic washout and saturation drift at $w = 15$, retaining sharp details and bioluminescent dynamic range, whereas $\gamma_2 = 1.0$ fades contrast and $\gamma_2 = 0.0$ crushes background shadows.

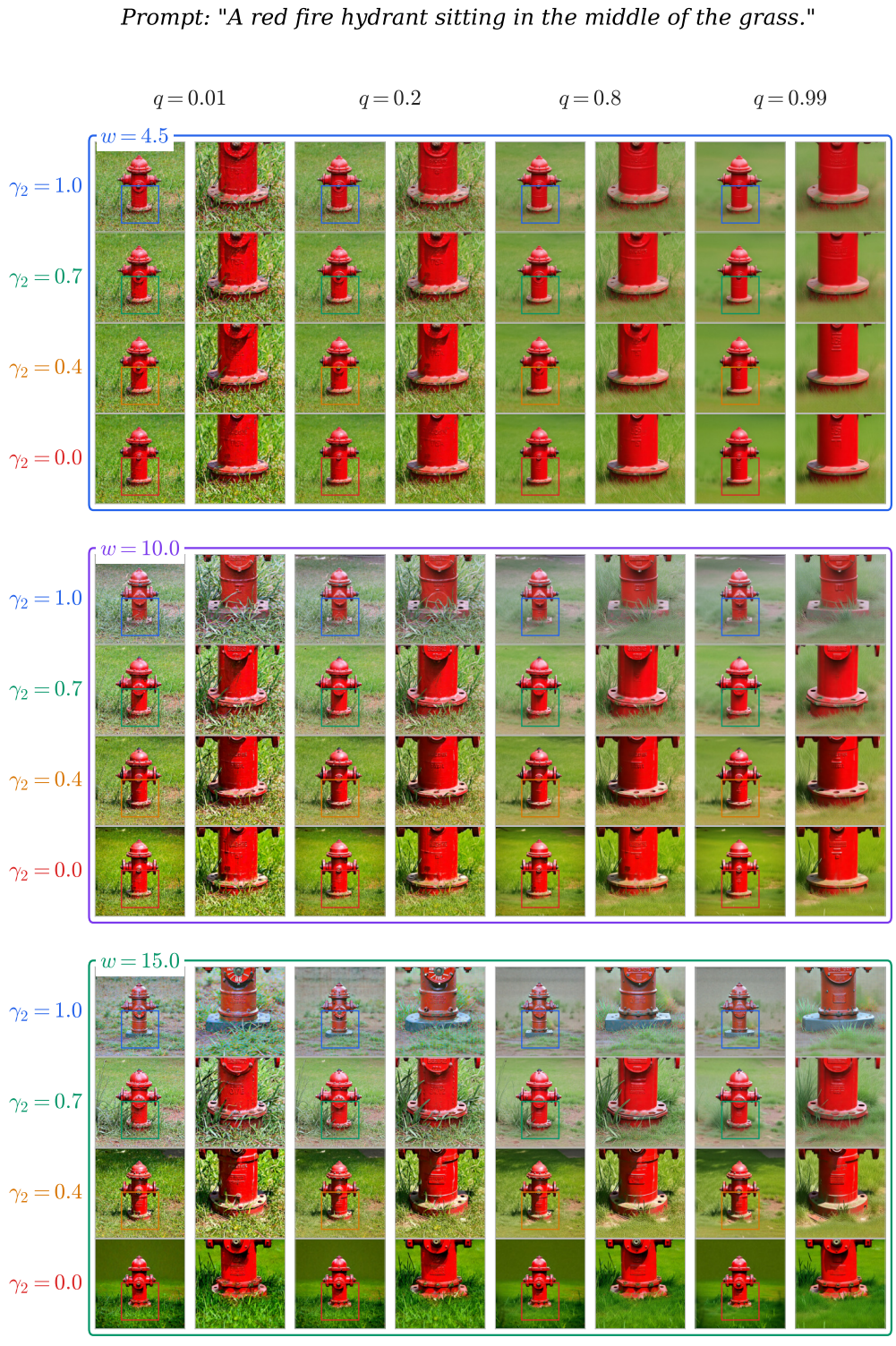


Figure 18: **Ablation of clean-space projection γ_2 across guidance scales in SD3.5 (Prompt 2).** Setting $\gamma_2 = 0.4$ renders balanced image contrast across $w \in \{10, 15\}$, avoiding both the fading of $\gamma_2 = 1.0$ and the contrast clipping of $\gamma_2 = 0.0$.

H.3 SWEEPING THE PARAMETER q

Figure 19 reports quantitative trajectories across density quantiles $q \in [0.05, 0.95]$ on SDXL (MS-COCO 5k) at $w \in \{7.5, 15.0, 20.0\}$, while Figs. 20 and 21 show the corresponding visual transitions at $w = 15.0$ and $w = 20.0$. Because D^2-CFG eliminates score-parallel leakage and bounds radial energy growth, varying q acts as a controlled complexity dial—transitioning smoothly from canonical compositions at high q to rich background and micro-texture detail at low q without triggering color burnout.

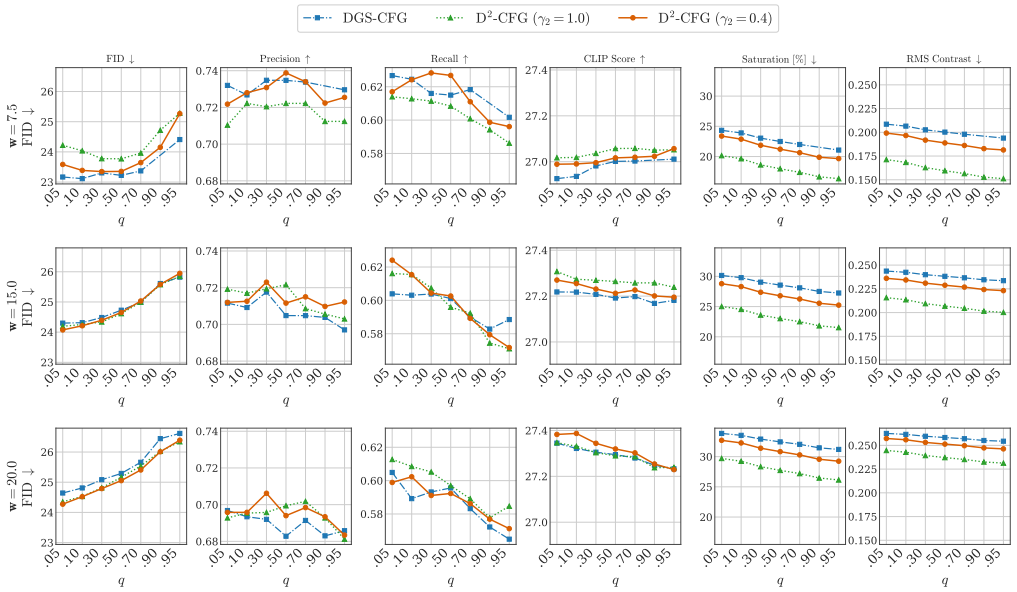
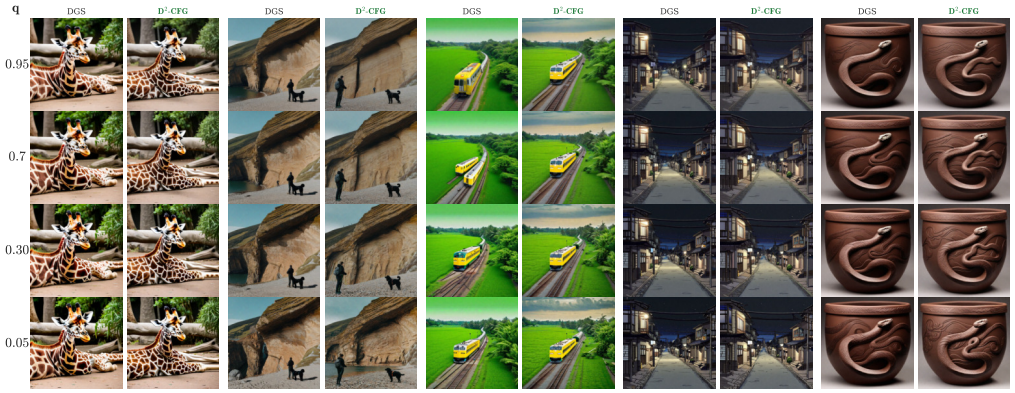
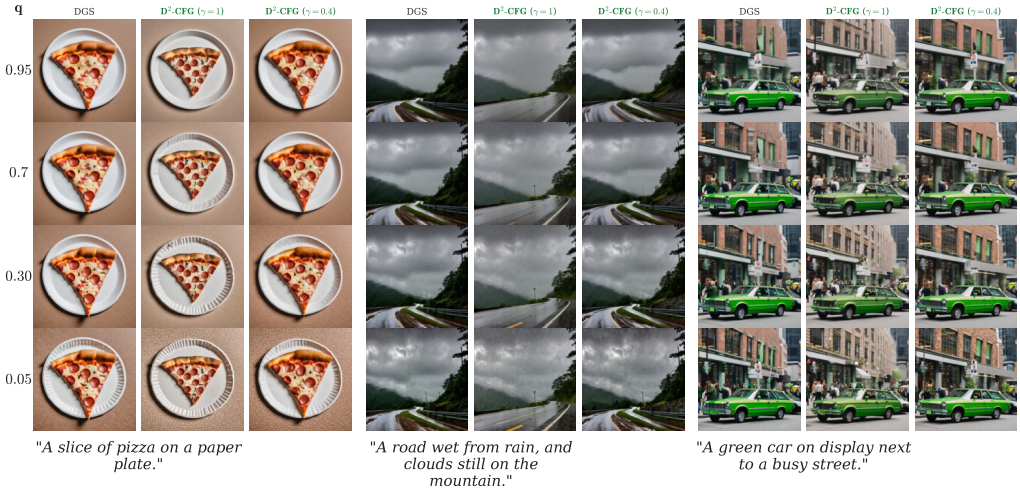


Figure 19: **Quantitative sweeps across density quantile** $q \in [0.05, 0.95]$ at guidance scales $w \in \{7.5, 15.0, 20.0\}$ on SDXL (MS-COCO 5k).

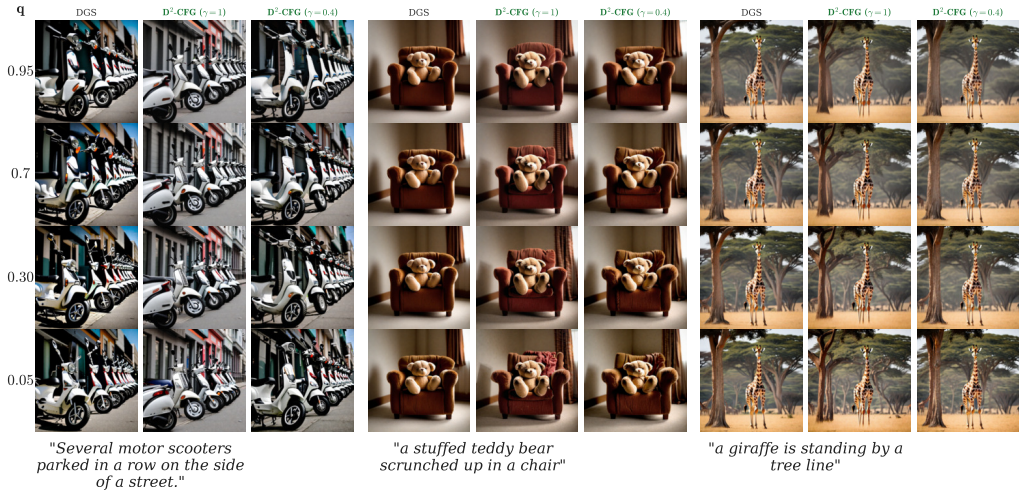


(a) Prompt set 1 at $w = 15.0$.



(b) Prompt set 2 at $w = 15.0$.

Figure 20: **Qualitative comparison across density quantiles $q \in \{0.95, 0.70, 0.30, 0.05\}$ at $w = 15.0$ (SDXL).** Comparing DGSCFG against D^2-CFG ($\gamma_2 = 0.4$); lowering q enriches high-frequency background and texture detail without inducing oversaturation.



(a) Prompt set 1 at $w = 20.0$.



(b) Prompt set 2 at $w = 20.0$.

Figure 21: **Qualitative comparison across density quantiles $q \in \{0.95, 0.70, 0.30, 0.05\}$ at extreme guidance $w = 20.0$ (SDXL).** Even at $w = 20.0$, D^2-CFG maintains density control and natural contrast where DGSCFG saturates.

I D^2-CFG++

In this section, we extend D^2-CFG to CFG++ (Chung et al., 2025).

I.1 Unintended Density Leakage in CFG++

In CFG++, a reverse step from t to $t-1$ uses two different noise predictions, a guided one to form Tweedie's clean estimate $\hat{\boldsymbol{x}}_0$ and the unconditional one to renoise:

$$\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) = \frac{\boldsymbol{x}_t - \sigma_t \boldsymbol{\epsilon}_{\text{denoise}}(\boldsymbol{x}_t)}{\alpha_t}, \qquad \boldsymbol{x}_{t-1} = \alpha_{t-1}\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t) + \sigma_{t-1}\boldsymbol{\epsilon}_{\text{renoise}}(\boldsymbol{x}_t), \tag{49}$$

with $\boldsymbol{\epsilon}_{\text{denoise}} = \boldsymbol{\epsilon}_u + \lambda\Delta\boldsymbol{\epsilon}$ for $\lambda \in [0,1]$ and $\boldsymbol{\epsilon}_{\text{renoise}} = \boldsymbol{\epsilon}_u$.

Proposition 4 (Effective single-noise DDIM and parallel leakage)**.** *Let $c_1 \coloneqq \frac{\alpha_{t-1}}{\alpha_t}\sigma_t$ and $c_2 \coloneqq \sigma_{t-1}$. The two-noise step (49) equals the standard DDIM step $\boldsymbol{x}_{t-1} = \frac{\alpha_{t-1}}{\alpha_t}\boldsymbol{x}_t - (c_1 - c_2)\,\boldsymbol{\epsilon}_{\text{eff}}(\boldsymbol{x}_t)$ driven by*

$$\boldsymbol{\epsilon}_{\text{eff}} = \boldsymbol{\epsilon}_{\text{renoise}} + \kappa_t\big(\boldsymbol{\epsilon}_{\text{denoise}} - \boldsymbol{\epsilon}_{\text{renoise}}\big), \qquad \kappa_t \coloneqq \frac{c_1}{c_1 - c_2} > 1. \tag{50}$$

For CFG++ this gives

$$\boldsymbol{\epsilon}_{\text{eff}}^{\text{CFG++}} = \boldsymbol{\epsilon}_u + \kappa_t\lambda\,\Delta\boldsymbol{\epsilon} = \boldsymbol{\epsilon}_c + w_{\text{eff}}\,\Delta\boldsymbol{\epsilon}, \qquad w_{\text{eff}} \coloneqq \kappa_t\lambda - 1, \tag{51}$$

i.e., standard CFG with the step-dependent weight w_{eff}, so by (8) the step injects the density forcing

$$R(\boldsymbol{x}_t, t; w_{\text{eff}}) = -\frac{\beta(t)\,w_{\text{eff}}}{2\sigma_t^2}\langle\boldsymbol{\epsilon}_c, \Delta\boldsymbol{\epsilon}\rangle \neq 0. \tag{52}$$

Proof. Substituting $\hat{\boldsymbol{x}}_0$ from (49) into $\boldsymbol{x}_{t-1}$ gives $\boldsymbol{x}_{t-1} = \frac{\alpha_{t-1}}{\alpha_t}\boldsymbol{x}_t - c_1\boldsymbol{\epsilon}_{\text{denoise}} + c_2\boldsymbol{\epsilon}_{\text{renoise}}$; matching with the standard step and dividing by $c_1 - c_2$ gives (50). Along the reverse trajectory $\sigma_{t-1} < \sigma_t$ and $\alpha_{t-1} > \alpha_t$, so $c_1 > c_2 > 0$ and $\kappa_t > 1$. Substituting the CFG++ predictions and $\boldsymbol{\epsilon}_u = \boldsymbol{\epsilon}_c - \Delta\boldsymbol{\epsilon}$ gives (51), and (52) is (8) with w replaced by w_{eff}. Since $\mathbb{E}_{\boldsymbol{x}_t|c}[\langle\boldsymbol{\epsilon}_c, \Delta\boldsymbol{\epsilon}\rangle] > 0$ for separated classes (Appendix E.1), CFG++ leaks an uncalibrated score-parallel drift toward high-density regions. □

Naive density guidance fails on CFG++ (DGS-CFG++). Rescaling (51) by $\eta_t(q)$ gives $\boldsymbol{\epsilon}_{\text{eff}} = \eta_t(q)\boldsymbol{\epsilon}_c + \eta_t(q)\,w_{\text{eff}}\,\Delta\boldsymbol{\epsilon}$, and by (9) the density rate splits into $R_{\text{target}}(\eta_t(q)) + \eta_t(q)\,R(\boldsymbol{x}_t, t; w_{\text{eff}})$: the prompt-dependent second term corrupts the dial, and since $\kappa_t \gg 1$ the leak exceeds that of standard CFG at the same λ. This analysis is per DDIM step: κ_t grows as the step size shrinks, so w_{eff} is tied to the number of sampling steps rather than to a continuous-time limit, and the density rates above are per-step quantities.

I.2 The D^2-CFG++ Update and Exact Cancellation

D^2-CFG++ applies the projected update to the clean estimate only and uses the density-rescaled score $\eta_t(q)\boldsymbol{\epsilon}_c$ both as the base prediction and for renoising:

$$\boldsymbol{\epsilon}_{\text{denoise}} = \eta_t(q)\boldsymbol{\epsilon}_c + \lambda\,\Delta\boldsymbol{\epsilon}_\perp^{\text{joint}}, \qquad \boldsymbol{\epsilon}_{\text{renoise}} = \eta_t(q)\boldsymbol{\epsilon}_c. \tag{53}$$

Proposition 5 (Exact density tracking in D^2-CFG++)**.** *Under (53), the effective prediction of a step is*

$$\boldsymbol{\epsilon}_{\text{eff}}^{\text{D}^2\text{-CFG++}} = \eta_t(q)\boldsymbol{\epsilon}_c + \kappa_t\lambda\,\Delta\boldsymbol{\epsilon}_\perp^{\text{joint}}, \tag{54}$$

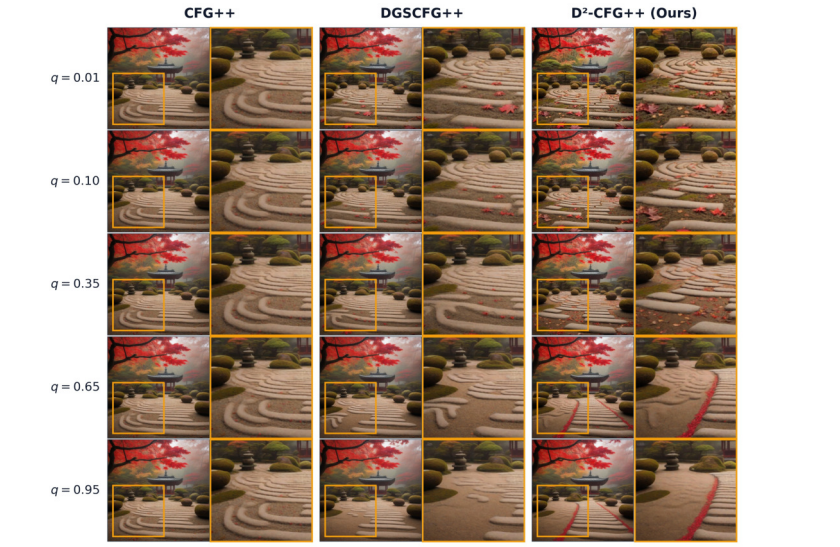
i.e., D^2-CFG with the step-dependent weight $w_t = \kappa_t\lambda$. With $\gamma_1 = 1$, $\langle\boldsymbol{\epsilon}_c, \Delta\boldsymbol{\epsilon}_\perp^{\text{joint}}\rangle = 0$, so the per-step density forcing vanishes and only the intended dial $R_{\text{target}}(\eta_t(q))$ remains.

Proof. Substituting (53) into (50) gives $\boldsymbol{\epsilon}_{\text{eff}} = \eta_t(q)\boldsymbol{\epsilon}_c + \kappa_t\big[(\eta_t(q)\boldsymbol{\epsilon}_c + \lambda\Delta\boldsymbol{\epsilon}_\perp^{\text{joint}}) - \eta_t(q)\boldsymbol{\epsilon}_c\big]$, which is (54). The forcing of this step is $-\frac{\beta(t)\kappa_t\lambda}{2\sigma_t^2}\langle\boldsymbol{\epsilon}_c, \Delta\boldsymbol{\epsilon}_\perp^{\text{joint}}\rangle$, and the inner product vanishes for $\gamma_1 = 1$ because $\hat{\boldsymbol{e}}_1 \propto \boldsymbol{\epsilon}_c$ and $\hat{\boldsymbol{e}}_2 \perp \boldsymbol{\epsilon}_c$ in (11)– (12). □

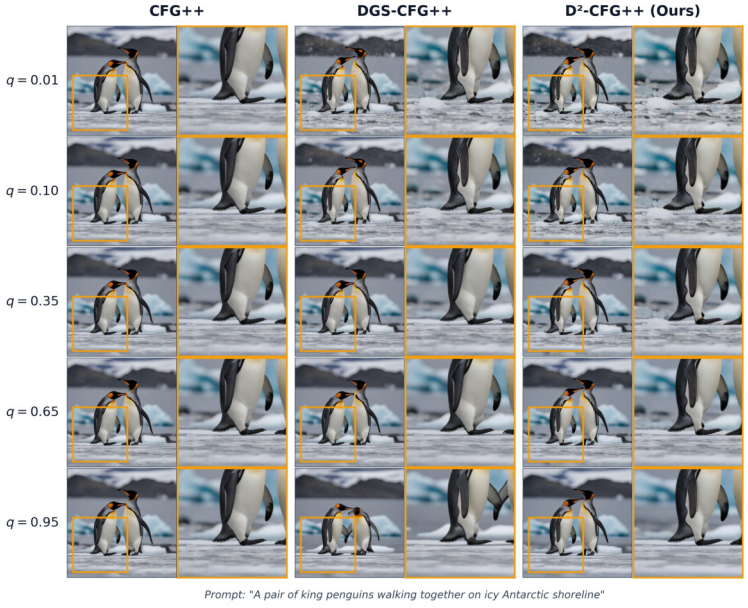
I.3 Experimental Results

We empirically validate Propositions 4 and 5 on SDXL using a 30-step DDIM sampler at guidance scale $\lambda = 0.8$, comparing standard CFG++ (Chung et al., 2025), naive DGS-CFG++, and our proposed D^2-CFG++ ($\gamma_1 = 1, \gamma_2 = 0.4$) across density quantiles $q \in \{0.01, 0.10, 0.35, 0.65, 0.95\}$ in Fig. 22.

- **Prompt adherence and texture traversal (Fig. 22a):** Standard CFG++ has no density dial and omits prompt details (the scattered maple leaves), and DGS-CFG++ inherits the amplified score-parallel forcing $\eta_t(q)\, R(\boldsymbol{x}_t, t; w_{\text{eff}})$ of (52), so the leaves stay suppressed and the raked sand flattens at $q \geq 0.65$. D^2-CFG++ cancels this forcing and recovers a controlled traversal from fine gravel and abundant leaves at low q to regular, smooth geometry at high q.
- **Morphological stability at high density (Fig. 22b):** In DGS-CFG++, the entanglement between $\eta_t(q)$ and the effective update $w_{\text{eff}}\Delta\boldsymbol{\epsilon}$ fuses the two penguins and spawns a phantom limb at $q = 0.95$. D^2-CFG++ keeps the anatomy intact over $q \in [0.01, 0.95]$ while modulating the shoreline texture from fractured ice floes to smooth ice sheets.



(a) Prompt 1: Japanese zen garden ($\lambda = 0.8$).



(b) Prompt 2: Pair of king penguins ($\lambda = 0.8$).

Figure 22: **Qualitative comparison of CFG++, DGS-CFG++, and D^2-CFG++ across density quantiles $q \in [0.01, 0.95]$ on SDXL ($\lambda = 0.8$).** Because $\kappa_t > 1$ amplifies score-parallel leakage ($R(\boldsymbol{x}_t, t; w_{\text{eff}}) \neq 0$), naive DGS-CFG++ washes out raked sand/stepping stones at $q \geq 0.65$ in **(a)** and distorts penguin anatomy at $q = 0.95$ in **(b)**. By enforcing exact score-space cancellation ($R \equiv 0$), D^2-CFG++ preserves structural fidelity and smoothly modulates fine-grained scene detail across q.