UNDERSTANDING INTENSITY SATURATION IN DIFFUSION MODELS THROUGH DATA COVARIANCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models exhibit approximately linear behavior at moderate to high noise levels, but how this linearity manifests in final generated samples remains less understood. We show that projections of initial noise and generated samples onto the same data covariance eigenvector generally exhibit a more linear relationship for eigenvectors with larger eigenvalues. In natural image datasets, leading eigenvectors with larger eigenvalues represent broad brightness variations. Large-magnitude initial noise projections along these directions can therefore contribute to spatially coherent regions becoming extremely bright or dark, a phenomenon we call *intensity saturation*. Classifier-free guidance (CFG) changes both the type of intensity saturation and the saturated area. Without guidance, saturation is more commonly dominated by either dark or bright regions, while their coexistence becomes more frequent as the CFG weight increases. Overall, increasing the CFG weight increases intensity saturation. However, it can increase or decrease the saturated area of individual samples. Our linear analysis suggests that these changes depend on the class and initial noise projections through changes in coefficient magnitudes along the leading eigenvectors. Motivated by these findings, we propose a method that selectively damps large coefficients along these directions to reduce intensity saturation, thereby improving image quality under guidance.

1 INTRODUCTION

Diffusion denoisers (Ho et al., 2020) exhibit approximately linear behavior at moderate to high noise levels. At these noise levels, trained diffusion denoisers are well approximated by the optimal linear denoiser for a multivariate Gaussian distribution defined by the empirical mean and covariance of the training data (Wang & Vastola, 2024; Li et al., 2024). Under this Gaussian approximation, the sampling dynamics decompose along eigenvectors of the data covariance matrix. Coefficients along eigenvectors with larger eigenvalues emerge earlier in the reverse process and contribute to coarse image structure (Weitzner et al., 2025). While this linear structure provides an analytically tractable description of early sampling, its visible consequences in final generated images remain less understood. We therefore ask what visual patterns are determined by these linear dynamics.

Wang & Vastola (2024) derived a linear relationship between initial noise projections and final coefficients along covariance eigenvectors under the Gaussian approximation, and showed that Gaussian linear denoisers capture coarse features of trained model outputs from the same noise. Motivated by these findings, we quantify this linearity in trained diffusion models and show that it generally increases with the covariance eigenvalue, with leading eigenvector coefficients in final samples remaining approximately linear in the corresponding initial noise projections.

These eigenvectors primarily represent broad brightness variations in natural images (Hancock et al., 1992; Simoncelli & Olshausen, 2001; Alsallakh et al., 2022). We show that large-magnitude initial projections along these directions can produce excessively bright or dark regions in final images. We refer to this phenomenon, widely observed in examples across prior studies (Ho & Salimans, 2022; Sadat et al., 2024; Kynkäänniemi et al., 2024; Zheng & Lan, 2024; Chung et al., 2025; Sadat et al., 2025a; Lehman Pavasovic et al., 2026; Xia et al., 2025; Gao et al., 2026) but whose underlying causes have remained unclear, as *intensity saturation*.

We also show that classifier-free guidance (CFG) (Ho & Salimans, 2022) changes the composition and strength of intensity saturation. We distinguish one-sided saturation, dominated by either dark or

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.

bright saturated regions, from two-sided saturation, in which both coexist. One-sided saturation is primarily associated with the top eigenvector, while two-sided saturation involves several leading eigenvectors. Without guidance, the larger eigenvalue associated with the top eigenvector contributes to larger coefficient magnitudes, making one-sided saturation more common. CFG can increase coefficient magnitudes along other leading eigenvectors as well, making two-sided saturation more frequent.

For individual samples, increasing the CFG weight can either increase or decrease the saturated area. Under a linear approximation along the leading few covariance eigenvectors, CFG decomposes into mean shift and variance scaling, both of which depend on the class. Mean shift can increase or decrease coefficient magnitudes depending on the initial noise projections, which tends to increase or reduce intensity saturation. Variance scaling increases or decreases these coefficients, tending to increase or decrease the saturated area, respectively.

This analysis motivates a targeted strategy for mitigating excessive intensity saturation under guidance. Our method adds a regulation step to sampling. This step selectively damps large coefficients along the leading covariance eigenvectors. At each sampling step, the damping strength depends on the noise scale and coefficient magnitude. Reducing extreme coefficients suppresses excessive intensity saturation, thereby improving generated image quality.

Our contributions are as follows.

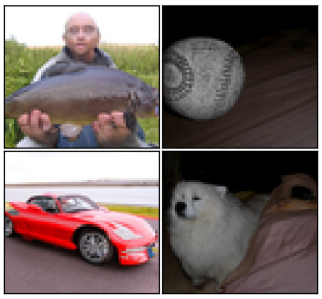
- We show that coefficients along leading covariance eigenvectors in generated samples are approximately linear functions of the corresponding initial noise projections. Large coefficients along these directions produce excessively bright or dark regions, which we define as intensity saturation.
- We characterize how CFG changes the composition of intensity saturation and explain why it increases saturation in some samples and decreases it in others.
- Based on these findings, we propose a method that selectively damps large coefficients along leading covariance eigenvectors, reducing intensity saturation and improving image quality.

2 Related Work

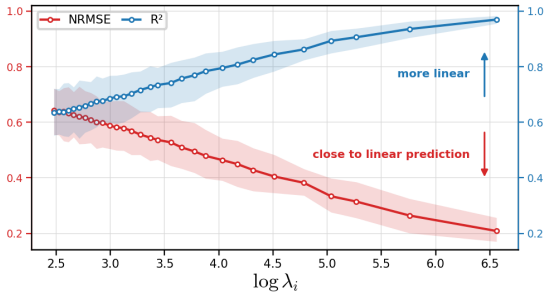
Linear interpretations of diffusion models. Diffusion models learn second-order statistics before higher-order structure, with high-variance covariance directions generally learned earlier than low-variance directions (Wang & Pehlevan, 2025; Bardone et al., 2026). Their denoisers become increasingly linear as they generalize (Li et al., 2024). Gaussian approximations capture denoisers and early sampling dynamics at moderate to high noise levels (Wang & Vastola, 2024; Pierret & Galerne, 2025), explaining the early emergence of coarse, low-frequency structure (Weitzner et al., 2025). Patch PCA reveals compositional cues shared across models (Wang et al., 2025), while image correlations explain locality in linear and deep denoisers (Lukoianov et al., 2025). Li et al. (2025) further analyze CFG through mean shift and variance scaling.

Relationships between initial noise and generated samples. Diffusion models sample initial noise from a standard Gaussian prior, with individual noise samples influencing generation quality and visual characteristics. Several studies improve generation by selecting, optimizing, or refining initial noise (Guo et al., 2024; Qi et al., 2024; Samuel et al., 2024; Eyring et al., 2024; Zhou et al., 2025; Ahn et al., 2026). Noise schedule and signal-leak analyses reveal effects on brightness and color (Lin et al., 2024; Everaert et al., 2024), while seed comparisons show consistent color and compositional patterns (Xu et al., 2025). Linear analyses further show that Gaussian denoisers and trained models produce similar coarse structure from the same initial noise (Wang & Vastola, 2024; Li et al., 2024). Building on these findings, we quantify initial-to-final coefficient linearity along covariance eigenvectors and relate it to *intensity saturation*.

Failures and mitigation at high CFG weights. Classifier-free guidance (CFG) improves conditional fidelity at the cost of diversity (Ho & Salimans, 2022), while large guidance weights can produce oversaturation and artifacts (Saharia et al., 2022; Sadat et al., 2025a). Existing approaches include dynamic thresholding of pixel intensities in predicted clean images (Saharia et al., 2022),



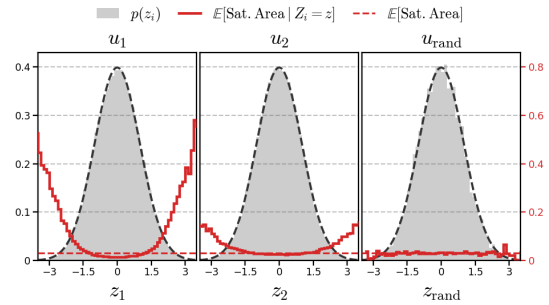
(a) Generated samples



(b) Initial-to-final projection linearity



(c) Predicted saturated regions



(d) Initial noise projection density

Figure 1: **Intensity saturation and its linearity from the initial noise.** (a) Generated samples with intensity saturation, defined as all three RGB channel values being greater than 0.95 or all three being less than -0.95 on $[-1, 1]$. Saturated area is the proportion of pixels in an image that exhibit intensity saturation. (b) Linearity R^2 between initial noise projections and final coefficients, and NRMSE of linear denoiser predictions, versus covariance eigenvalue λ_i. Each point represents one eigenvector, averaged over 50 classes with 10,000 noise samples per class with the EDM (Karras et al., 2022). (c) Saturated regions predicted from initial noise. (d) Initial noise projection densities along the top two eigenvectors and a random unit vector. Further details are provided in the Appendix A.

projection and rescaling of guidance updates (Zhang et al., 2024; Sadat et al., 2025a), frequency-dependent guidance (Sadat et al., 2025b), and manifold-constrained sampling (Chung et al., 2025). We study *intensity saturation*, in which image regions become excessively dark or bright, through coefficients along leading data covariance eigenvectors.

3 Predictable Visual Patterns in Diffusion Sampling

In this section, we investigate how the linear dynamics of diffusion sampling produce characteristic spatial patterns in generated images. We first examine the linearity of final coefficients along leading covariance eigenvectors from the corresponding initial noise projections. We then relate these coefficients to broad intensity patterns and examine whether saturated regions can be predicted from the initial noise.

3.1 Linearity of Eigenvector Coefficients

Prior work shows that, at high and moderate noise levels, trained diffusion denoisers are well approximated by the optimal linear denoiser obtained from the Gaussian mean and covariance of the training data (Wang & Vastola, 2024; Li et al., 2024). For class c, let μ_c and $\Sigma_c = \sum_i \lambda_{c,i} u_{c,i} u_{c,i}^\top$ denote the corresponding mean and covariance. Under this approximation, the linear denoiser is

$$D_L(x_t, \sigma_t) = \mu_c + \sum_i \frac{\lambda_{c,i}}{\lambda_{c,i} + \sigma_t^2} u_{c,i}^\top (x_t - \mu_c) u_{c,i}. \tag{1}$$

The corresponding probability-flow trajectory from (x_T, σ_T) is

$$x_t = \mu_c + \sum_i \sqrt{\frac{\lambda_{c,i} + \sigma_t^2}{\lambda_{c,i} + \sigma_T^2}}\, u_{c,i}^\top (x_T - \mu_c) u_{c,i}. \tag{2}$$

Prior analyses of these dynamics further show that low-frequency, high-variance features emerge early in the reverse process, with their emergence governed by the covariance eigenvalues (Wang & Vastola, 2024; Weitzner et al., 2025). These findings motivate the hypothesis that final coefficients along eigenvectors with larger eigenvalues have higher linear predictability from the corresponding initial noise projections. We test this hypothesis using initial noise samples paired with their corresponding generated samples.

For class c, we define the linearity of the final coefficient along eigenvector $u_{c,i}$ from the corresponding initial noise projection as the squared correlation

$$R_{c,i}^2 \coloneqq \mathrm{Corr}^2\left(u_{c,i}^\top (x_T - \mu_c), u_{c,i}^\top (x_0 - \mu_c)\right), \tag{3}$$

where the correlation is computed across initial noise samples paired with their corresponding generated samples. Within each class, the linearity of the final coefficients largely follows the ordering of the covariance eigenvalues (Figure 14). The analysis and empirical results in Appendix B support this approximate ordering, which we summarize as

$$\lambda_{c,i} \geq \lambda_{c,j} \quad \Longrightarrow \quad R_{c,i}^2 \gtrsim R_{c,j}^2, \tag{4}$$

where $\gtrsim$ indicates that the ordering holds for a large majority of the evaluated eigenvector pairs within classes, with exceptions quantified in Appendix B. This relationship is consistent with the earlier emergence of features along eigenvectors with larger eigenvalues during reverse sampling. These features emerge at higher noise levels, where the trained denoiser is more closely approximated by the linear denoiser. The approximately linear dynamics at these noise levels help explain the strong linear relationship between the corresponding initial noise projections and final coefficients.

Figure 1(b) supports this eigenvalue dependent ordering. For eigenvectors with larger λ_i, the initial and final projections exhibit higher R_i^2. The final coefficients also more closely match the predictions of the linear denoiser, as measured by lower NRMSE. These relationships are also observed in the latent space of latent diffusion models (Appendix C).

3.2 Visual Patterns from Eigenvector Coefficients



u_1 u_2 u_3

Figure 2: **Eigenvectors of ImageNet.**

The linearity established above concerns scalar coefficients, and its visible consequence in generated images depends on the spatial patterns that the leading eigenvectors represent. Prior analyses of natural-image covariance show that leading eigenvectors represent global brightness variations, including brightness contrasts between large image regions (Hancock et al., 1992; Simoncelli & Olshausen, 2001; Alsallakh et al., 2022). Figure 2 shows the three leading eigenvectors of the covariance matrix computed from ImageNet (Deng et al., 2009). The leading covariance eigenvectors for individual classes closely align with those of the overall data distribution (Appendix D) and therefore also represent brightness variations. Leading covariance eigenvectors in latent space also correspond to global brightness variations (Appendix E).

Because these components represent global brightness variations, the approximate linearity of their coefficients implies that the corresponding brightness patterns in the final image vary approximately linearly with the initial noise projections. Through this linear relationship, a large-magnitude initial projection onto a leading eigenvector can contribute to excessively dark or bright intensities over a spatially coherent image region.

Figure 1(a) shows such dark or bright regions even in samples generated without guidance. We refer to this phenomenon as *intensity saturation*. Figure 1(d) shows that the mean saturated area increases with the magnitude of the initial noise projections onto leading eigenvectors, with a particularly pronounced increase along the top eigenvector. This association, together with the linear predictability of the final coefficients along these directions established in Section 3.1, suggests that saturated regions are partially predictable from the initial noise.

We therefore evaluate how accurately the linear denoiser approximation predicts saturated regions directly from the initial noise. Figure 1(c) illustrates the spatial correspondence between predicted and observed saturated regions, with more examples in Appendix F. Table 1 compares linear prediction with random Gaussian samples matched to the mean and covariance of generated images. Metrics are evaluated over all generated samples and over samples whose saturated area is at least 10% of the image. The higher precision and recall of linear prediction support the partial predictability of saturated regions from the initial noise. Further experimental details are provided in Appendix G.

Table 1: **Prediction of saturated regions from initial noise.**

Satu. Area		Precision	Recall
All	Random Gaussian	0.035	0.036
	Linear prediction	0.217	0.329
$\geq 10\%$	Random Gaussian	0.261	0.044
	Linear prediction	0.800	0.313

4 How Classifier-Free Guidance Modulates Intensity Saturation

Building on Section 3, we examine how CFG changes the type and saturated area of intensity saturation across guidance weights. We then relate these changes to initial noise projections, using linear CFG to interpret how guidance can increase or reduce saturation in individual samples.

4.1 Effects of CFG on One-Sided and Two-Sided Saturation



Figure 3: Examples of one-sided and two-sided saturation.

We examine how guidance changes the relative proportions of two types of intensity saturation (Figure 3). We distinguish *one-sided* saturation, dominated by either dark or bright saturated regions, from *two-sided* saturation, in which substantial dark and bright saturated regions coexist within the same sample.

These two types of saturation exhibit different associations with coefficients along the leading covariance eigenvectors. The regional decomposition in Figure 5(a) shows that the contribution along the leading eigenvector u_1 is dominant in one-sided saturation. Consistently, removing the component along u_1 from the initial noise substantially reduces the corresponding saturated area, as shown in Figure 5(b).

For two-sided saturation, Figure 5(a) shows substantial contributions along several leading eigenvectors. Removing the component along u_1 or another leading eigenvector from the initial noise yields comparable reductions in the saturated area in Figure 5(b), supporting the involvement of multiple leading eigenvectors. This observation is consistent with the spatial brightness contrasts represented by additional leading covariance eigenvectors (Figure 2).

CFG increases mean coefficient magnitudes along multiple leading covariance eigenvectors in generated samples, as shown in Appendix H. Since several leading eigenvectors contribute substantially to two-sided saturation, this increase helps explain why larger CFG weights lead to a higher proportion of saturated samples exhibiting two-sided saturation.

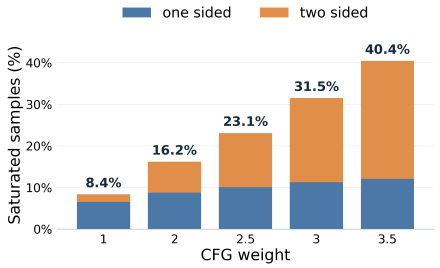
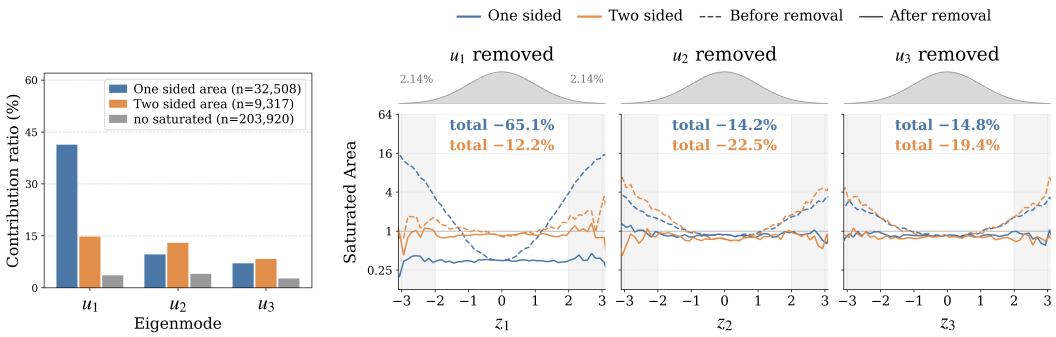


Figure 4: Intensity saturation across CFG weights

Figure 4 shows the percentage of generated samples with a saturated area of at least 10% of the image, grouped by saturation type. As the CFG weight increases, the total percentage increases substantially, with the percentage exhibiting two-sided saturation increasing more rapidly than that exhibiting one-sided saturation.

Although increasing the CFG weight increases the percentage of saturated samples, it can increase or decrease the saturated area of individual samples. For some initial noise samples, CFG reduces the



(a) Eigenmode contributions

(b) Normalized saturated area before and after mode removal

Figure 5: **Leading eigenvector contributions to intensity saturation.** Saturation is *one-sided* when the smaller dark or bright saturated area is below 5% of the larger, and *two-sided* otherwise. (a) We use the same 10,000 initial noise samples per class across 50 classes as in Figure 1(b,d). Samples with saturated area $\geq 10\%$ are classified by type. Those below 1% form the "no saturated" group. Contributions to saturated-region pixel magnitudes concentrate on u_1 for one-sided saturation and span several leading eigenvectors for two-sided saturation. The no-saturation group shows more evenly distributed contributions. (b) $\mathbb{E}[A \mid Z_i = z_i]/\mathbb{E}[A]$, where A is saturated area, before and after removing the initial noise component along u_i. Removing u_1, u_2, u_3 reduces one-sided saturated area by $65.1\%, 14.2\%, 14.8\%$ and two-sided area by $12.2\%, 22.5\%, 19.4\%$, respectively. Results for additional eigenvectors appear in the Appendix I.

saturated area compared with the sample generated from the same initial noise without guidance. We next examine how changes in coefficients along leading eigenvectors help explain how this response varies with the initial noise projections.

4.2 Effects of CFG on Saturation across Initial Noise Projections

Large coefficient magnitudes along leading covariance eigenvectors contribute to intensity saturation, as established in Sections 3 and 4.1. Overall, CFG increases coefficient magnitudes and saturated area. For individual samples, however, CFG can increase or decrease coefficient magnitudes depending on the class and initial noise. Increases in coefficient magnitudes under CFG tend to increase saturated area, while decreases tend to yield a smaller increase or a reduction in saturated area. Appendix J examines this relationship along the top eigenvector.

We focus on the top eigenvector because of its substantial contribution to intensity saturation without guidance (Figure 5). Its strong initial-to-final coefficient linearity and close agreement with linear denoiser predictions (Figure 1(b)), together with the alignment between conditional and unconditional eigenvectors (Appendix D), motivate using linear CFG to analyze changes in its coefficient and trace these tendencies to the initial noise projections. We empirically evaluate the accuracy of the linear CFG approximation in Appendix K.

For conditional and unconditional denoisers D_c and D_{uc}, CFG is defined as

$$D_{\text{cfg}} = D_{uc} + w(D_c - D_{uc}), \tag{5}$$

where $w = 1$ gives unguided conditional sampling. Following the linear CFG analysis of Li et al. (2025), we assume that the conditional and unconditional covariances share eigenvectors $u_{c,i}$, with corresponding eigenvalues $\lambda_{c,i}$ and $\lambda_{uc,i}$. The conditional linear trajectory has coefficient

$$a_i(\sigma_t) = \sqrt{\frac{\lambda_{c,i} + \sigma_t^2}{\lambda_{c,i} + \sigma_T^2}}\, u_{c,i}^\top (x_T - \mu_c), \tag{6}$$

where μ_c is the conditional mean. Linear CFG changes this coefficient as

$$a_i^{\text{cfg}}(\sigma_t) = \underbrace{\left(\frac{\lambda_{uc,i} + \sigma_T^2}{\lambda_{c,i} + \sigma_T^2}\,\frac{\lambda_{c,i} + \sigma_t^2}{\lambda_{uc,i} + \sigma_t^2}\right)^{(w-1)/2} a_i(\sigma_t)}_{\text{variance scaling}} + \underbrace{(w-1)B_i(\sigma_t, c)\, u_{c,i}^\top(\mu_c - \mu_{uc})}_{\text{mean shift}}. \tag{7}$$

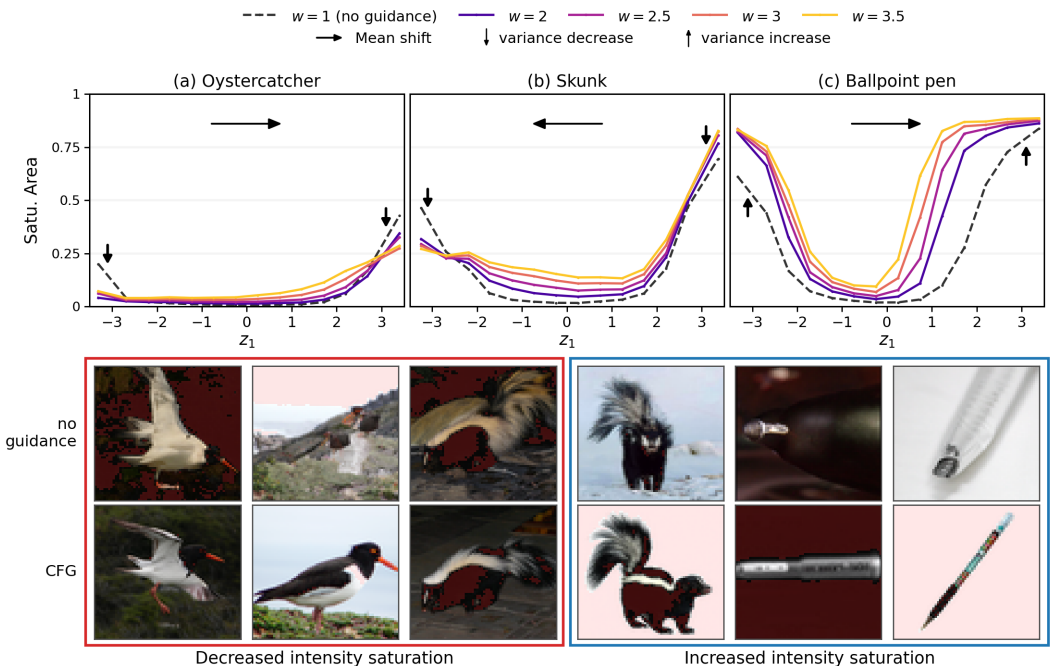


Figure 6: **Initial noise projections and changes in saturated area with CFG.** We evaluate mean shift and variance scaling predictions using the same 10,000 initial noise samples per class across 50 classes at $w = 1, 2, 2.5, 3, 3.5$. Curves show mean saturated area versus the initial noise projection z_1 onto the top eigenvector for three representative classes. Image pairs compare $w = 1$ and CFG samples from the same initial noise at $z_1 = -2, 2$, with saturated regions marked in red. Saturated area generally increases more in the mean shift direction and less, or decreases, in the opposite direction. At large $|z_1|$, changes align with predicted decreases or increases in variance, illustrating how CFG can increase or decrease saturation depending on the initial noise projection.

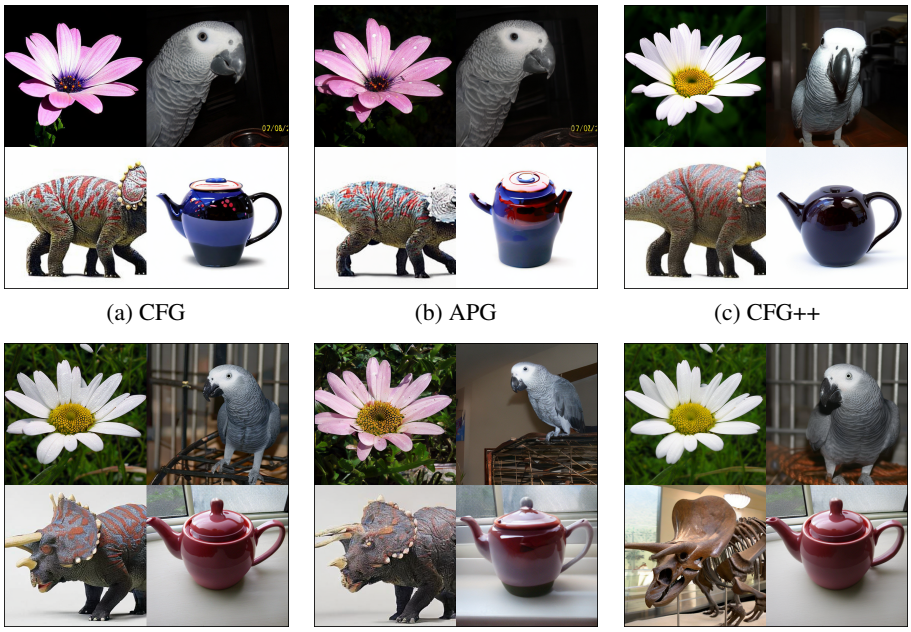
Here, μ_{uc} is the unconditional mean, and $B_i(\sigma_t, c)$ depends on the covariance eigenvalues, CFG weight, and noise levels, independently of x_T, with further details on this expression provided in Appendix L.

For the same initial noise, Equation 7 changes the unguided coefficient through two effects.

- **Mean shift.** The additive term is constant across initial noise within each class, increasing the coefficient magnitude for initial projections in the shift direction and decreasing it for those in the opposite direction.
- **Variance scaling.** For $w > 1$, the multiplicative factor decreases coefficient magnitudes when $\lambda_{c,i} < \lambda_{uc,i}$ and increases them when $\lambda_{c,i} > \lambda_{uc,i}$, with larger changes at larger initial projection magnitudes.

Under this approximation, their combined effect predicts whether the coefficient magnitude increases or decreases for a given initial projection. Increases in the top eigenvector coefficient magnitude tend to increase saturated area, while decreases can yield a smaller increase or a reduction in saturated area.

Figure 6 illustrates these tendencies for three classes. Saturated area generally increases more in the mean shift direction and less, or decreases, in the opposite direction. At large projection magnitudes, changes also follow the predicted decrease or increase in the magnitude of the top eigenvector coefficient. These tendencies are broadly observed across classes, with results for all 50 classes provided in Appendix M. Although approximation errors and contributions from other leading eigenvectors that substantially affect intensity saturation under CFG limit this analysis, it captures qualitative tendencies in how saturation changes with the initial noise projections.



(a) CFG (b) APG (c) CFG++

(d) CFG with Ours (e) APG with Ours (f) CFG++ with Ours

Figure 7: Images from the same EDM2-XXL experiments reported in Table 2, generated using (a) CFG, (b) APG, and (c) CFG++, with (d), (e), and (f) showing the corresponding guidance methods augmented with our method. In these examples, our method reduces the intensity saturation observed in the images generated using the original guidance methods.

5 Mitigating Intensity Saturation

Building on the analysis in Sections 3 and 4, we propose adding a regulation step to sampling to reduce excessive intensity saturation under strong guidance in diffusion models. A related approach is dynamic thresholding (Saharia et al., 2022), which adaptively clips and rescales the predicted clean image in pixel space. In our experiments, applying dynamic thresholding directly in latent space reduces intensity saturation while substantially degrading image quality (Appendix N).

Our regulation step uses the relationship between coefficients along leading covariance eigenvectors and intensity saturation examined in Sections 3 and 4. Large coefficient magnitudes along these directions contribute to broad intensity saturation, while variation along the same directions contributes to brightness diversity. To reduce excessive saturation while limiting changes to smaller coefficients, we apply stronger regulation to larger coefficient magnitudes and weaker regulation to smaller ones. To reduce the cost of estimating statistics for every class, we approximate the mean and covariance of each class with the unconditional mean and covariance.

We apply the regulation step as follows.

$$x_t \leftarrow x_t + \sum_{k=1}^{r} \big(g_k(\sigma_t) - 1\big)\tilde{a}_k(\sigma_t)u_k, \tag{8}$$

where $\tilde{a}_k(\sigma_t) = u_k^\top (x_t - \mu_{uc})$ is the coefficient measured during sampling using the unconditional mean μ_{uc} and covariance eigenvector u_k, and r is the number of leading eigenvectors used for regulation. The factor $0 < g_k(\sigma_t) \leq 1$ reduces large coefficients more strongly according to their magnitude relative to the reference scale $\sqrt{\lambda_k + \sigma_t^2}$, where λ_k is the corresponding unconditional covariance eigenvalue. Further method details are provided in Appendix O.

Table 2 compares CFG, APG (Sadat et al., 2025a), CFG++ (Chung et al., 2025), and guidance interval (Kynkäänniemi et al., 2024) with and without our method on EDM2-S (Karras et al., 2024), EDM2-XXL, and DiT-XL/2 (Peebles & Xie, 2023). We measure intensity saturation using our proposed Excess Clip Area at a 10% saturated area threshold (ECA@10), with lower values indicating

Table 2: Evaluation of CFG, APG, CFG++, and guidance interval with and without our regulation step. Metrics are evaluated using 50,000 generated images. Bold values indicate improvements over the corresponding baseline without our method. Detailed experimental settings are provided in Appendix P.

	CFG		APG		CFG++		guidance interval	
Metric	Base	+ ours	Base	+ ours	Base	+ ours	Base	+ ours
EDM2-S ($w = 4$)								
ECA@10 ↓	8.83	**6.26**	2.24	**0.84**	9.66	**6.82**	3.56	**2.10**
HSV sat. ↓	0.44	**0.42**	0.32	**0.29**	0.46	**0.44**	0.38	**0.36**
Precision ↑	0.79	**0.80**	0.83	**0.84**	0.81	**0.82**	0.80	**0.81**
HPSv2.1 ↑	24.74	**24.94**	23.85	**23.95**	24.50	**24.76**	24.60	**24.68**
FD_{DINOv2} ↓	94.5	**92.3**	66.3	68.5	108.0	**102.9**	55.4	55.7
Recall ↑	0.37	0.36	0.49	0.47	0.29	0.29	0.49	0.48
EDM2-XXL ($w = 2$)								
ECA@10 ↓	3.01	**1.47**	1.77	**0.61**	3.25	**1.63**	2.01	**0.94**
HSV sat. ↓	0.37	**0.34**	0.32	**0.29**	0.38	**0.35**	0.33	**0.31**
Precision ↑	0.85	**0.87**	0.78	**0.79**	0.87	**0.88**	0.80	**0.81**
HPSv2.1 ↑	24.97	**25.14**	23.36	**23.52**	24.49	**24.65**	24.03	**24.12**
FD_{DINOv2} ↓	32.6	32.8	44.7	46.0	50.8	**48.7**	32.2	32.2
Recall ↑	0.49	0.47	0.60	0.58	0.40	0.39	0.58	0.57
DiT-XL/2 ($w = 4$)								
ECA@10 ↓	8.00	**4.89**	5.73	**3.17**	5.01	**2.97**	3.64	**1.15**
HSV sat. ↓	0.45	**0.44**	0.42	**0.40**	0.42	**0.40**	0.38	**0.35**
Precision ↑	0.89	**0.91**	0.88	**0.89**	0.88	**0.90**	0.65	0.65
HPSv2.1 ↑	25.62	**26.05**	25.03	**25.38**	25.72	**25.98**	22.47	**22.58**
FD_{DINOv2} ↓	96.4	**93.8**	64.4	**62.2**	76.19	76.59	83.3	85.7
Recall ↑	0.20	0.19	0.32	0.31	0.26	0.25	0.64	0.64

less excessive intensity saturation. Experimental settings and metric definitions are provided in Appendix P.

Overall, our method reduces ECA@10, consistent with its intended effect of reducing intensity saturation. These reductions are generally accompanied by improved precision (Kynkäänniemi et al., 2019) and HPSv2.1 (Wu et al., 2023), suggesting an association between reduced intensity saturation and improved image quality. Samples with larger reductions in excess saturated area generally show greater improvements in image quality, as examined in Appendix Q. Our method also lowers HSV saturation (Sadat et al., 2025a), consistent with reduced color oversaturation.

Recall generally decreases, suggesting a trade-off between image quality and diversity. The results in Appendix T support the interpretation that brightness changes induced by our method contribute to part of the recall decrease. The effects of CFG weight and our method's regulation hyperparameters on performance are examined in Appendices R and S. Our method also reduces intensity saturation while improving image quality in SD 3.5 (Esser et al., 2024) (Appendix U).

6 Discussion and Conclusion

Intensity saturation has been widely observed in prior studies, while its underlying causes have remained unclear. To our knowledge, we are the first to explain this phenomenon through a linear analysis of diffusion models. Our analysis accounts for intensity saturation without guidance and captures qualitative tendencies in how classifier-free guidance increases or decreases saturated area. Based on these findings, we propose a regulation method that reduces intensity saturation and improves image quality. Our linear analysis can guide the development of more effective mitigation methods. We also plan to investigate how intensity saturation influences the later, nonlinear stages of diffusion sampling.

AI USE STATEMENT

In this work, we used generative AI tools to provide feedback on the design of our experiments and to implement parts of the experimental code. We have not used generative AI tools to propose or refine hypotheses, to generate synthetic datasets, to develop theoretical models or conceptual frameworks, to formulate mathematical claims or assist in proving them, or to interpret results, and the remaining required disclosure tasks (translation, dataset cleaning and reformatting, qualitative data analysis) are not applicable to this work. Additionally, we used generative AI tools to edit the manuscript for grammar and readability. We have reviewed all AI-assisted work. Experimental designs suggested by AI tools were evaluated and finalized by the authors, all AI-generated code was reviewed line by line by one of the authors and tested for correctness before use, and all edited text was compared with the original draft to confirm that no technical content was altered. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

All experiments in this work use the publicly available ImageNet dataset, captions from MS-COCO, and publicly available pretrained models (DiT-XL/2, EDM, EDM2-S, EDM2-XXL, FLUX.1, and Stable Diffusion 3.5 Medium), in accordance with their respective licenses and terms of use. The study uses existing resources and involves no new data collection or recruitment of human participants. Our contribution concerns the analysis of saturation behavior in the sampling process of existing diffusion models and introduces no new capabilities beyond those of the underlying models. We are aware that generative models carry general risks of misuse, such as producing misleading content, and that large-scale text-to-image models may reflect biases present in their web-scale training data. We see no ethical concerns specific to this work beyond these general risks.

REPRODUCIBILITY STATEMENT

All experiments use the publicly available ImageNet dataset, captions from MS-COCO, and publicly available pretrained models (DiT-XL/2, EDM, EDM2-S, EDM2-XXL, FLUX.1, and Stable Diffusion 3.5 Medium). Experimental settings for the main method evaluation, including model configurations, sampler settings, and hyperparameters, are described in Appendix. Source code for the method described in Section 5 is included in the anonymized supplementary material.

REFERENCES

Donghoon Ahn, Jiwon Kang, Sanghyun Lee, Jaewon Min, Minjae Kim, Wooseok Jang, Hyoungwon Cho, Sayak Paul, SeonHwa Kim, Eunju Cha, Kyong Hwan Jin, and Seungryong Kim. A noise is worth diffusion guidance. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=xEWooSOgaz.

Bilal Alsallakh, Pamela Bhattacharya, Vanessa Feng, Narine Kokhlikyan, Orion Reblitz-Richardson, Rahul Rajan, and David Yan. A tour of visualization techniques for computer vision datasets. *ArXiv*, abs/2204.08601, 2022. URL https://api.semanticscholar.org/CorpusID:248240058.

Lorenzo Bardone, Claudia Merger, and Sebastian Goldt. A theory of learning data statistics in diffusion models, from easy to hard. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=xPdpfcJ65T.

Hyungjin Chung, Jeongsol Kim, Geon Yeong Park, Hyelin Nam, and Jong Chul YE. CFG++: Manifold-constrained classifier free guidance for diffusion models. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 30824–30850, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/4c9477b9e2c7ec0ad3f4f15077aaf85a-Paper-Conference.pdf.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. IEEE, 2009.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12606–12633. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/esser24a.html.

Martin Nicolas Everaert, Athanasios Fitsios, Marco Bocchio, Sami Arpa, Sabine Süsstrunk, and Radhakrishna Achanta. Exploiting the signal-leak bias in diffusion models. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 4025–4034, January 2024.

Luca Eyring, Shyamgopal Karthik, Karsten Roth, Alexey Dosovitskiy, and Zeynep Akata. ReNO: Enhancing one-step text-to-image models through reward-based noise optimization. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 125487–125519. Curran Associates, Inc., 2024. doi: 10.52202/079017-3987. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/e31bdea0a93741c2157eea705dd219eb-Paper-Conference.pdf.

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

Xiefan Guo, Jinlin Liu, Miaomiao Cui, Jiankai Li, Hongyu Yang, and Di Huang. Initno: Boosting text-to-image diffusion models via initial noise optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9380–9389, June 2024.

Peter J B Hancock, Roland J Baddeley, and Leslie S Smith. The principal components of natural images. *Network: Computation in Neural Systems*, 3(1):61–70, 1992. doi: 10.1088/0954-898X_3_1_008. URL https://doi.org/10.1088/0954-898X_3_1_008.

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. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022. URL https://arxiv.org/abs/2207.12598.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 26565–26577. Curran Associates, Inc., 2022. doi: 10.52202/068431-1926. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/a98846e9d9cc01cfb87eb694d946ce6b-Paper-Conference.pdf.

Tero Karras, Miika Aittala, Jaakko Lehtinen, Janne Hellsten, Timo Aila, and Samuli Laine. Analyzing and improving the training dynamics of diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24174–24184, June 2024.

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/0234c510bc6d908b28c70ff313743079-Paper.pdf.

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 A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 122458–122483. Curran Associates, Inc., 2024. doi: 10.52202/079017-3892. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/dd540e1c8d26687d56d296e64d35949f-Paper-Conference.pdf.

Krunoslav Lehman Pavasovic, Jakob Verbeek, Giulio Biroli, and Marc Mezard. Overshoot and shrinkage in classifier-free guidance: From theory to practice. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 132722–132771, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/d7168ad82b4780797ffa8202c531ef6a-Paper-Conference.pdf.

Xiang Li, Yixiang Dai, and Qing Qu. Understanding generalizability of diffusion models requires rethinking the hidden Gaussian structure. In Amir Globerson, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/69e68611ec4d8c0ae4a4b2bece165f5f-Abstract-Conference.html.

Xiang Li, Rongrong Wang, and Qing Qu. Towards understanding the mechanisms of classifier-free guidance. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 62835–62889. Curran Associates, Inc., 2025. doi: 10.52202/085713-2106. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/5ac55a8d65fb5ecd9ccaa852e21325db-Paper-Conference.pdf.

Shanchuan Lin, Bingchen Liu, Jiashi Li, and Xiao Yang. Common diffusion noise schedules and sample steps are flawed. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5404–5411, January 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 David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (eds.), *Computer Vision – ECCV 2014*, pp. 740–755, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10602-1.

Artem Lukoianov, Chenyang Yuan, Justin Solomon, and Vincent Sitzmann. Locality in image diffusion models emerges from data statistics. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 95121–95157. Curran Associates, Inc., 2025. doi: 10.52202/085713-3180. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/88ad9774ffcb7a272457e9396f793a07-Paper-Conference.pdf.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4195–4205, October 2023.

Emile Pierret and Bruno Galerne. Diffusion models for Gaussian distributions: Exact solutions and Wasserstein errors. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 49355–49381. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/pierret25a.html.

Zipeng Qi, Lichen Bai, Haoyi Xiong, and Zeke Xie. Not all noises are created equally: Diffusion noise selection and optimization. *arXiv preprint arXiv:2407.14041*, 2024.

Seyedmorteza Sadat, Jakob Buhmann, Derek Bradley, Otmar Hilliges, and Romann M. Weber. CADS: Unleashing the diversity of diffusion models through condition-annealed sampling. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=zMoNrajk2X.

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

Seyedmorteza Sadat, Tobias Vontobel, Farnood Salehi, and Romann M. Weber. Guidance in the frequency domain enables high-fidelity sampling at low CFG scales, 2025b. URL https://arxiv.org/abs/2506.19713.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, Jonathan Ho, David Fleet, and Mohammad Norouzi. Photorealistic text-to-image diffusion models with deep language understanding. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 36479–36494. Curran Associates, Inc., 2022. doi: 10.52202/068431-2643. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/ec795aeadae0b7d230fa35cbaf04c041-Paper-Conference.pdf.

Dvir Samuel, Rami Ben-Ari, Simon Raviv, Nir Darshan, and Gal Chechik. Generating images of rare concepts using pre-trained diffusion models. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024. ISBN 978-1-57735-887-9. doi: 10.1609/aaai.v38i5.28270. URL https://doi.org/10.1609/aaai.v38i5.28270.

Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1216, 2001. ISSN 1545-4126. doi: https://doi.org/10.1146/annurev.neuro.24.1.1193. URL https://www.annualreviews.org/content/journals/10.1146/annurev.neuro.24.1.1193.

George Stein, Jesse Cresswell, Rasa Hosseinzadeh, Yi Sui, Brendan Ross, Valentin Villecroze, Zhaoyan Liu, Anthony L Caterini, Eric Taylor, and Gabriel Loaiza-Ganem. Exposing flaws of generative model evaluation metrics and their unfair treatment of diffusion models. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 3732–3784. Curran Associates, Inc., 2023. doi: 10.52202/075280-0165. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/0bc795afae289ed465a65a3b4b1f4eb7-Paper-Conference.pdf.

Binxu Wang and Cengiz Pehlevan. An analytical theory of spectral bias in the learning dynamics of diffusion models. In Danielle Belgrave, Cheng Zhang, Laura N. Montoya, Hsuan-Tien Lin, Razvan Pascanu, Piotr Koniusz, Marzyeh Ghassemi, Nancy Chen, Iván Vladimir Meza Ruíz, and Arturo Loaiza-Bonilla (eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2025, NeurIPS 2025, San Diego, CA, USA, December 2-7, 2025 / Mexico City, Mexico, November 30 - December 5, 2025*, 2025. URL http://papers.nips.cc/paper_files/paper/2025/hash/8a0d3f77bb435817807d463c5dcef1ab-Abstract-Conference.html.

Binxu Wang and John Vastola. The unreasonable effectiveness of Gaussian score approximation for diffusion models and its applications. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=I0uknSHM2j.

Qingsong Wang, Zhengchao Wan, Misha Belkin, and Yusu Wang. Seeds of structure: Patch PCA reveals universal compositional cues in diffusion models. In Danielle Belgrave, Cheng Zhang, Laura N. Montoya, Hsuan-Tien Lin, Razvan Pascanu, Piotr Koniusz, Marzyeh Ghassemi, Nancy

Chen, Iván Vladimir Meza Ruíz, and Arturo Loaiza-Bonilla (eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2025, NeurIPS 2025, San Diego, CA, USA, December 2-7, 2025 / Mexico City, Mexico, November 30 - December 5, 2025*, 2025. URL http://papers.nips.cc/paper_files/paper/2025/hash/a9a053cf205783386db4d5673c4cfcd4-Abstract-Conference.html.

Dana Weitzner, Mauricio Delbracio, Peyman Milanfar, and Raja Giryes. The diffusion process as a correlation machine: Linear denoising insights. *Transactions on Machine Learning Research*, 2025. URL https://openreview.net/forum?id=FGDJOc27rt.

Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *CoRR*, abs/2306.09341, 2023. doi: 10.48550/ARXIV.2306.09341. URL https://doi.org/10.48550/arXiv.2306.09341.

Mengfei Xia, Nan Xue, Yujun Shen, Ran Yi, Tieliang Gong, and Yong-Jin Liu. Rectified diffusion guidance for conditional generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13371–13380, June 2025.

Katherine Xu, Lingzhi Zhang, and Jianbo Shi. Good seed makes a good crop: Discovering secret seeds in text-to-image diffusion models. In *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2025, Tucson, AZ, USA, February 26 - March 6, 2025*, pp. 3024–3034. IEEE, 2025. doi: 10.1109/WACV61041.2025.00299. URL https://doi.org/10.1109/WACV61041.2025.00299.

Kai Zhang, Fujun Luan, Sai Bi, and Jianming Zhang. EP-CFG: Energy-preserving classifier-free guidance, 2024. URL https://arxiv.org/abs/2412.09966.

Candi Zheng and Yuan Lan. Characteristic guidance: Non-linear correction for diffusion model at large guidance scale. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 61386–61412. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/zheng24f.html.

Zikai Zhou, Shitong Shao, Lichen Bai, Shufei Zhang, Zhiqiang Xu, Bo Han, and Zeke Xie. Golden noise for diffusion models: A learning framework. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17688–17697. IEEE, 2025.

A Experimental Details for Intensity Saturation

Estimation of image statistics. The limited number of samples per ImageNet class constrains the accuracy of dataset-based estimates of the mean and covariance matrix. We therefore generated 50,000 images per class and used them to estimate the class-conditional mean and covariance. We also estimated the unconditional mean and covariance from 50,000 unconditionally generated images. All images were generated using EDM models, with `edm-imagenet-64x64-cond-adm.pkl` for conditional generation and `network-snapshot-048000.pkl` for unconditional generation. We used the Heun sampler with 32 sampling steps and seeds from 0 to 49,999 for each class and for unconditional generation.

Linear prediction error. For Fig. 1(b), we evaluate the normalized root mean squared error (NRMSE) of the final coefficients predicted by the linear denoiser. For each class c, let

$$\tilde{a}_i^{(n)} = u_i^\top \left(x_{\text{gen}}^{(n)} - \mu_c\right) \tag{9}$$

denote the coefficient of generated sample n along the class covariance eigenvector u_i, where μ_c is the class mean. Let $a_i^{(n)}$ denote the corresponding predicted coefficient obtained from Eq. 2 at $\sigma_t = 0$ using the same initial noise. We compute

$$\text{NRMSE}_i = \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(a_i^{(n)} - \tilde{a}_i^{(n)}\right)^2}}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\tilde{a}_i^{(n)}\right)^2}}. \tag{10}$$

The denominator is the root mean square of the observed coefficients. We use $N = 10{,}000$ samples per class across 50 classes and average the per-class NRMSE values at each eigenvector rank.

Prediction of saturated regions. To obtain the predicted saturated regions in Fig. 1(c), we evaluated Eq. 2 at $\sigma_t = 0$, using the initial noise sample used for image generation as x_T. For the resulting prediction, we classified a pixel as saturated if all three RGB channel values were greater than 0.95 or all three were less than -0.95, on the $[-1, 1]$ scale. The predicted saturated area was defined as the fraction of pixels satisfying this criterion. The same procedure was used for Table 1.

Saturated area and initial noise projections. We define the centered initial noise projection as $z_i = u_i^\top (x_T - \mu_c)$, with Z_i denoting the corresponding random variable. In Fig. 1(d), the solid red curve shows $\mathbb{E}[\text{Saturated area} \mid Z_i = z]$. The dashed red line shows $\mathbb{E}[\text{Saturated area}]$.

B Eigenvalue-Dependent Linear Predictability of Coefficients along Covariance Eigenvectors

This appendix provides the conditions, derivation, and empirical evidence for the eigenvalue-dependent ordering of the linear predictability of the final coefficients used in Section 3.1. All ordering statements compare quantities associated with covariance eigenvectors within a fixed class. The level of linear predictability can vary across classes.

B.1 Definitions

For class c, let $u_{c,i}$ denote an eigenvector of the class-conditional covariance matrix Σ_c and let $\lambda_{c,i}$ denote its corresponding eigenvalue. We define the initial noise projection and the corresponding coefficient along the sampling trajectory as

$$q_{c,i} = u_{c,i}^\top (x_T - \mu_c), \qquad a_{c,i}(\tau) = u_{c,i}^\top (x(\tau) - \mu_c), \tag{11}$$

where $\tau = \log(\sigma_T/\sigma)$ increases during reverse sampling. We define the linear predictability of the trajectory coefficient from the corresponding initial noise projection as the squared correlation across initial noise samples,

$$R_{c,i}^2(\tau) = \text{Corr}^2[q_{c,i}, a_{c,i}(\tau)]\,. \tag{12}$$

Its final value is denoted by $R^2_{c,i,f}$.

The population least squares projection of $a_{c,i}(\tau)$ on $q_{c,i}$ gives

$$a_{c,i}(\tau) - \mathbb{E}[a_{c,i}(\tau)] = \beta_{c,i}(\tau)\,(q_{c,i} - \mathbb{E}[q_{c,i}]) + \eta_{c,i}(\tau), \tag{13}$$

$$\mathrm{Cov}[q_{c,i}, \eta_{c,i}(\tau)] = 0. \tag{14}$$

Consequently,

$$R^2_{c,i}(\tau) = \frac{\beta^2_{c,i}(\tau)\,\mathrm{Var}(q_{c,i})}{\beta^2_{c,i}(\tau)\,\mathrm{Var}(q_{c,i}) + \mathrm{Var}(\eta_{c,i}(\tau))}. \tag{15}$$

Thus, $R^2_{c,i}(\tau)$ measures the fraction of coefficient variance associated linearly with the corresponding initial noise projection.

We define the decorrelation loss, its final amplitude, and its normalized progress as

$$D_{c,i}(\tau) = 1 - R^2_{c,i}(\tau), \qquad L_{c,i} = 1 - R^2_{c,i,f}, \qquad P_{c,i}(\tau) = \frac{D_{c,i}(\tau)}{L_{c,i}}. \tag{16}$$

These definitions give the amplitude and timing decomposition

$$R^2_{c,i}(\tau) = 1 - L_{c,i} P_{c,i}(\tau). \tag{17}$$

For $p \in (0, 1)$, the transition location is

$$\tau^{(p)}_{c,i} = \inf\left\{\tau \mid P_{c,i}(\tau) \geq p\right\}. \tag{18}$$

B.2 Eigenvalue Dependent Transition Timing

Along eigenvector $u_{c,i}$, the data variance is $\lambda_{c,i}$ and the diffusion noise variance is σ^2. The corresponding signal to noise ratio is

$$\mathrm{SNR}_{c,i}(\sigma) = \frac{\lambda_{c,i}}{\sigma^2}. \tag{19}$$

Suppose that data dependent nonlinear dynamics become active near an effective threshold χ. The predicted transition scale then satisfies

$$\frac{\lambda_{c,i}}{\sigma^2_{c,i,\mathrm{tr}}} = \chi, \qquad \log \sigma_{c,i,\mathrm{tr}} = \frac{1}{2}\log \lambda_{c,i} - \frac{1}{2}\log \chi. \tag{20}$$

This relation motivates the alignment coordinate

$$z_{c,i} = \log \sigma - \frac{1}{2}\log \lambda_{c,i}. \tag{21}$$

A larger eigenvalue corresponds to a transition at a larger noise level and an earlier position in reverse sampling. This argument establishes a timing prediction. The endpoint ordering follows from the additional conditions in the next subsection.

We evaluate the timing prediction using 50 classes, the 10 leading covariance eigenvectors of each class, 1024 seeds (100000 through 101023) per class, and 33 saved sampling states, with conditional sampling at $w = 1$. Five fold class held out evaluation measures normalized decorrelation progress over the active transition region $0.05 \leq P_{c,i} \leq 0.95$; raw R^2 is evaluated over all saved positive-noise states.

Raw R^2 combines transition timing with final amplitude that varies across eigenvectors. Its class held out Q^2 is 0.6004 on $\log \sigma$ and 0.5597 under the one half eigenvalue shift. The optimal alignment coefficient for raw R^2 is 0.000. Figure 8 shows the remaining final amplitude variation after the prespecified horizontal shift.

The signal to noise coordinate substantially aligns normalized decorrelation timing across classes and eigenvectors. The fitted coefficient 0.400 provides the best held out alignment in this experiment. Figures 9 and 10 show the aligned progress curves and the alignment coefficient scan. The transition crossing regression

$$\log \sigma^{(p)}_{\mathrm{cross}} = b_p + \gamma_p \log \lambda_{c,i} \tag{22}$$

gives consistent positive exponents.

Figure 11 shows the transition crossing regressions at the three progress levels.

The normalized timing relation is approximately shared across classes. Final amplitudes $L_{c,i}$ vary across classes and eigenvectors.

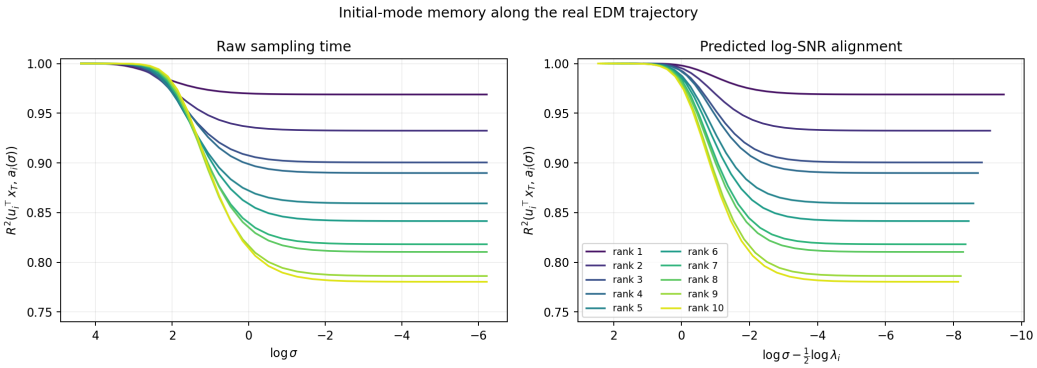


Figure 8: Class-averaged R^2 curves for each eigenvalue rank, shown against raw log noise level and the prespecified signal-to-noise alignment coordinate. Final amplitude variation remains visible across eigenvectors.

Table 3: Held-out prediction of normalized decorrelation progress using three alignment coordinates. Metrics are evaluated over the active transition region across five class folds.

Alignment coordinate	Held out RMSE	Held out Q^2
$\log \sigma$	0.1128	0.8602
$\log \sigma - \frac{1}{2} \log \lambda_{c,i}$	0.0373	0.9847
$\log \sigma - 0.400 \log \lambda_{c,i}$	0.0248	0.9933

B.3 Class-Conditional Ordering of the Linear Predictability of Final Coefficients

We transform decorrelation into the odds

$$H_{c,i}(\tau) = \frac{1 - R^2_{c,i}(\tau)}{R^2_{c,i}(\tau)}. \tag{23}$$

The following conditions connect transition timing to the final value of R^2.

Transition ordering For covariance eigenvectors within a fixed class,

$$\lambda_{c,j} \geq \lambda_{c,i} \quad \Longrightarrow \quad \tau_{c,j} \leq \tau_{c,i}. \tag{24}$$

Separable generation of decorrelation odds For a nonnegative transition kernel k_c and an effective envelope ℓ_c,

$$\frac{dH_{c,i}}{d\tau} = \ell_c(\tau) k_c(\tau - \tau_{c,i}). \tag{25}$$

The functions may vary across classes and are shared across eigenvectors within each class.

Envelope ordering The effective envelope is nondecreasing over the union of the transition supports,

$$\tau_1 \leq \tau_2 \quad \Longrightarrow \quad \ell_c(\tau_1) \leq \ell_c(\tau_2). \tag{26}$$

Table 4: Transition crossing regressions at three levels of normalized decorrelation progress. The fitted exponent γ_p relates the transition noise level to the class covariance eigenvalue.

p	γ_p	95% interval	Fit R^2	Median rank correlation
0.1	0.420	$[0.411, 0.428]$	0.970	0.872
0.5	0.403	$[0.393, 0.413]$	0.960	0.825
0.9	0.371	$[0.358, 0.384]$	0.912	0.535

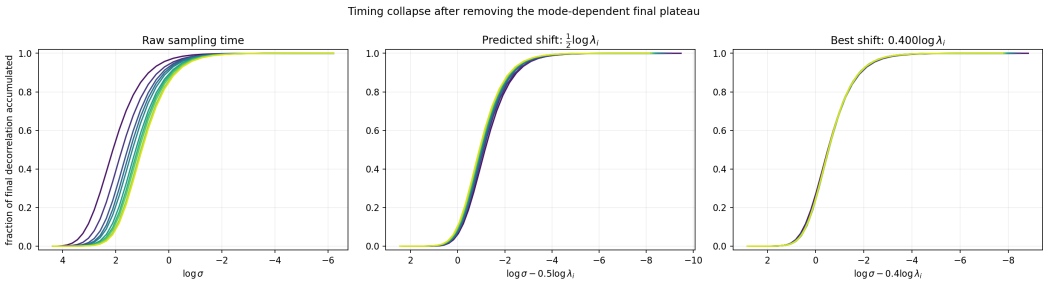


Figure 9: Normalized decorrelation progress in raw sampling time, under the prespecified one half shift, and under the empirical 0.400 shift.

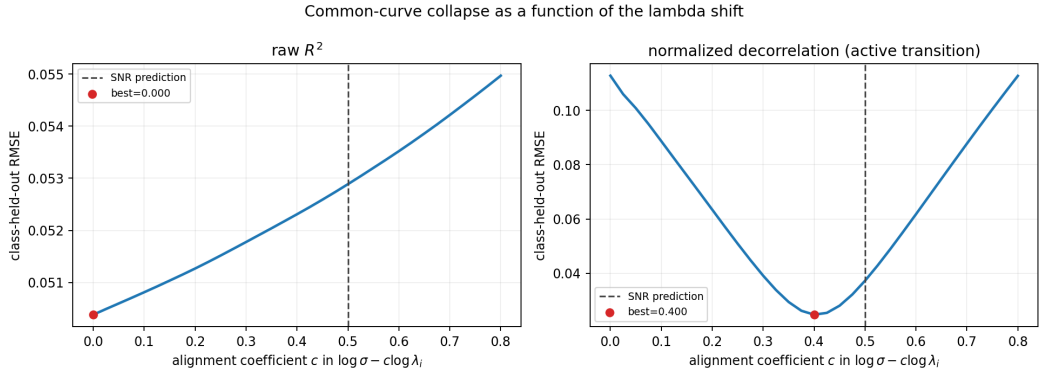


Figure 10: Alignment coefficient scans for raw R^2 over all saved positive-noise states and normalized progress over the active transition region. Normalization separates final amplitude variation from transition timing variation.

Since $H_{c,i}(0) = 0$, integration gives

$$H_{c,i,f} = \int k_c(s)\ell_c(\tau_{c,i} + s)\, ds. \tag{27}$$

For $\lambda_{c,j} \geq \lambda_{c,i}$, transition ordering and envelope ordering yield

$$H_{c,j,f} - H_{c,i,f} = \int k_c(s)\left[\ell_c(\tau_{c,j} + s) - \ell_c(\tau_{c,i} + s)\right] ds \tag{28}$$

$$\leq 0. \tag{29}$$

The map $H \mapsto H/(1+H) = 1 - R^2$ is strictly increasing. Therefore,

$$\lambda_{c,j} \geq \lambda_{c,i} \quad \Longrightarrow \quad R^2_{c,j,f} \geq R^2_{c,i,f}. \tag{30}$$

This is the within class ordering stated in Section 3.1.

A strict ordering follows when

$$\Delta_{c,j,i} = \int k_c(s)\left[\ell_c(\tau_{c,i} + s) - \ell_c(\tau_{c,j} + s)\right] ds > 0. \tag{31}$$

This condition gives $H_{c,j,f} < H_{c,i,f}$ and $R^2_{c,j,f} > R^2_{c,i,f}$. Equality remains possible when the effective envelope is constant over the kernel weighted comparison region.

Transition timing alone leaves the final amplitude unconstrained. For example, $D_{c,i}(\tau) = L_{c,i}F(\tau - \tau_{c,i})$ permits independent choices of $\tau_{c,i}$ and $L_{c,i}$. The separable generation and envelope ordering conditions provide the link between timing and the linear predictability of the final coefficients.

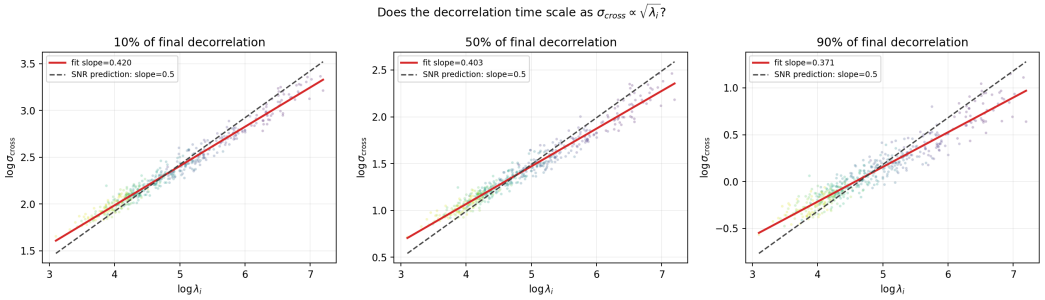


Figure 11: Transition noise levels as a function of the class covariance eigenvalue at 10%, 50%, and 90% of final decorrelation.

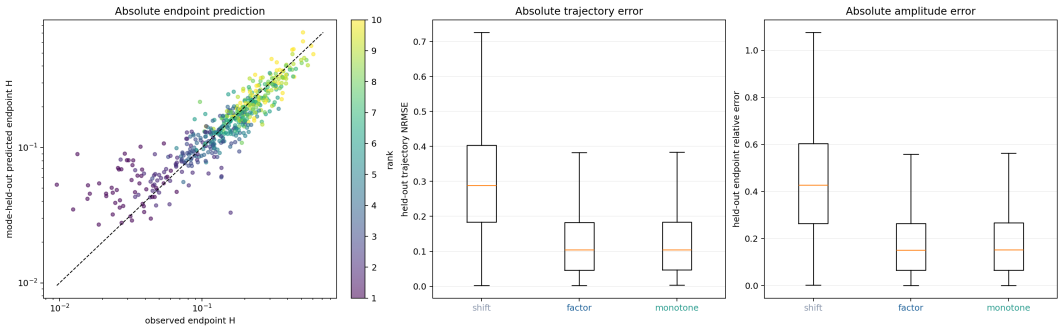


Figure 12: Held out endpoint prediction together with trajectory and amplitude errors for the shifted kernel, absolute time factorization, and monotone envelope factorization.

B.4 Empirical Validation of the Ordering Conditions

Separable odds generation For each class, we fit independent ℓ_c and k_c functions using the H trajectories associated with nine eigenvectors. We evaluate the predicted H trajectory and final value for the held-out tenth eigenvector. Each of the 10 leading eigenvectors is held out in turn in each of the 50 classes.

Table 5: Held-out prediction of decorrelation odds trajectories and endpoints for three factorization models across 50 classes. Each of the 10 leading eigenvectors is held out in turn within each class.

Metric	Shifted kernel	Absolute time factorization	Monotone envelope factorization
Trajectory NRMSE	0.288	0.105	0.104
Endpoint error	0.427	0.151	0.151
Endpoint correlation	0.218	0.923	0.924
Median class correlation	−0.934	0.908	0.910

The absolute time factorization predicts the H trajectory for the held-out eigenvector substantially more accurately than the shifted kernel model. Its median endpoint relative error is 0.151, which identifies the factorization as an accurate approximation with residual amplitude variation across eigenvectors. Figure 12 shows the endpoint predictions and the corresponding trajectory and amplitude errors.

Envelope ordering The fitted envelope has a positive slope in every class. The median slope is 1.428 with a class bootstrap 95% interval of $[1.342, 1.507]$. The median Spearman association between τ and $\log \ell_c$ is 1.000, and 96.0% of classes are locally nondecreasing at every sampled support point. Enforcing a monotone envelope changes the median trajectory NRMSE from 0.1049

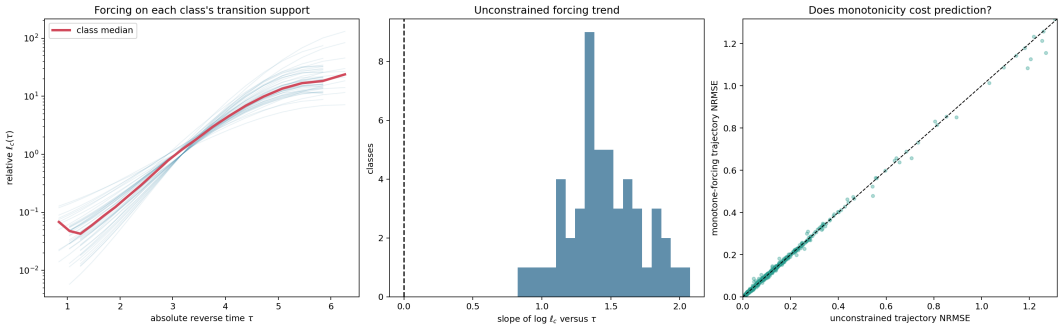


Figure 13: Effective envelope estimates over observed transition supports, their slope distribution, and the prediction errors obtained under the monotone envelope constraint.

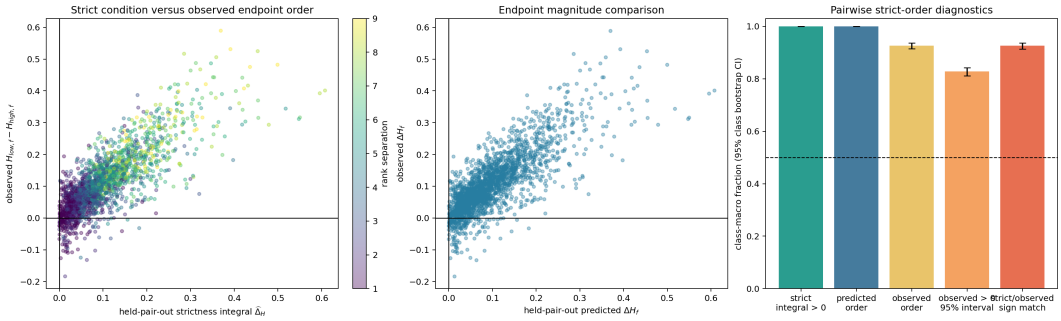


Figure 14: Pairwise strictness integrals, predicted final differences, observed final differences, and class bootstrap ordering frequencies.

to 0.1045. Figure 13 shows the fitted envelopes, their slope distribution, and the predictive effect of enforcing monotonicity.

Strict ordering For each class, we fit the factorization using the H trajectories associated with eight eigenvectors and evaluate the predicted final ordering on the remaining pair of eigenvectors. Among the 2250 eigenvector pairs, 97.1% satisfy the eigenvalue-ordered timing relation. Within this timing-ordered set of eligible pairs, every fitted strictness integral is positive. Observed strict final ordering holds for 92.5% of eligible pairs, and 82.7% have an unadjusted paired-seed interval that is strictly positive.

Across 2184 timing ordered pairs, the fitted strictness integral correlates with the observed final difference in H at 0.803, with a class bootstrap 95% interval of $[0.764, 0.837]$. The linear predictability of the final coefficients follows the strict ordering in 77.5% of eligible pairs with adjacent eigenvalue ranks. Their smaller timing separation produces a smaller predicted effect. Significant reversals occur in 2.4% of eligible pairs. These results support strict ordering as a strong population tendency. Figure 14 summarizes the pairwise prediction and the observed final ordering.

Dependence mediated by transition timing After removing class fixed effects, the slope of $\log H_{c,i,f}$ with respect to $\log \lambda_{c,i}$ is -0.810. The held out factorization residual has a slope of -0.161 with a class bootstrap 95% interval of $[-0.181, -0.141]$. The timing and envelope factorization accounts for 80.1% of the total slope. Figure 15 shows the total and residual associations and their descriptive decomposition.

B.5 Residual Association with Local Denoiser Nonlinearity

We measure local affine superposition error along each covariance eigenvector as

$$\nu_{c,i}(\tau) = \log\big(1 - \mathrm{Corr}^2\big[u_{c,i}^\top D(\alpha x_1 + \beta x_2), u_{c,i}^\top \{\alpha D(x_1) + \beta D(x_2)\}\big]\big)\,. \tag{32}$$

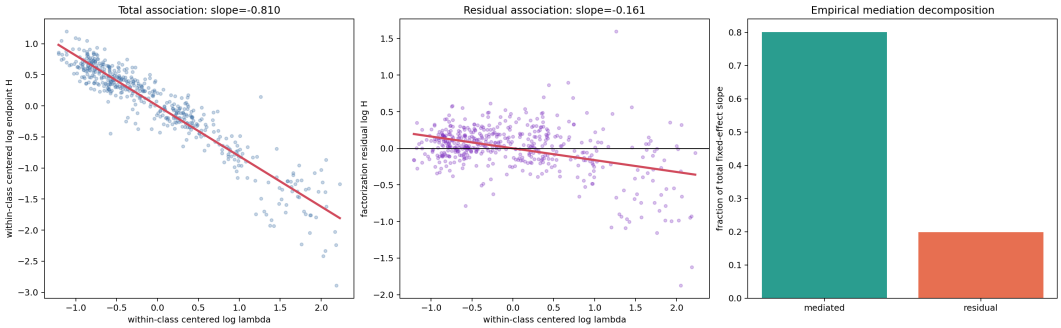


Figure 15: Total and residual associations between eigenvalues and final decorrelation odds after the timing and envelope factorization.

We add $b_c \nu_{c,i}(\tau)$ to the class specific model for $\log(dH_{c,i}/d\tau)$. Measurements associated with the target eigenvector are excluded from fitting. Measurements of H and ν use disjoint halves of the 1024 seeds with a reversed split for replication.

Table 6: Held-out prediction errors and residual eigenvalue associations for the timing and envelope model and its extension with local affine superposition error.

Model	Trajectory NRMSE	Endpoint error	Residual slope	Explained slope
Timing and envelope	0.111	0.161	-0.163	79.9%
With local affine error	0.093	0.134	-0.042	94.8%

The local affine error accounts for 74.3% of the residual left by the timing and envelope model, with a class bootstrap 95% interval of $[70.6\%, 78.1\%]$. All 1000 held out coefficients have the predicted positive sign. The association is concentrated in the first 16 EDM intervals, whose left-endpoint noise levels range from approximately 80 to 2.90. The early interval model reaches an endpoint error of 0.142 and a trajectory NRMSE of 0.100. Figure 16 shows the residual associations, predictive errors, and descriptive slope decomposition.

The augmented residual slope remains -0.042 with a 95% interval of $[-0.050, -0.034]$. Approximately 5.2% of the total eigenvalue dependent slope remains unexplained. These results establish a predictive association between local denoiser nonlinearity and the remaining decorrelation. Causal identification requires a direct intervention on the local dynamics.

B.6 Independent Consistency Check of Final Coefficient Linearity

An independent experiment evaluates the same 50 classes, the 50 leading covariance eigenvectors, and seeds from 0 through 9999. Every class has a negative within class slope of final decorrelation loss against $\log \lambda$. The median per class linear fit has $R^2 = 0.9368$.

Table 7: Mean covariance eigenvalues, final coefficient linearity, and final decorrelation loss at selected eigenvector ranks across 50 classes, evaluated using seeds 0 through 9999.

Rank	Mean λ	Mean final R^2	Mean decorrelation loss
1	713.93	0.9689	0.0311
2	317.46	0.9343	0.0657
3	195.61	0.9047	0.0953
5	119.78	0.8596	0.1404
10	48.13	0.7797	0.2203

Figure 17 shows the class specific final loss curves and their pooled summary across covariance eigenvalues.

This independent result confirms that final coefficients along leading covariance eigenvectors have the highest linear predictability from the corresponding initial noise projections. The complete analysis

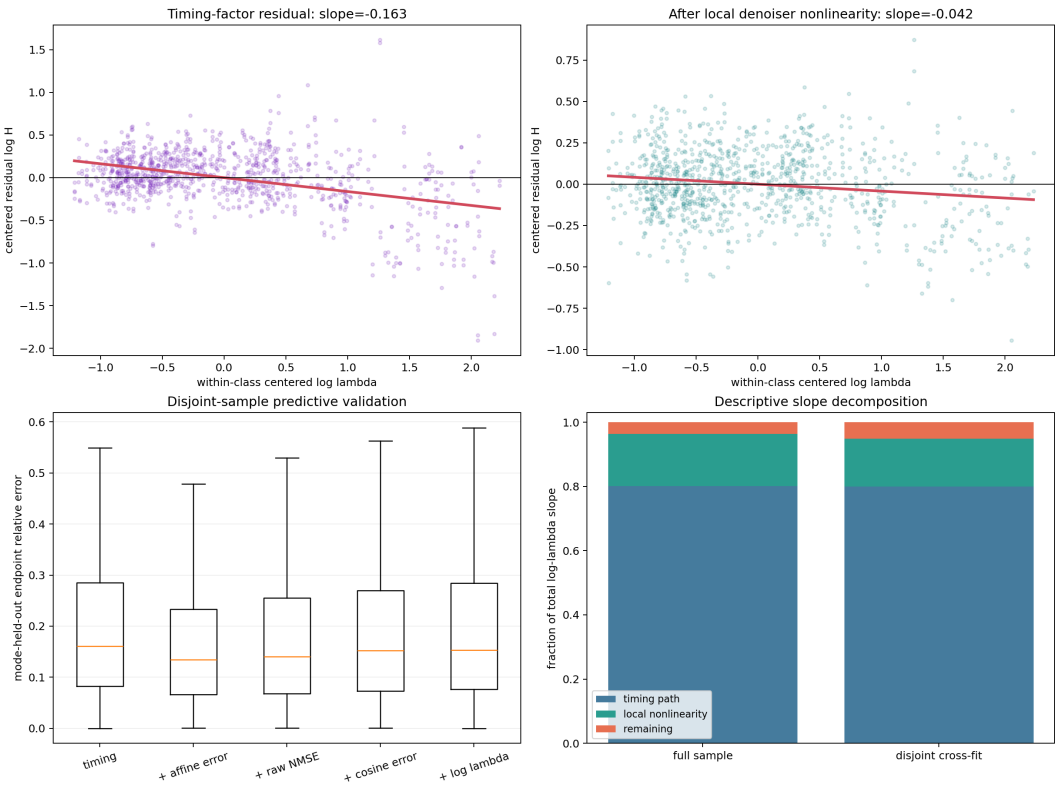


Figure 16: Residual eigenvalue association and predictive error before and after including local affine superposition error.

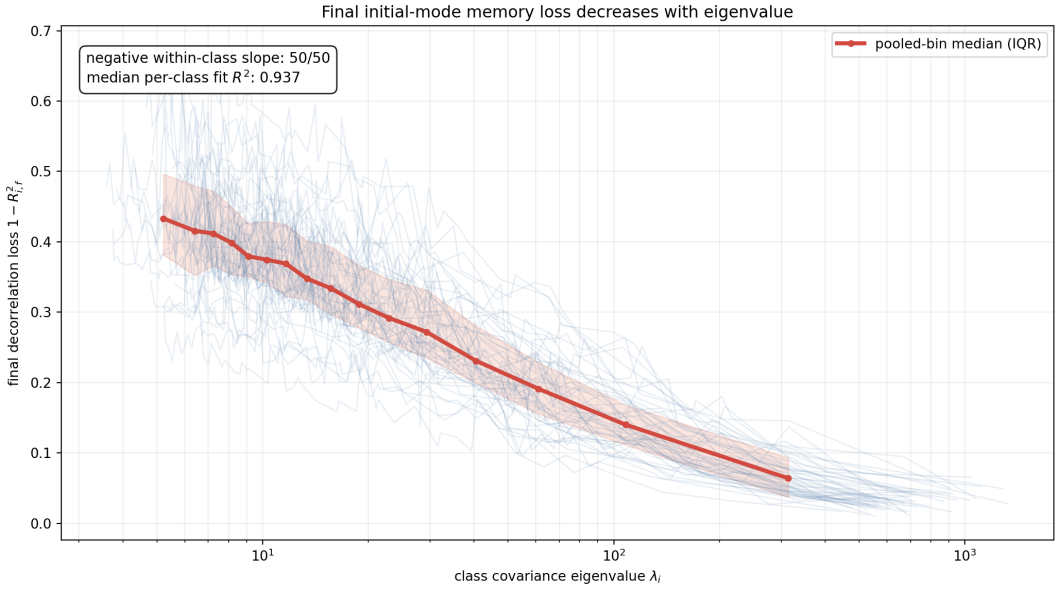


Figure 17: Final decorrelation loss across covariance eigenvalues for 50 classes and the corresponding pooled summary.

supports an approximate timing law across classes and a conditional final ordering within each class. The factorization and local nonlinearity analyses quantify the scope and residual error of this ordering.

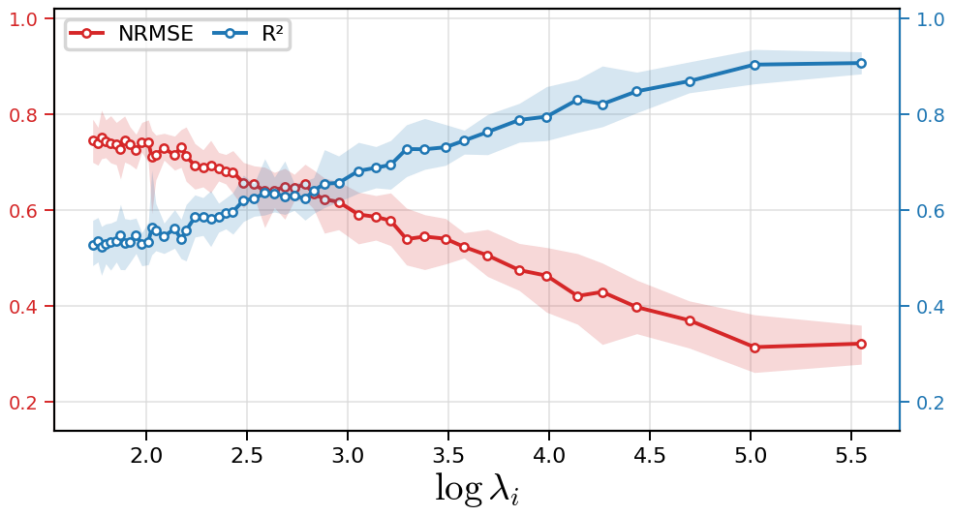


Figure 18: **Initial-to-final projection linearity in EDM2-S latent space.** Results use 5,000 generated samples per class across 10 classes. Squared correlation R^2 (blue, right axis) and linear prediction NRMSE (red, left axis) are plotted against the log covariance eigenvalue. Leading covariance eigenvectors exhibit high R^2 and low NRMSE, extending the trend in Figure 1(b) to latent space.

C Initial-to-Final Projection Linearity in EDM2 Latent Space

We extend the analysis in Figure 1(b) to the latent space of EDM2-S. We generate 5,000 images for each of 10 classes with class indices $0, 207, 31, 430, 483, 504, 64, 817, 933, 947$, yielding 50,000 samples in total. Using the corresponding initial noise and final latent representations, we evaluate the squared correlation R^2 between their projections onto covariance eigenvectors, as defined in Equation 3, and the NRMSE of the linear denoiser predictions.

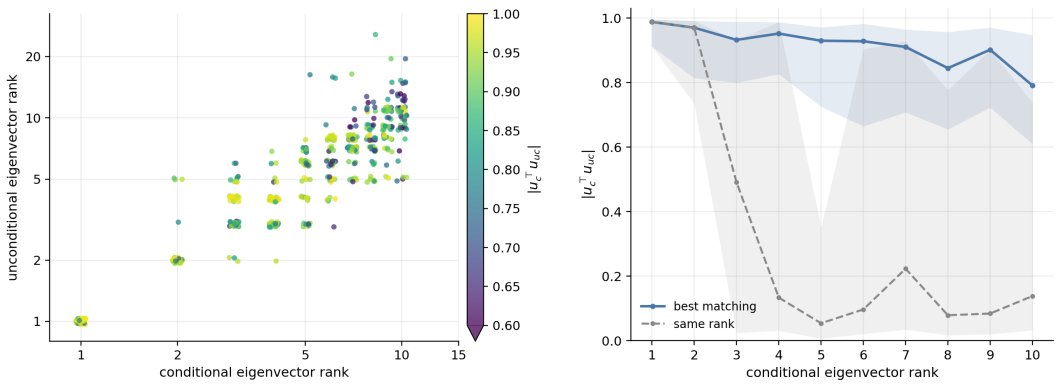
Figure 18 shows that eigenvectors with larger eigenvalues generally exhibit higher R^2 and lower NRMSE. The leading eigenvectors retain strong initial-to-final projection linearity in latent space, consistent with the trend in Figure 1(b). Compared with the leading directions in that figure, R^2 is somewhat lower and the linear prediction error is somewhat higher. These results support the persistence of strong linearity along leading covariance eigenvectors in the evaluated EDM2-S latent-space setting.

D Alignment of Conditional and Unconditional Eigenvectors

We examine the alignment between conditional and unconditional covariance eigenvectors across 50 classes to support the interpretation of leading conditional eigenvectors as brightness directions in Section 3.2. Let $u_{c,i}$ and $u_{uc,j}$ denote unit eigenvectors of the conditional covariance Σ_c and unconditional covariance Σ_{uc}, respectively, with each indexed in decreasing eigenvalue order. We measure their alignment using absolute cosine similarity $|u_{c,i}^\top u_{uc,j}|$, which accounts for the arbitrary sign of each eigenvector. For each conditional eigenvector, we identify the unconditional eigenvector with the highest similarity. We also compare eigenvectors at the same eigenvalue rank using $|u_{c,i}^\top u_{uc,i}|$.

Figure 19(a) shows particularly strong alignment for the top two eigenvectors. In all 50 classes, the top conditional eigenvector has its highest similarity with the top unconditional eigenvector, with a mean similarity of 0.966 and a median of 0.988 across classes. For the second conditional eigenvector, the highest similarity occurs at the second unconditional eigenvector in 92% of classes. The highest similarity has a mean of 0.939 and a median of 0.970 across classes.

The correspondence in eigenvalue rank decreases substantially from the third eigenvector onward. For the third conditional eigenvector, the unconditional eigenvector with the highest similarity has the same rank in 46% of classes. Across all 50 classes, its highest similarity occurs within the top six unconditional eigenvectors, with a mean similarity of 0.904 and a median of 0.933 across classes.



(a) Eigenvector Rank Correspondence (b) Eigenvector Similarity

Figure 19: **Alignment of conditional and unconditional covariance eigenvectors across 50 classes.** (a) Each point pairs a conditional eigenvector rank with the rank of the unconditional eigenvector having the highest absolute cosine similarity. Color indicates this similarity. (b) Median absolute cosine similarity across classes as a function of conditional eigenvector rank. The solid curve uses the unconditional eigenvector with the highest similarity for each class and rank. The dashed curve compares conditional and unconditional eigenvectors at the same eigenvalue rank.

The additional leading conditional eigenvectors shown in Figure 19(b) also retain high similarity with leading unconditional eigenvectors across different ranks.

These observations support the interpretation that leading conditional eigenvectors represent the broad brightness variations associated with leading unconditional eigenvectors in Section 3.2. The top two eigenvectors exhibit especially strong directional alignment and correspondence in eigenvalue rank. Subsequent leading conditional eigenvectors retain substantial directional alignment with leading unconditional eigenvectors across different ranks.

E Brightness Variations along Leading Covariance Eigenvectors

This appendix visualizes the image variations associated with the five leading covariance eigenvectors for EDM, DiT-XL/2, EDM2-S, EDM2-XXL, FLUX, and Stable Diffusion 3.5. The visualizations complement Section 3 by showing how changes in individual eigenvector coefficients affect global brightness and brightness contrasts between large image regions, including for models operating in latent space.

Visualization procedure. For each model, let μ denote the mean of the representation used to compute the covariance, and let (λ_i, u_i) denote its eigenpairs, with $\lambda_1 \geq \lambda_2 \geq \cdots$ and $\|u_i\|_2 = 1$. The eigenvectors $u_1, u_2, \ldots, u_5$ are the first five principal component directions, ordered by decreasing variance. We visualize

$$v_i(\alpha) = \mu + \alpha\sqrt{\lambda_i}\, u_i, \qquad I_i(\alpha) = \mathcal{V}(v_i(\alpha))\,, \qquad \alpha \in \{-2, -1, 0, 1, 2\}, \tag{33}$$

for $i = 1, \ldots, 5$. Here $\sqrt{\lambda_i}$ is the standard deviation of the coefficient along u_i. The map $\mathcal{V}$ converts the representation into an image. Pixel-space representations directly specify pixel values. Latent-space representations are decoded using the corresponding image decoder. Each row varies a single coefficient while keeping all other coefficients at their mean values. The center column therefore shows $\mathcal{V}(\mu)$ for every row of a given model. Eigenvector signs are arbitrary, so the direction from dark to bright can be reversed by changing the sign convention.

Observed brightness patterns. Figure 20 shows consistent overall brightness changes along u_1 across all six models. Other leading components capture regional brightness contrasts and broad color variations, supporting the claim in Section 3 that leading latent covariance eigenvectors correspond to broad brightness variations in image space.

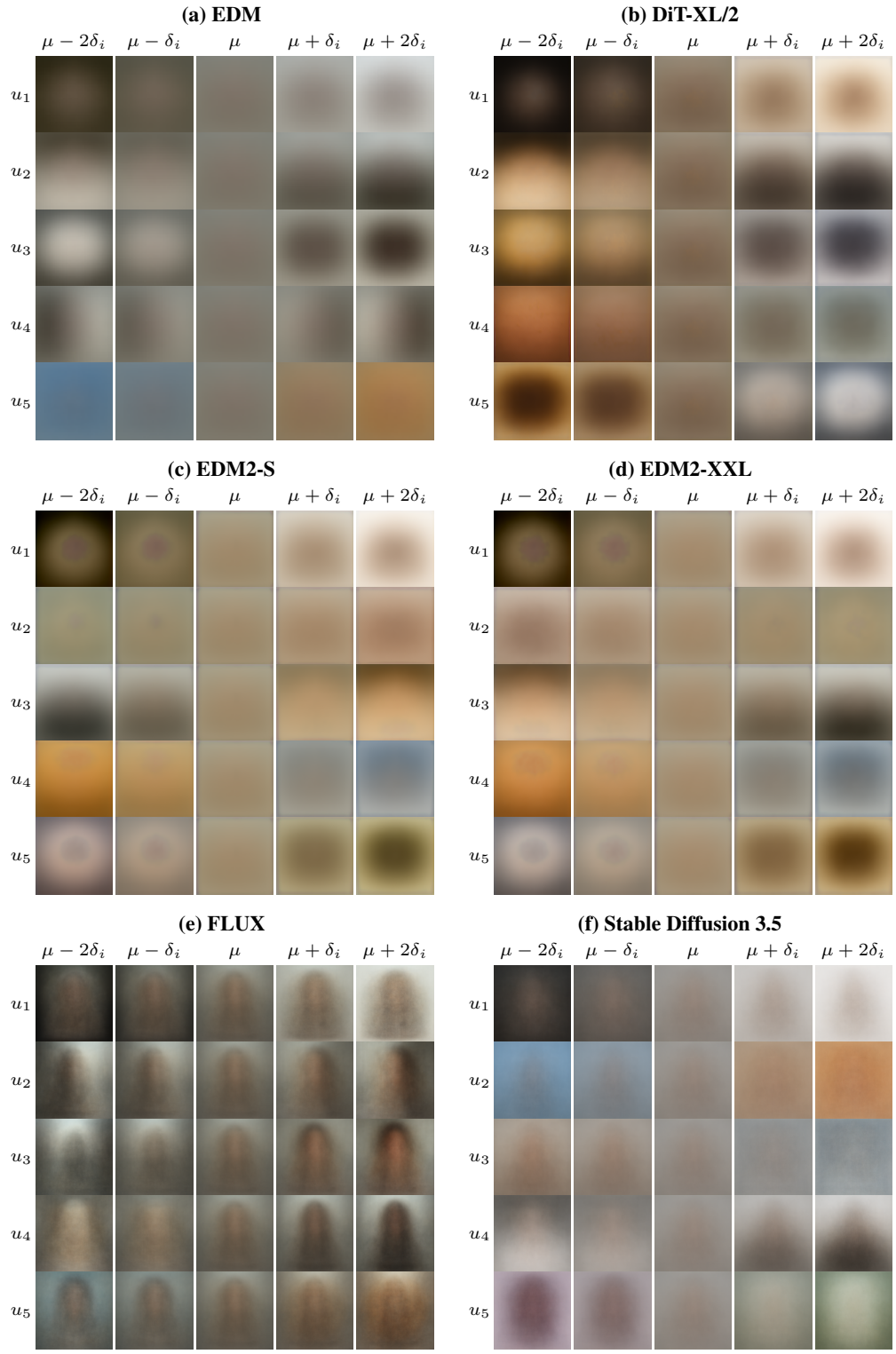


Figure 20: **Image variations along leading covariance eigenvectors across six models.** Panels show (a) EDM, (b) DiT-XL/2, (c) EDM2-S, (d) EDM2-XXL, (e) FLUX, and (f) Stable Diffusion 3.5. Rows correspond to $u_1, \ldots, u_5$ in decreasing eigenvalue order. Columns show $\mathcal{V}(\mu + k\sqrt{\lambda_i}u_i)$ for $k = -2, -1, 0, 1, 2$, where μ is the mean representation. Across models, u_1 captures global brightness variations, while other leading eigenvectors capture regional brightness contrasts and color variations. Eigenvector signs are arbitrary.

F Additional Examples of Saturation Prediction

Figure 21 provides additional examples of the samples and predicted saturated regions shown in Figure 1(a) and (c).

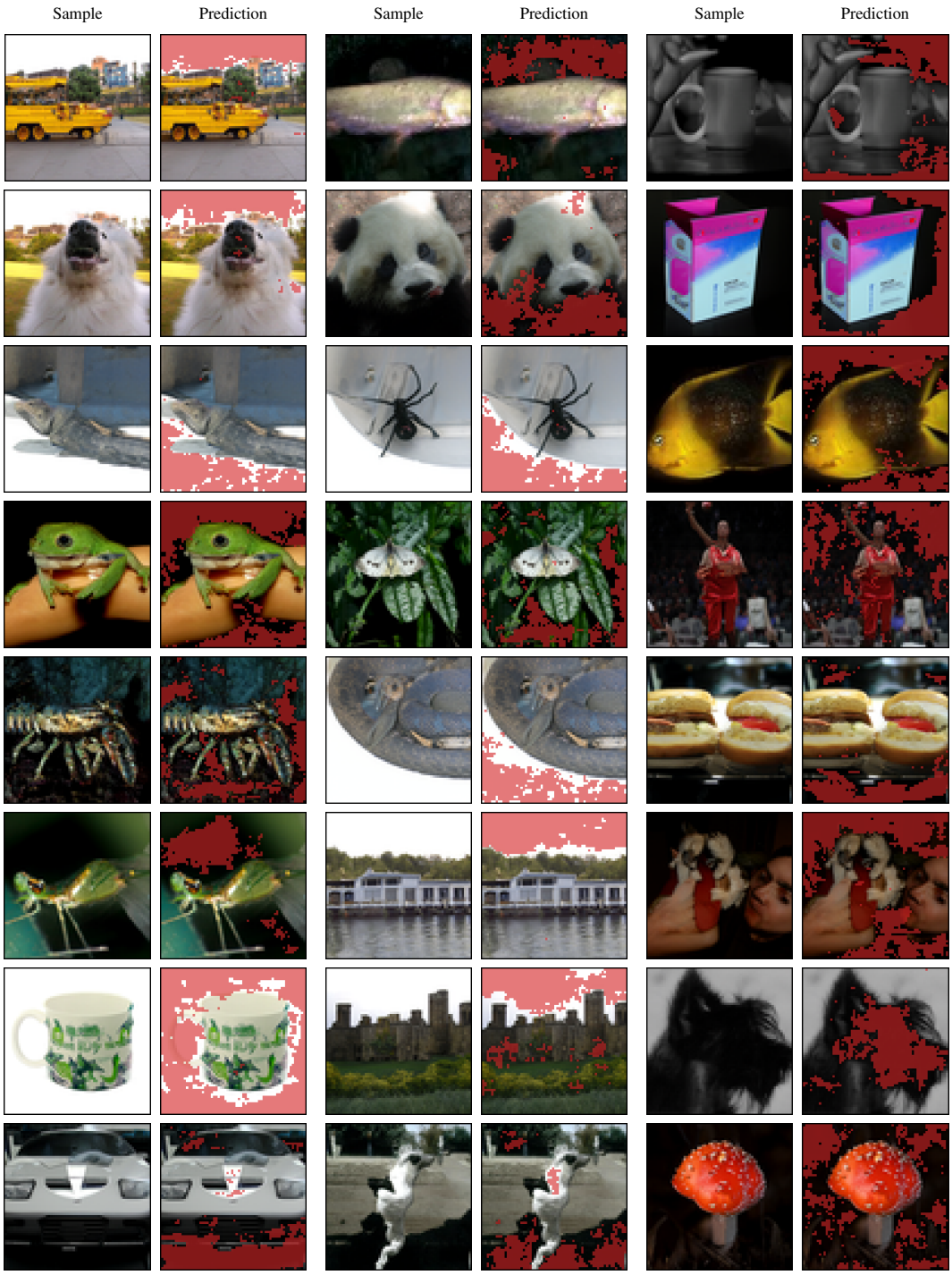


Figure 21: **Additional examples of saturation prediction.** Each pair shows a sample on the left and the same sample with predicted saturated regions overlaid in red on the right. The regions are predicted from the initial noise using the linear denoiser approximation, following Figure 1(c).

G PREDICTING SATURATED REGIONS

We evaluate how accurately the linear denoiser approximation predicts saturated regions from the initial noise. This section provides the experimental setup, the Gaussian baseline, and the evaluation procedures used for Table 1, together with additional results.

Experimental setup. We generated images using the EDM model without guidance from 10,000 initial noise samples for each of 50 classes, yielding 500,000 images in total. Let Ω_i denote the pixel set of generated image i, and let $x_{i,c}(p)$ denote the value of RGB channel c at pixel p, expressed on the $[-1, 1]$ scale. We define the pixel clip indicator as

$$C_i(p) = \mathbf{1}\left\{ \min_{c\in\{R,G,B\}} x_{i,c}(p) > 0.95 \ \vee \ \max_{c\in\{R,G,B\}} x_{i,c}(p) < -0.95 \right\}. \tag{34}$$

Here $\mathbf{1}\{\cdot\}$ equals one when its condition holds and zero otherwise. The saturated region of generated image i is

$$S_i = \{p \in \Omega_i : C_i(p) = 1\}.$$

A pixel belongs to this region when all three normalized RGB values exceed 0.95 or all three fall below -0.95.

Prediction methods. For each initial noise sample, we use Equation (2) to predict the final image under the linear denoiser approximation. Applying the same clip indicator to the predicted image gives $\widehat{C}_i(p)$ and the predicted saturated region $\widehat{S}_i = \{p \in \Omega_i : \widehat{C}_i(p) = 1\}$ for Linear prediction.

For the Random Gaussian baseline, we construct a Gaussian distribution with the mean and covariance of the generated images, which are also used in Linear prediction. For each generated image, we draw a sample from this distribution independently of its initial noise and apply the same intensity thresholds to obtain a predicted saturated region. This baseline measures the precision and recall attainable through independent Gaussian sampling with these matched statistics.

Evaluation metrics. Let S_i denote the observed saturated region of generated image i, and let $\widehat{S}_i$ denote the region predicted by either method. We compute precision and recall for each sample as

$$\text{Precision}_i = \frac{|S_i \cap \widehat{S}_i|}{|\widehat{S}_i|}, \qquad \text{Recall}_i = \frac{|S_i \cap \widehat{S}_i|}{|S_i|},$$

where $|\cdot|$ denotes the number of pixels in a set. Table 1 reports the averages of these per sample metrics. We evaluate both the full set of generated samples and the subset whose observed saturated area is at least 10% of the image.

For these averages, precision is undefined when the predicted region is empty ($|\widehat{S}_i| = 0$), and recall is undefined when the observed region is empty ($|S_i| = 0$). We exclude samples with a zero denominator separately for each metric and average over the remaining valid samples. Among all 500,000 samples, precision uses 483,418 valid samples for Random Gaussian and 482,496 for Linear prediction; recall uses 379,515 valid samples for both methods. The subset with an observed saturated area of at least 10% contains 41,825 samples, all of which are used for recall by both methods. In this subset, precision uses 41,022 valid samples for Random Gaussian and 41,791 for Linear prediction.

We additionally compute pooled precision and recall by aggregating pixel counts across all samples in each evaluation set:

$$\text{Precision}_{\text{pooled}} = \frac{\sum_i |S_i \cap \widehat{S}_i|}{\sum_i |\widehat{S}_i|}, \qquad \text{Recall}_{\text{pooled}} = \frac{\sum_i |S_i \cap \widehat{S}_i|}{\sum_i |S_i|}.$$

The pooled metrics include all 500,000 samples in the full evaluation set and all 41,825 samples in the $\geq 10\%$ subset, including samples with empty predicted or observed regions. The averages of per sample metrics assign equal weight to each valid sample for the corresponding metric. Pooled precision weights samples by the number of predicted saturated pixels, while pooled recall weights samples by the number of observed saturated pixels.

Pooled evaluation. Table 8 compares the pooled precision and recall of Linear prediction and Random Gaussian over all generated samples and over samples whose observed saturated area is at least 10% of the image.

Table 8: Pooled precision and recall for predicting saturated regions. The $\geq 10\%$ subset contains samples whose observed saturated area is at least 10% of the image.

Evaluation set	Method	Precision	Recall
All samples	Linear prediction	0.373	0.307
	Random Gaussian	0.044	0.042
$\geq 10\%$ subset	Linear prediction	0.823	0.282
	Random Gaussian	0.280	0.043

Evaluation by saturation type. In Section 4.1, we distinguish two types of intensity saturation: one sided and two sided saturation. To evaluate prediction for each type, we compute the precision and recall of Linear prediction separately for the two groups. Table 9 compares the pooled metrics for the two saturation types. Table 10 compares the averages of per sample metrics. Both tables use the same two evaluation conditions.

Table 9: Pooled precision and recall of Linear prediction for one sided and two sided intensity saturation, as defined in Section 4.1. The $\geq 10\%$ subset contains samples whose observed saturated area is at least 10% of the image.

Evaluation set	Saturation type	Precision	Recall
All samples	One sided	0.443	0.306
	Two sided	0.342	0.313
$\geq 10\%$ subset	One sided	0.844	0.279
	Two sided	0.746	0.295

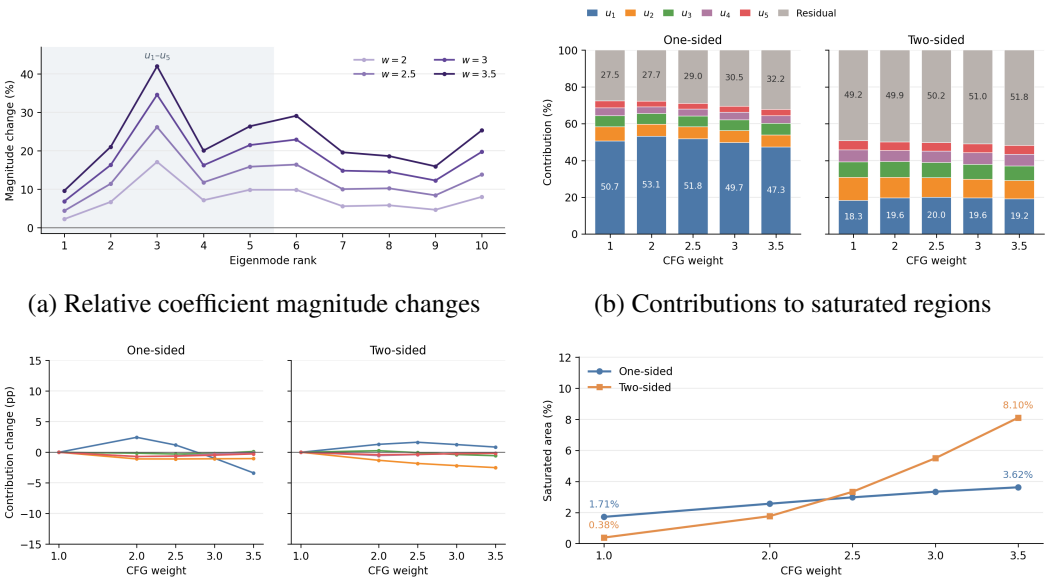
Table 10: Precision and recall of Linear prediction averaged over individual samples for one sided and two sided intensity saturation, as defined in Section 4.1. The $\geq 10\%$ subset contains samples whose observed saturated area is at least 10% of the image.

Evaluation set	Saturation type	Precision	Recall
All samples	One sided	0.285	0.340
	Two sided	0.268	0.302
$\geq 10\%$ subset	One sided	0.818	0.312
	Two sided	0.736	0.320

H Effects of CFG on Eigenvector Coefficients and Saturated Area

We first compare leading covariance eigenvector contributions to intensity saturation under CFG and unguided sampling, then examine how changes in their coefficient magnitudes relate to the increase in two-sided saturation. We evaluate 10,000 generated images per class across 50 classes at each CFG weight $w \in \{1, 2, 2.5, 3, 3.5\}$, with unguided conditional sampling ($w = 1$) as the baseline. We use the definitions of intensity saturation and its one-sided and two-sided types from Figure 5.

To examine which components account for the saturated-region pixel magnitudes under CFG, we repeat the regional decomposition in Figures 5(a) and 23(a) at each guidance weight. Figure 22(b) shows that the contribution profiles remain similar within each saturation type. The top eigenvector u_1 contributes 47.3–53.1% for one-sided saturation and 18.3–20.0% for two-sided saturation. Several leading eigenvectors continue to contribute substantially to two-sided saturation. Figure 22(c)



(a) Relative coefficient magnitude changes

(b) Contributions to saturated regions

(c) Changes in contribution ratios

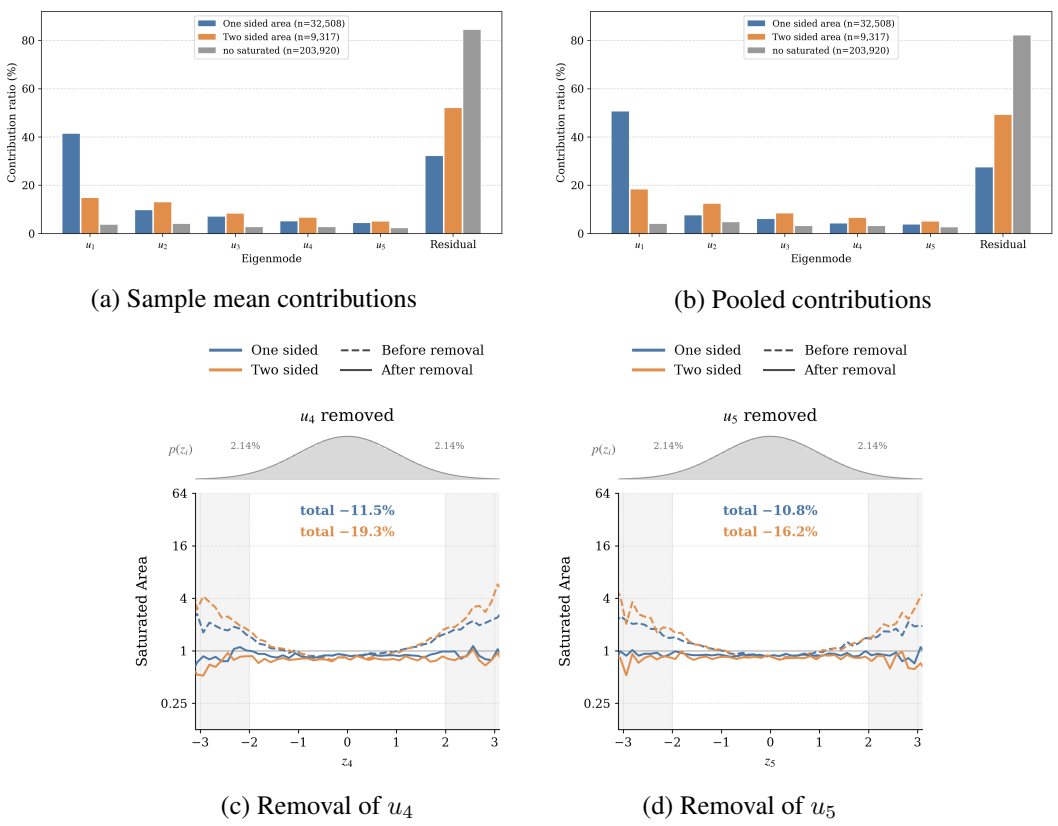
(d) Mean saturated area

Figure 22: **Eigenvector contributions to intensity saturation across CFG weights.** Measurements use 10,000 images per class across 50 classes. (a) Percentage changes in mean coefficient magnitudes relative to unguided conditional sampling ($w = 1$). (b) Contributions of $u_1, \ldots, u_5$ and the residual to saturated-region pixel magnitudes, evaluated separately for one-sided and two-sided saturation at each weight. (c) Changes in the contribution ratios of $u_1, \ldots, u_5$ from $w = 1$, in percentage points, using the colors in (b). (d) Mean saturated area for each saturation type over all generated samples. Larger CFG weights increase mean saturated area while the eigenvector contribution profiles remain similar to those at $w = 1$.

quantifies these changes relative to $w = 1$. Each of the five leading eigenvector contributions changes by less than four percentage points over the evaluated weights. These similar contribution profiles support the interpretation that intensity saturation under CFG remains associated with large coefficients along the leading eigenvectors that contribute to saturation in unguided sampling. The top eigenvector remains dominant in one-sided saturation, and multiple leading eigenvectors contribute substantially to two-sided saturation.

Figure 22(a) shows the relative change in mean coefficient magnitude along leading covariance eigenvectors from the unguided baseline. Mean magnitudes increase across the leading eigenvectors, including u_1, with larger CFG weights producing larger increases. This increase in coefficient magnitudes across multiple leading eigenvectors helps explain the more rapid increase in two-sided saturation, whose saturated-region pixel magnitudes receive substantial contributions from several of these directions.

Figure 22(d) measures the mean saturated area over all generated samples, separately for the two saturation types. From $w = 1$ to $w = 3.5$, this mean increases from 1.71% to 3.62% for one-sided saturation and from 0.38% to 8.10% for two-sided saturation. Figure 4 measures the number of samples whose saturated area is at least 10%. Together, these results show that increasing the CFG weight increases both the frequency of samples with substantial saturation and the mean saturated area across samples, with two-sided saturation increasing more rapidly than one-sided saturation. This pattern is consistent with the increase in coefficient magnitudes across multiple leading eigenvectors and the sustained contributions of these directions to two-sided saturation.



(a) Sample mean contributions

(b) Pooled contributions

(c) Removal of u_4

(d) Removal of u_5

Figure 23: **Additional eigenvector contributions and ablation results.** (a),(b) Sample mean and pooled contribution ratios, respectively, for $u_1, \ldots, u_5$ and the residual in the one-sided, two-sided, and "no saturated" groups. (c),(d) Normalized saturated area $\mathbb{E}[A \mid Z_i = z_i]/\mathbb{E}[A]$ before (dashed) and after (solid) removing the initial noise component along u_4 and u_5, respectively. Removing u_4 reduces the saturated areas of the two groups by 11.5% and 19.3%, while removing u_5 reduces them by 10.8% and 16.2%.

I Additional Results on Eigenvector Contributions to Intensity Saturation

Figure 23 extends Figure 5 with the contributions of $u_1, \ldots, u_5$ and the residual under two aggregation schemes, and with individual ablations of u_4 and u_5. We use the same experimental settings and group definitions as Figure 5.

Contribution calculation. Let x_s be the flattened original generated image for sample s, before centering and clipping, and let u_i be a unit covariance eigenvector of its class. The mode coefficient and reconstructed component are

$$a_{s,i} = u_i^\top x_s, \qquad x_s^{(i)} = a_{s,i} u_i. \tag{35}$$

We use the original image values because intensity saturation is defined by these values, and the class mean also contributes to saturated-region pixel magnitudes. Each coefficient therefore includes both the class mean projection and the deviation from it, $a_{s,i} = u_i^\top \mu_c + u_i^\top (x_s - \mu_c)$, where μ_c is the class mean.

Let R_s contain all channel coordinates in the region used for the contribution analysis, namely the intensity-saturated region for the one-sided and two-sided groups. Define the signed mode contribution $C_{s,i}$ and regional magnitude D_s by

$$C_{s,i} = \sum_{p \in R_s} \operatorname{sgn}(x_{s,p}) a_{s,i} u_{i,p}, \qquad D_s = \sum_{p \in R_s} |x_{s,p}|. \tag{36}$$

Table 11: **Grouped contribution ratios by saturation type.** The u_2–u_5 column sums the contributions of these four eigenvectors. The sample mean and pooled ratios follow Equations (38) and (39), respectively. Each row sums to 100%.

Group	u_1	u_2–u_5	Residual
Sample mean			
One-sided	41.41%	26.38%	32.21%
Two-sided	14.83%	33.08%	52.09%
No saturated	3.69%	11.85%	84.46%
Pooled			
One-sided	50.67%	21.81%	27.52%
Two-sided	18.35%	32.43%	49.22%
No saturated	4.10%	13.74%	82.16%

The sample's contribution ratio is

$$c_{s,i} = 100\,\frac{C_{s,i}}{D_s} \quad (\%), \qquad D_s > 0. \tag{37}$$

The sign factor measures each component's contribution in the direction of the original pixel value. For the residual, we apply the same formula to $r_s = x_s - \sum_{i=1}^{5} a_{s,i} u_i$. Consequently, the contributions of the first five modes and the residual sum to 100%.

Sample mean and pooled contributions. For a group $\mathcal{S}$ of N evaluated samples with $D_s > 0$, we compute

$$\bar{c}_i^{\text{sample}} = \frac{100}{N} \sum_{s \in \mathcal{S}} \frac{C_{s,i}}{D_s}, \tag{38}$$

$$c_i^{\text{pooled}} = 100\,\frac{\sum_{s \in \mathcal{S}} C_{s,i}}{\sum_{s \in \mathcal{S}} D_s}. \tag{39}$$

The sample mean assigns equal weight to each sample. The pooled ratio combines all evaluated regions within the group and equals the sample contribution ratios weighted by D_s. Figure 5(a) and Figure 23(a) report the sample mean. Figure 23(b) reports the pooled contributions. Table 11 gives the values for both aggregation schemes, grouped into u_1, u_2–u_5, and the residual.

Initial noise component removal. We use the same 10,000 initial noise samples per class across 50 classes as in Figure 1. For each initial noise sample x_T, we separately remove the component along each of $u_1, \ldots, u_5$. Here, removal means setting the coefficient along the selected eigenvector to the corresponding class mean projection,

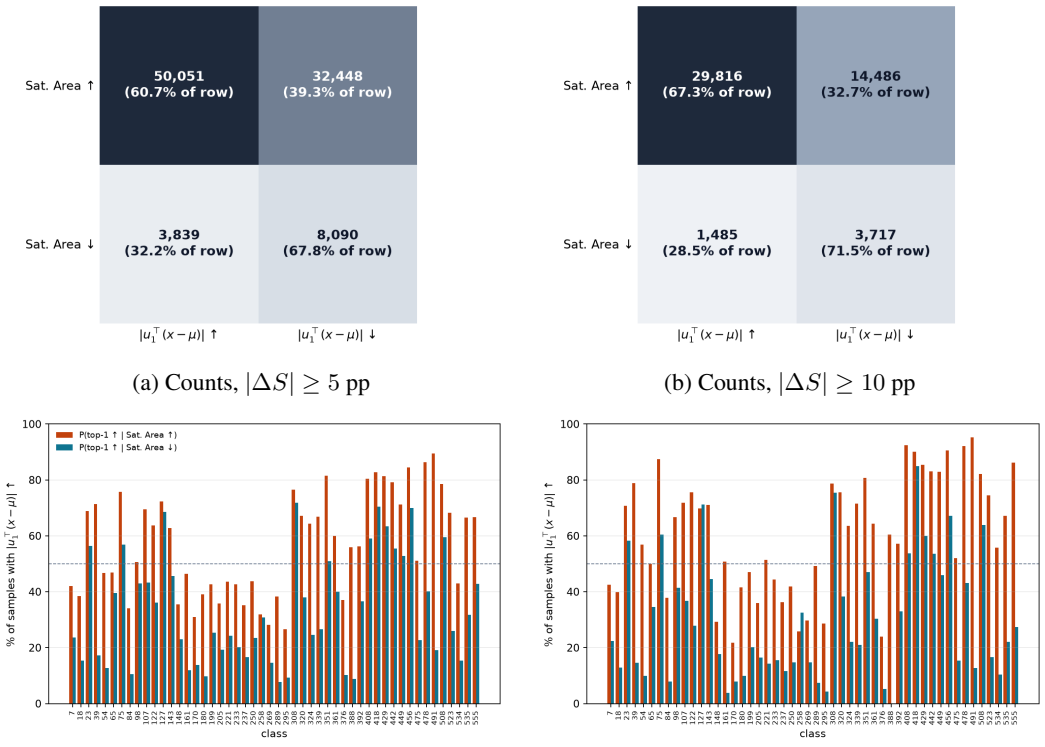
$$x_T^{(i)} = x_T - u_i^\top (x_T - \mu_c) u_i, \qquad i = 1, \ldots, 5, \tag{40}$$

where μ_c is the class mean. This gives $u_i^\top x_T^{(i)} = u_i^\top \mu_c$ and preserves all orthogonal components. We generate paired samples from the original and modified initial noise using the same sampling settings and compare their saturated areas.

Saturated-area reduction. The reductions reported in Figure 5(b) measure the decrease in total saturated area within each evaluated group. For paired samples before and after mode removal, the percentage reduction is

$$100\,\frac{\sum_s A_s^{\text{before}} - \sum_s A_s^{\text{after}}}{\sum_s A_s^{\text{before}}} \quad (\%), \tag{41}$$

where each A_s includes both dark and bright saturated areas.



(a) Counts, $|\Delta S| \geq 5$ pp

(b) Counts, $|\Delta S| \geq 10$ pp

(c) Per class, $|\Delta S| \geq 5$ pp

(d) Per class, $|\Delta S| \geq 10$ pp

Figure 24: **Top eigenvector coefficient changes and intensity saturation under CFG.** Paired EDM samples compare $w = 1$ and $w = 2$ across 50 classes, with 10,000 images per class and both saturation types included. (a,b) Sign-based counts with row percentages. (c,d) Per-class percentages of coefficient increases, grouped by saturation increase or decrease. Columns use thresholds of 5 and 10 pp, respectively.

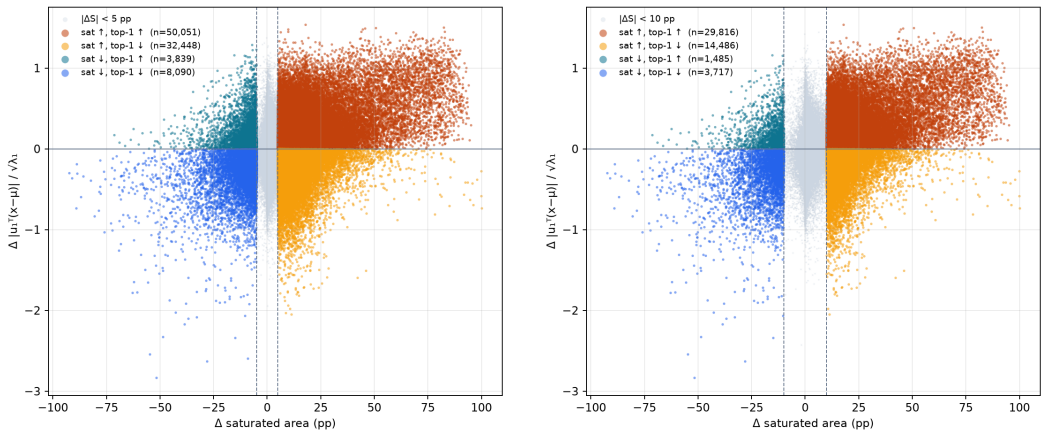
J Changes in Top Eigenvector Coefficient Magnitude and Intensity Saturation

We examine how changes in the top eigenvector coefficient magnitude under CFG relate to changes in intensity saturation. Using EDM, we compare unguided conditional samples ($w = 1$) with guided samples ($w = 2$) generated from the same initial noise, with 10,000 images per class across 50 classes at each weight. Saturated area includes both one-sided and two-sided saturation throughout this analysis.

Let $S^{(w)}$ denote the saturated area as a percentage of the image and $T^{(w)} = |u_1^\top(x_0^{(w)} - \mu)|$ the top eigenvector coefficient magnitude. We measure $\Delta S = S^{(2)} - S^{(1)}$ in percentage points (pp) and $\Delta T = T^{(2)} - T^{(1)}$. We evaluate samples with $|\Delta S| \geq 5$ pp and $|\Delta S| \geq 10$ pp separately. The scatter plots show $\Delta T/\sqrt{\lambda_1}$ to normalize coefficient changes by the standard deviation along the top eigenvector.

Figure 24(a,b) groups samples by whether CFG increases or decreases the coefficient magnitude and saturated area. At the 5 pp threshold, 60.7% of samples with increased saturated area also have increased coefficient magnitude, while 67.8% of samples with decreased saturated area also have decreased coefficient magnitude. These proportions rise to 67.3% and 71.5%, respectively, at the 10 pp threshold. The greater counts for changes in the same direction support a positive association between the two quantities.

Figure 24(c,d) examines this association within each class. The bars compare the proportion of samples with increased coefficient magnitude among samples with increased saturated area and



(a) Changes, $|\Delta S| \geq 5$ pp

(b) Changes, $|\Delta S| \geq 10$ pp

Figure 25: **Changes in saturated area and top eigenvector coefficient magnitude.** Samples follow the setup in Figure 24. Colors indicate the signs of ΔS and ΔT; gray points fall below the indicated threshold. Spearman correlations are 0.32 for (a) and 0.39 for (b).

among those with decreased saturated area, using the corresponding threshold. Across most classes, the former proportion is higher. Thus, the association observed in the pooled counts is broadly shared across classes, with variation in its strength and some exceptions.

Figure 25 shows the changes for individual samples. For samples with absolute saturated-area changes of at least 5 and 10 pp, the Spearman rank correlations are 0.32 and 0.39, respectively, indicating a moderate positive association between saturated-area changes and normalized top eigenvector coefficient magnitude changes. These results support the tendency for CFG-driven coefficient magnitude increases and decreases to accompany corresponding changes in intensity saturation. The correlations describe the selected samples with substantial saturated-area changes, while contributions from other leading eigenvectors also affect the response of individual samples.

K Accuracy of the Linear CFG Approximation

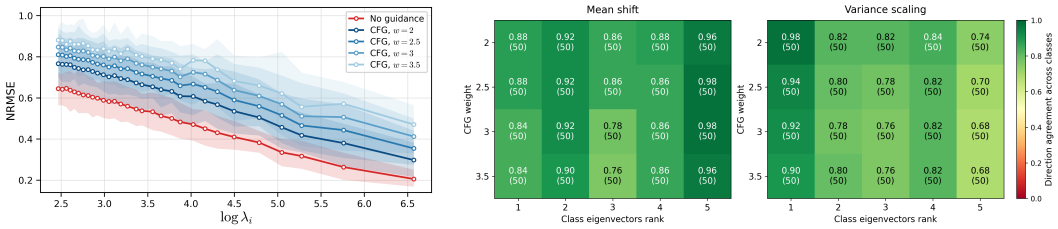
We evaluate the linear CFG approximation in Equation 7 through coefficient prediction error and agreement with the observed directions of mean shift and variance scaling. Following the setting of Figure 1(b), we use 50 classes with 10,000 generated samples per class and evaluate CFG weights $w \in \{2, 2.5, 3, 3.5\}$, with unguided conditional sampling ($w = 1$) as the baseline.

Coefficient prediction accuracy. Figure 26(a) reports the normalized root mean squared error (NRMSE) of the predicted final eigenvector coefficients. The error increases with the CFG weight and is smaller for eigenvectors with larger eigenvalues. The unguided baseline exhibits the same dependence on eigenvalue, indicating that the accuracy of the underlying linear denoiser approximation already contributes to this pattern. The guided prediction error therefore reflects both the baseline approximation error and the approximation of guidance-induced changes. Linear CFG retains appreciable errors in individual coefficient values, particularly at larger guidance weights and along eigenvectors with smaller eigenvalues.

Directional agreement of the two effects. We separately evaluate the directional predictions of mean shift and variance scaling for the five leading eigenvectors across the same 50 classes. For mean shift, we measure the fraction of classes in which the predicted shift direction agrees with the observed direction. For variance scaling, we measure the fraction in which the predicted decrease or increase in variance agrees with the observed effect in generated samples.

Figure 26(b) shows that mean shift directions are predicted with consistently high agreement, ranging from 76% to 98% across the evaluated eigenvectors and weights. Variance scaling agreement is

highest for the top eigenvector at every evaluated weight, decreasing from 98% at $w = 2$ to 90% at $w = 3.5$. Agreement generally decreases as the guidance weight increases, with some variation across eigenvectors. These results support using linear CFG to interpret the directions of guidance-induced changes, particularly variance scaling along the top eigenvector. The coefficient prediction errors delimit the quantitative precision of this interpretation.



(a) Coefficient prediction error

(b) Directional agreement

Figure 26: **Accuracy of the linear CFG approximation.** Results use 10,000 generated samples per class across 50 classes. (a) NRMSE of predicted final eigenvector coefficients versus the log covariance eigenvalue. The red curve shows the unguided conditional baseline ($w = 1$), and the blue curves show CFG weights $w \in \{2, 2.5, 3, 3.5\}$. Prediction error increases with guidance weight and decreases toward larger eigenvalues. (b) Agreement between predicted and observed mean shift directions (left) and variance decreases or increases (right) for the five leading eigenvectors. Each cell reports the fraction of agreeing classes, with the number of evaluated classes in parentheses. Mean shift agreement remains high across eigenvectors, and variance scaling agreement is highest for the top eigenvector throughout the evaluated guidance weights.

L Linear CFG under the Common Principal Components Assumption

We summarize the linear CFG solution of Li et al. (2025, Theorem 1 and Appendix B.2) in the notation of Section 4.2. Their guidance parameter γ corresponds to $w - 1$ in our formulation.

Common principal components. The common principal components assumption states that the conditional and unconditional data covariances admit a shared orthonormal eigenbasis:

$$\Sigma_c = U_c \Lambda_c U_c^\top, \qquad \Sigma_{uc} = U_c \Lambda_{uc} U_c^\top, \tag{42}$$

where $U_c = [u_{c,1}, \ldots, u_{c,d}]$, $\Lambda_c = \mathrm{diag}(\lambda_{c,i})$, and $\Lambda_{uc} = \mathrm{diag}(\lambda_{uc,i})$. The unconditional eigenvalues are indexed by their corresponding shared eigenvectors, so their ordering can differ from their magnitude ordering. This assumption permits each coefficient to evolve independently. It imposes a strong structural constraint on the full covariances; its use for leading directions in Section 4.2 is an approximation motivated by their conditional–unconditional alignment.

Variance scaling. Li et al. (2025) describe the covariance-dependent effect as *contrastive principal component (CPC) guidance*. Under the shared-basis assumption,

$$\Sigma_c - \Sigma_{uc} = U_c(\Lambda_c - \Lambda_{uc})U_c^\top. \tag{43}$$

Thus, the shared data PCs also form an eigenbasis for the covariance contrast. Positive and negative CPCs correspond to shared PCs with $\lambda_{c,i} > \lambda_{uc,i}$ and $\lambda_{c,i} < \lambda_{uc,i}$, respectively. In this setting, guidance rescales coefficients along these fixed data PCs. We therefore use the term *variance scaling* to describe its effect in Equation 7. Specifically, defining

$$S_i(\sigma_t, c) = \left(\frac{\lambda_{uc,i} + \sigma_T^2}{\lambda_{c,i} + \sigma_T^2} \frac{\lambda_{c,i} + \sigma_t^2}{\lambda_{uc,i} + \sigma_t^2} \right)^{(w-1)/2}, \tag{44}$$

Equation 7 gives, over the initial noise distribution at fixed c,

$$\mathrm{Var}_{x_T}\left[a_i^{\mathrm{cfg}}(\sigma_t)\right] = S_i(\sigma_t, c)^2 \,\mathrm{Var}_{x_T}[a_i(\sigma_t)]\,. \tag{45}$$

The additive mean shift term is constant across initial noise samples. For $w > 1$ and $\sigma_t < \sigma_T$, the scaling increases variance when $\lambda_{c,i} > \lambda_{uc,i}$ and decreases it when $\lambda_{c,i} < \lambda_{uc,i}$.

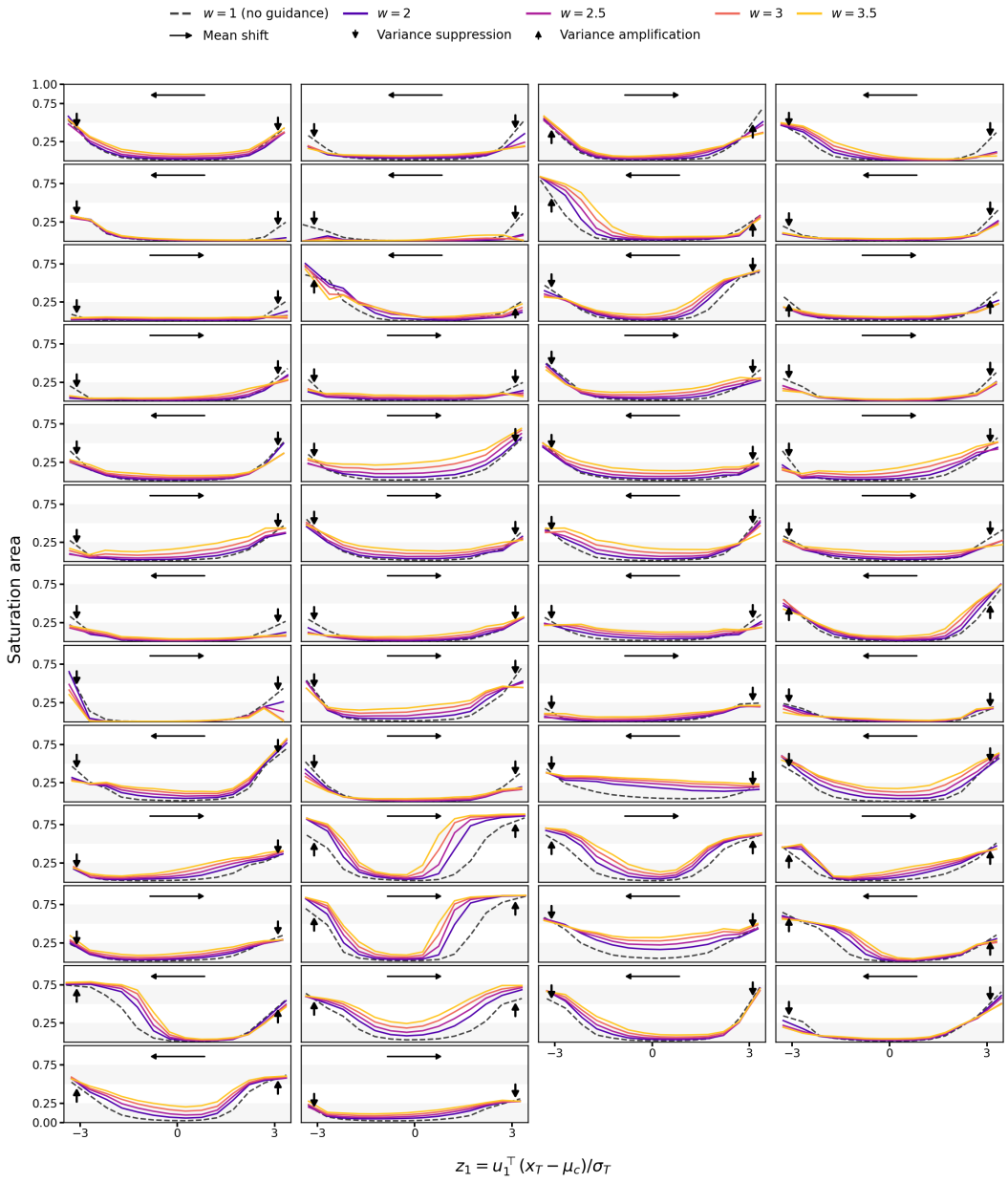


Figure 27: **Intensity saturation across initial noise projections for 50 ImageNet classes.** Each panel shows mean saturated area versus the initial noise projection z_1 onto the top eigenvector for CFG weights $w = 1, 2, 2.5, 3, 3.5$. The dashed curve denotes unguided conditional sampling. Horizontal arrows indicate the predicted mean shift direction, and vertical arrows indicate predicted decreases or increases in variance.

Mean shift coefficient. Writing $\Delta\mu_i = u_{c,i}^\top(\mu_c - \mu_{uc})$, the projected linear CFG ODE is

$$\frac{da_i^{\text{cfg}}}{d\sigma} = \sigma\left(\frac{w}{\lambda_{c,i}+\sigma^2} - \frac{w-1}{\lambda_{uc,i}+\sigma^2}\right) a_i^{\text{cfg}} - \frac{(w-1)\sigma}{\lambda_{uc,i}+\sigma^2}\Delta\mu_i. \tag{46}$$

Applying an integrating factor with the initial condition $a_i^{\text{cfg}}(\sigma_T) = u_{c,i}^\top(x_T - \mu_c)$ yields Equation 7, with

$$\boxed{B_i(\sigma_t, c) = \frac{(\lambda_{c,i} + \sigma_t^2)^{w/2}}{(\lambda_{uc,i} + \sigma_t^2)^{(w-1)/2}} \int_{\sigma_t}^{\sigma_T} \frac{s\,(\lambda_{uc,i} + s^2)^{(w-3)/2}}{(\lambda_{c,i} + s^2)^{w/2}}\, ds.} \tag{47}$$

This is the coefficient given in Appendix B.2 of Li et al. (2025), expressed using $\gamma = w - 1$. Its dependence on w and σ_T is implicit in our notation. It depends only on the paired covariance eigenvalues, guidance weight, and noise endpoints, and is independent of x_T. Consequently, $(w - 1)B_i(\sigma_t, c)\Delta\mu_i$ produces the same additive shift for every initial noise sample at fixed c.

M Saturation Changes across Classes

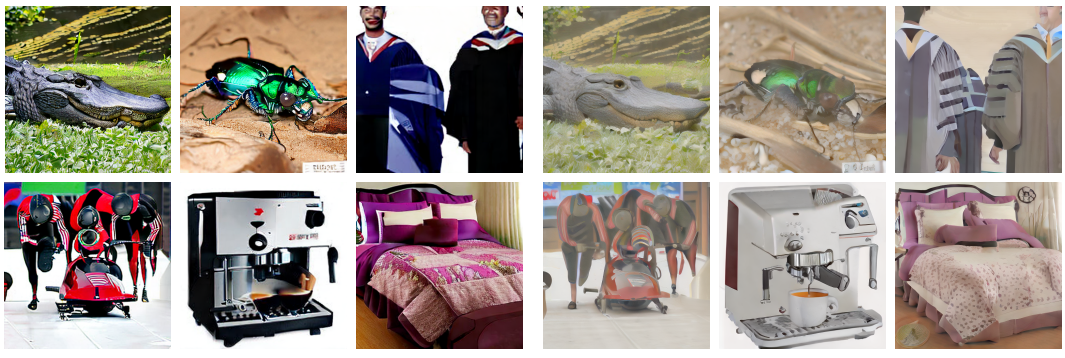
Figure 27 extends the analysis in Figure 6 to all 50 classes. We use the same 10,000 initial noise samples per class with EDM at CFG weights $w \in \{1, 2, 2.5, 3, 3.5\}$, where $w = 1$ denotes unguided conditional sampling. Each panel plots mean saturated area against the initial noise projection onto the top eigenvector. The ImageNet class indices are 7, 18, 23, 39, 54, 65, 75, 84, 98, 107, 122, 127, 143, 148, 161, 170, 180, 199, 205, 221, 233, 237, 250, 258, 269, 289, 295, 308, 320, 324, 339, 351, 361, 376, 388, 392, 408, 418, 429, 442, 449, 456, 475, 478, 491, 508, 523, 534, 535, and 555.

We assess agreement with the linear CFG predictions by comparing $w = 1$ with $w = 2$. In 41 of the 50 classes, initial noise projections in the predicted mean shift direction exhibit a larger increase in mean saturated area than projections in the opposite direction.

For variance scaling, we first require agreement in both tails of the initial noise projection. Mean saturated area decreases in both tails under a predicted decrease in variance and increases in both tails under a predicted increase in variance. This criterion holds in 21 classes. We then account for the contribution of mean shift when one tail disagrees with the variance scaling prediction. We allow an increase in saturated area in the mean shift direction or a decrease in the opposite direction, while retaining agreement with variance scaling in the other tail. Under this criterion, 35 classes agree with the predicted changes. Of these, 30 also satisfy the mean shift criterion above.

These results support qualitative tendencies across a broad range of classes in the evaluated diffusion model, with class-dependent exceptions. Approximation errors and contributions from other leading eigenvectors limit the analysis of individual samples through the top eigenvector alone, while the observed agreement supports its use in interpreting how intensity saturation changes with initial noise projections.

N Dynamic Thresholding in Latent Space



(a) CFG (b) CFG with dynamic thresholding

Figure 28: Qualitative comparison on EDM2-S with CFG weight $w = 4$. (a) and (b) show CFG and CFG with dynamic thresholding applied directly in latent space at quantile 0.997, respectively. Dynamic thresholding reduces intensity saturation while substantially degrading image quality, with muted colors, reduced contrast, and degraded fine detail.

We evaluate the effect of applying dynamic thresholding directly in latent space using the EDM2-S model used in Table 2. We use the same CFG weight, $w = 4$, and evaluate dynamic thresholding at quantiles 0.95, 0.97, 0.99, 0.995, 0.999, and 0.9999 (the 95th, 97th, 99th, 99.5th, 99.9th, and 99.99th percentiles, respectively). The procedure adaptively clips and rescales the predicted clean latent during sampling. We generate 50,000 images for each setting and compare them with the CFG baseline using the metrics described in Appendix P.

Table 12 shows that dynamic thresholding reduces intensity saturation while substantially degrading image quality across all evaluated quantiles, as reflected in higher $\mathrm{FD}_{\mathrm{DINOv2}}$ and lower precision and HPSv2.1 relative to the CFG baseline. In particular, quantile 0.95 yields an ECA@10 reduction comparable to that of our method, lowering ECA@10 from 8.83% to 6.30%, while still substantially degrading image quality. At this quantile, $\mathrm{FD}_{\mathrm{DINOv2}}$ increases from 94.5 to 111.6, precision decreases from 0.79 to 0.76, and HPSv2.1 decreases from 24.74 to 24.31. Higher quantiles yield further increases in $\mathrm{FD}_{\mathrm{DINOv2}}$ and further decreases in HPSv2.1. Figure 28 illustrates the qualitative degradation at quantile 0.997, including muted colors, reduced contrast, and degraded fine detail.

Table 12: Effect of dynamic thresholding in latent space on EDM2-S with CFG weight $w = 4$, evaluated on 50,000 generated images per setting. DT denotes dynamic thresholding, and q denotes its quantile. Across all evaluated quantiles, dynamic thresholding reduces intensity saturation while substantially degrading image quality, as reflected in higher $\mathrm{FD}_{\mathrm{DINOv2}}$ and lower precision and HPSv2.1.

		CFG with DT		
Metric	CFG	$q = 0.95$	$q = 0.97$	$q = 0.99$
$\mathrm{FD}_{\mathrm{DINOv2}} \downarrow$	94.5	111.6	116.7	138.8
Precision $\uparrow$	0.79	0.76	0.77	0.75
Recall $\uparrow$	0.37	0.38	0.38	0.38
HPSv2.1 $\uparrow$	24.74	24.31	24.04	22.84
ECA@10 (%) $\downarrow$	8.83	6.30	3.94	0.61
HSV saturation $\downarrow$	0.44	0.40	0.32	0.23

		CFG with DT		
Metric	CFG	$q = 0.995$	$q = 0.999$	$q = 0.9999$
$\mathrm{FD}_{\mathrm{DINOv2}} \downarrow$	94.5	159.1	241.8	454.3
Precision $\uparrow$	0.79	0.72	0.58	0.32
Recall $\uparrow$	0.37	0.37	0.38	0.41
HPSv2.1 $\uparrow$	24.74	21.97	19.79	16.98
ECA@10 (%) $\downarrow$	8.83	0.03	0.00	0.00
HSV saturation $\downarrow$	0.44	0.20	0.17	0.19

O METHOD DETAILS

This appendix describes the coefficient normalization, gain, and application procedure for the magnitude-dependent damping in Section 5. The method reduces large coefficients along leading unconditional covariance eigenvectors, with weaker corrections to smaller coefficients.

Coefficient normalization. Let μ_{uc} denote the unconditional mean, and let u_k and λ_k denote the eigenvectors and eigenvalues of the unconditional covariance, ordered by decreasing eigenvalue. We use the first r eigenvectors, with $u_k^\top u_\ell = \delta_{k\ell}$. At sampling step i, we compute

$$\tilde{a}_k(\sigma_i) = u_k^\top (x_i - \mu_{uc}), \qquad s_{i,k}^2 = \lambda_k + \sigma_i^2, \qquad z_{i,k} = \frac{\tilde{a}_k(\sigma_i)}{s_{i,k}}. \tag{48}$$

In coordinates $x_\sigma = x_0 + \sigma\epsilon$, independent noise $\epsilon \sim \mathcal{N}(0, I)$ contributes variance σ^2 along each eigenvector. The reference scale therefore accounts for differences in variance across eigenvectors and noise levels. The normalized coefficient measures the magnitude of the current coefficient relative

to this scale. Along a guided sampling trajectory, the actual coefficient variance can differ from the reference variance.

Magnitude-dependent gain. The gain in Equation 8 is

$$g_k(\sigma_i) = \big(1 + 2\alpha w_{i,k} z_{i,k}^2\big)^{-1/2}, \tag{49}$$

where $\alpha > 0$ controls the damping strength and $w_{i,k} \geq 0$ allocates the correction across sampling steps. The corrected state is

$$\widetilde{x}_i = x_i + \sum_{k=1}^{r} \big(g_k(\sigma_i) - 1\big)\tilde{a}_k(\sigma_i) u_k. \tag{50}$$

By orthonormality, the corrected coefficient along u_k is $g_k(\sigma_i)a_k(\sigma_i)$, and the components orthogonal to the selected eigenvectors remain unchanged during the correction. For finite inputs, $0 < g_k(\sigma_i) \leq 1$, so the correction preserves coefficient signs. For fixed $\alpha w_{i,k}$, the relative reduction increases with $|z_{i,k}|$. For small $2\alpha w_{i,k} z_{i,k}^2$,

$$1 - g_k(\sigma_i) = \alpha w_{i,k} z_{i,k}^2 + O\big((\alpha w_{i,k} z_{i,k}^2)^2\big)\,. \tag{51}$$

Thus, the correction becomes weak as the coefficient approaches zero.

To obtain Equation 49, we hold the noise level and weight fixed and integrate

$$\frac{d\tilde{a}}{d\eta} = -\frac{\alpha w_{i,k}}{s_{i,k}^2}\tilde{a}^3, \qquad \tilde{a}(0) = \tilde{a}_k(\sigma_i), \qquad 0 \leq \eta \leq 1. \tag{52}$$

Here η parameterizes the correction at a fixed sampling step. The exact solution is

$$\tilde{a}(1) = \frac{\tilde{a}(0)}{\sqrt{1 + 2\alpha w_{i,k} a(0)^2 / s_{i,k}^2}}, \tag{53}$$

which gives Equation 49. A zero coefficient remains zero. The exact solution describes this individual correction; the subsequent sampler update determines the next sampling state.

Weights across sampling steps. We define the noise-dependent weights

$$\omega_{i,k} = \frac{\sigma_i^2}{(\lambda_k + \sigma_i^2)^2}. \tag{54}$$

This choice is motivated by the signal and noise fractions of the reference variance:

$$\frac{\lambda_k}{\lambda_k + \sigma_i^2}\frac{\sigma_i^2}{\lambda_k + \sigma_i^2} = \lambda_k \omega_{i,k}. \tag{55}$$

For $\lambda_k > 0$, the continuous weighting function is largest at $\sigma^2 = \lambda_k$, where the two contributions are equal, and decreases toward both high and low noise levels. The realized gain additionally depends on the current normalized coefficient.

For N sampling steps indexed by $i = 0, \ldots, N-1$, the supplied implementation applies the correction at $A = \{1, \ldots, N-1\}$. Step zero uses the initialized state directly. We define

$$s_{*,k}^2 = \lambda_k + \min_{j \in A} \sigma_j^2 \tag{56}$$

and normalize the weights independently for each selected eigenvector:

$$w_{i,k} = \frac{\omega_{i,k}}{\sum_{j \in A} \omega_{j,k} s_{*,k}^2 / s_{j,k}^2}, \qquad \sum_{i \in A} w_{i,k} \frac{s_{*,k}^2}{s_{i,k}^2} = 1. \tag{57}$$

The factor λ_k in the product of signal and noise fractions cancels under this normalization. The scale factors account for the change produced by an individual correction. For a nonzero input coefficient a and its corrected value $\widetilde{a}$, Equation 53 gives

$$\frac{1}{\widetilde{a}^2} - \frac{1}{\tilde{a}^2} = \frac{2\alpha w_{i,k}}{s_{i,k}^2}. \tag{58}$$

Algorithm 1 Magnitude-dependent damping during sampling

Require: Unconditional mean μ_{uc}; leading eigenvectors and eigenvalues $\{(u_k, \lambda_k)\}_{k=1}^{r}$
Require: Noise schedule $\sigma_0 > \cdots > \sigma_{N-1} > \sigma_N = 0$ with $N \geq 2$; strength $z_{\max} > 0$; sampler updates $\{\mathcal{S}_i\}_{i=0}^{N-1}$
Ensure: Final sampling state x_N

$A \leftarrow \{1, \ldots, N-1\}$; $\alpha \leftarrow 1/(2z_{\max}^2)$
for $k = 1, \ldots, r$ **do**
  $s_{*,k}^2 \leftarrow \lambda_k + \min_{j \in A} \sigma_j^2$
  for $i \in A$ **do**
    $s_{i,k}^2 \leftarrow \lambda_k + \sigma_i^2$
    $\omega_{i,k} \leftarrow \sigma_i^2 / (s_{i,k}^2)^2$
  end for
  for $i \in A$ **do**
    $w_{i,k} \leftarrow \omega_{i,k} / \left(\sum_{j \in A} \omega_{j,k} s_{*,k}^2 / s_{j,k}^2\right)$
  end for
end for
Draw $\epsilon \sim \mathcal{N}(0, I)$; $x_0 \leftarrow \sigma_0 \epsilon$
for $i = 0, \ldots, N-1$ **do**
  $\widetilde{x}_i \leftarrow x_i$
  if $i \in A$ **then**
    for $k = 1, \ldots, r$ **do**
      $\tilde{a}_k(\sigma_i) \leftarrow u_k^\top (x_i - \mu_{uc})$
      $z_{i,k} \leftarrow \tilde{a}_k(\sigma_i) / s_{i,k}$
      $g_k(\sigma_i) \leftarrow (1 + 2\alpha w_{i,k} z_{i,k}^2)^{-1/2}$
    end for
    $\widetilde{x}_i \leftarrow x_i + \sum_{k=1}^{r} (g_k(\sigma_i) - 1) \tilde{a}_k(\sigma_i) u_k$
  end if
  $x_{i+1} \leftarrow \mathcal{S}_i(\widetilde{x}_i)$
end for
return x_N

The weights therefore account for the reference scale at each sampling step. When the schedule or correction steps change, the weights and reference scales are recomputed.

The implementation specifies the strength through $z_{\max} > 0$:

$$\alpha = \frac{1}{2z_{\max}^2}, \qquad g_k(\sigma_i) = \left(1 + \frac{w_{i,k} z_{i,k}^2}{z_{\max}^2}\right)^{-1/2}. \tag{59}$$

Smaller $z_{\max}$ gives stronger damping for the same state and weights. The number of selected eigenvectors r and $z_{\max}$ specify the method once the basis, schedule, and correction steps are fixed. The final coefficient magnitudes depend on the full sampling trajectory, since each sampler update changes the coefficients used in the next correction.

Application during sampling. Algorithm 1 gives the application procedure. Each $\mathcal{S}_i$ denotes one sampler update, including the chosen guidance method. The correction is applied once at the start of each selected outer step, and the coefficients and gains are recomputed from the current state. It introduces no additional denoiser evaluations.

The covariance eigenvectors, mean, and corrected state use the same coordinates and scaling. For a variance-preserving parameterization $x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon$, the correction is computed in coordinates $y_t = x_t / \sqrt{\bar{\alpha}_t}$ with $\sigma_t = \sqrt{(1 - \bar{\alpha}_t)/\bar{\alpha}_t}$ at steps with $\bar{\alpha}_t > 0$. The corrected state is multiplied by $\sqrt{\bar{\alpha}_t}$ before the sampler update. The experimental settings for Table 2 are presented in the following appendix.

Table 13: Guidance parameters used in Table 2. APG additionally uses $\eta = 0$, $r_{\mathrm{APG}} = 2.5$, and $\beta = -0.75$ for all models.

Model	CFG weight	APG weight	CFG++ coefficient
EDM2-S	4	4	1.22
EDM2-XXL	2	2	0.6
DiT-XL/2	4	4	0.09

Table 14: Settings for guidance interval, shared by the baseline and its counterpart augmented with our method. The CFG weight is 1 outside the listed interval. Both EDM2 models use second-order Heun sampling with no churn.

Model	$(\sigma_{\mathrm{lo}}, \sigma_{\mathrm{hi}}]$	w_{in}	Sampler	Steps
EDM2-S	$(0.60, 5.00]$	4	Heun	32
EDM2-XXL	$(0.60, 5.00]$	2	Heun	32
DiT-XL/2	$(0.45, 1.23]$	4	iDDPM	250

P Experimental Settings for Table 2

This appendix describes the models, guidance parameters, unconditional statistics, and sample construction used for Table 2. The correction and its application during sampling are described in Appendix O.

Models and guidance parameters. We evaluate CFG, APG, CFG++, and guidance interval with and without our method on EDM2-S, EDM2-XXL, and DiT-XL/2 at 512×512 resolution. The checkpoints are

- EDM2-S: `edm2-img512-s-2147483-0.070.pkl`.
- EDM2-XXL: `edm2-img512-xxl-0939524-0.015.pkl`.
- DiT-XL/2: `facebook/DiT-XL-2-512`.

The CFG weights are $w = 4$ for EDM2-S, $w = 2$ for EDM2-XXL, and $w = 4$ for DiT-XL/2. APG uses the corresponding model-specific guidance weight together with $\eta = 0$, $r_{\mathrm{APG}} = 2.5$, and $\beta = -0.75$. Here r_{APG} denotes the APG parameter, and r in Appendix O denotes the number of eigenvectors used by our method. The CFG++ guidance coefficients are 1.22 for EDM2-S, 0.6 for EDM2-XXL, and 0.09 for DiT-XL/2. Table 13 summarizes these settings.

Samplers and sampling steps. For EDM2-S and EDM2-XXL, CFG and APG use a deterministic second-order Heun sampler with no churn, and CFG++ uses the DDIM CFG++ sampler with $\eta = 0$. All three guidance methods use 32 sampling steps for both EDM2 models. For DiT-XL/2, CFG and APG use DDIM with $\eta = 0$, and CFG++ uses DDIM CFG++ with $\eta = 0$; all three guidance methods use 250 sampling steps.

Guidance interval. We evaluate guidance interval (Kynkäänniemi et al., 2024) by applying the model-specific CFG weight only within a prescribed noise interval. Specifically,

$$w(\sigma) = \begin{cases} w_{\mathrm{in}}, & \sigma_{\mathrm{lo}} < \sigma \leq \sigma_{\mathrm{hi}}, \\ 1, & \text{otherwise.} \end{cases} \tag{60}$$

Table 14 gives the interval, weight, sampler, and step count for each model. For guidance interval augmented with our method, the regulation step is also applied only when $\sigma_{\mathrm{lo}} < \sigma \leq \sigma_{\mathrm{hi}}$, using $r = 5$ and $z_{\max} = 2$. Outside this interval, sampling proceeds with $w = 1$ and the regulation step is inactive. All remaining experimental and evaluation settings, including checkpoints, unconditional statistics, initial noise, and class labels, follow the common protocol described in this appendix.

Unconditional statistics and damping strength. The unconditional mean and covariance used by our method are estimated from 50,000 generated images for each model. Their representations use the same coordinates and scaling as the sampling state to which the correction is applied. The leading eigenvectors and corresponding eigenvalues of this covariance provide the directions and reference scales defined in Appendix O. We use the five leading eigenvectors ($r = 5$) for all models and guidance methods augmented with our method. We use $z_{\max} = 2$ for every setting augmented with our method, giving $\alpha = 1/(2z_{\max}^2) = 1/8$. The weights across sampling steps are computed from the eigenvalues and the noise schedule using Equation 57.

Evaluation samples. For each model and method setting, we generate 50,000 images, with 50 images for each of the 1,000 ImageNet classes. The 50,000 initial noise samples are generated using seeds 0–49,999. Within each model, the compared methods use the same initial noise for each corresponding sample and the same class label. This pairing also applies to each guidance method and its counterpart augmented with our method.

Evaluation metrics. We evaluate each 50,000-image set using ECA@10, HSV saturation, precision, HPSv2.1, $\text{FD}_{\text{DINOv2}}$ (Stein et al., 2023), and recall, as reported in Table 2. ECA@10 is defined below and measures the average clipped area exceeding 10% of an image. HSV saturation measures color saturation. Precision, HPSv2.1, $\text{FD}_{\text{DINOv2}}$, and recall provide the additional image-quality and diversity measurements used in the comparison.

Excess Clip Area (ECA@10). Let $v_{i,c}(p) \in \{0, \dots, 255\}$ be the stored RGB value of channel c at pixel p in generated image i. We express this value on $[-1, 1]$ as $x_{i,c}(p) = 2v_{i,c}(p)/255 - 1$ and define the pixel-level clip indicator by

$$C_i(p) = \mathbf{1}\left\{ \min_{c\in\{R,G,B\}} x_{i,c}(p) > 0.95 \ \vee \ \max_{c\in\{R,G,B\}} x_{i,c}(p) < -0.95 \right\}. \tag{61}$$

Here $\mathbf{1}\{\cdot\}$ equals one when its condition holds and zero otherwise. A pixel is counted when all three channels have normalized values greater than 0.95 or all three have normalized values less than -0.95. For 8-bit images, this is equivalent to requiring $v_{i,c}(p) \le 6$ or $v_{i,c}(p) \ge 249$ for each channel. The indicator can therefore be evaluated directly on the stored integer values.

For the pixel set Ω_i of image i, its clipped area fraction is

$$f_i = \frac{1}{|\Omega_i|} \sum_{p\in\Omega_i} C_i(p), \qquad f_i \in [0, 1]. \tag{62}$$

With $N = 50{,}000$ evaluation images and area threshold $\tau = 0.10$, Excess Clip Area is

$$\text{ECA@10} = \frac{1}{N} \sum_{i=1}^{N} \max(f_i - \tau, 0). \tag{63}$$

Every generated image contributes to the denominator. An image with $f_i \le \tau$ contributes zero, and an image with $f_i > \tau$ contributes its clipped area fraction above the threshold. For example, an image with $f_i = 0.25$ contributes 0.15 before averaging. Table 2 reports 100 times Equation 63, expressing the mean excess area in percentage points. Lower values indicate a smaller average excess clipped area.

To describe how this score combines frequency and area, define $M_\tau = \sum_{i=1}^{N} \mathbf{1}\{f_i > \tau\}$. The Clip Rate and Conditional Area are

$$\text{CR@10} = \frac{M_\tau}{N}, \tag{64}$$

$$\text{CA@10} = \frac{1}{M_\tau} \sum_{i:f_i>\tau} f_i, \qquad M_\tau > 0. \tag{65}$$

Clip Rate is the fraction of images whose clipped area exceeds 10%, and Conditional Area is the average clipped area among those images. In fractional units, these quantities satisfy

$$\text{ECA@10} = \text{CR@10}\big(\text{CA@10} - \tau\big), \qquad M_\tau > 0. \tag{66}$$

When $M_\tau = 0$, both Clip Rate and Excess Clip Area are zero, and Conditional Area is undefined. ECA@10 thus combines the fraction of images exceeding the threshold and the average excess area within those images in a single score.

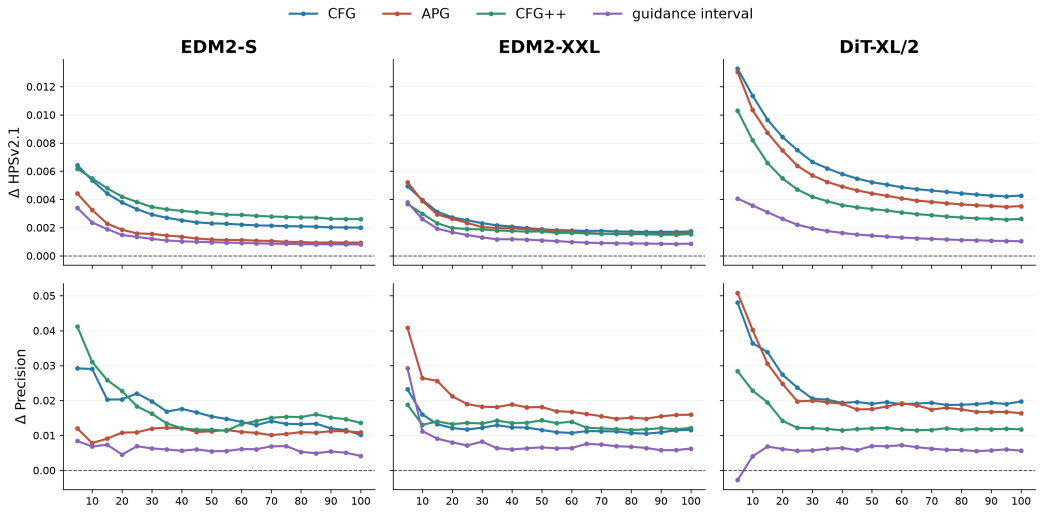


Figure 29: Mean changes in HPSv2.1 and precision among the top $n\%$ of samples ranked by their contribution to the ECA@10 reduction in Table 2. The horizontal axis gives n, ranging from 5 to 100 in increments of 5. Changes are measured relative to the corresponding guidance method without our method. Positive values indicate improvements. Samples with larger reductions in excess clipped area generally show greater improvements in image quality.

Q Intensity Saturation and Image Quality

We examine how reductions in excessive intensity saturation relate to changes in image quality using the experiments reported in Table 2. For each model and guidance method, we pair samples generated with and without our method using the same initial noise and class label. We rank these pairs by their contribution to the reduction in ECA@10, measured as the decrease in each sample's excess clipped area above the 10% threshold. For the top $n\%$ of ranked pairs, we compute the mean changes in HPSv2.1 and precision after applying our method.

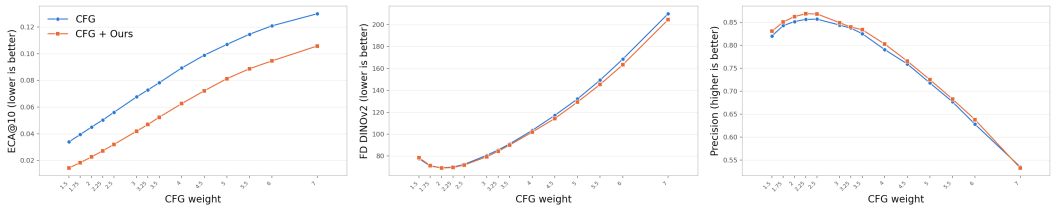
Figure 29 shows that samples contributing most to the reduction in ECA@10 generally exhibit larger improvements in image quality. The mean HPSv2.1 improvement is largest among the samples with the greatest reductions in excess clipped area and decreases as more samples are included. Precision exhibits a similar overall tendency across most settings, with a slight decrease among samples with large reductions in excess clipped area for DiT-XL/2 with guidance interval. Overall, these results support an association between greater reductions in excessive intensity saturation and larger improvements in image quality.

R Effect of CFG Weight

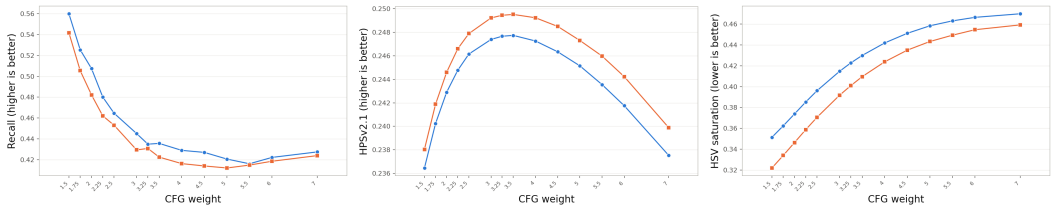
We compare CFG with and without our method on EDM2-S at CFG weights $w \in \{1.5, 1.75, 2, 2.25, 2.5, 3, 3.25, 3.5, 4, 4.5, 5, 5.5, 6, 7\}$. For each weight and method, we generate 20,000 images and compute the evaluation metrics. Our method uses $z_{\max} = 2$ and $r = 5$, matching the correction parameters used in Table 2.

Figure 30 compares ECA@10, $\text{FD}_{\text{DINOv2}}$, precision, recall, HPSv2.1, and HSV saturation at the same CFG weight. Our method lowers ECA@10 and HSV saturation and raises HPSv2.1 at every evaluated weight. Precision is higher for $w \leq 6$ and comparable at $w = 7$. $\text{FD}_{\text{DINOv2}}$ is comparable for $w \leq 2.5$ and lower for $w \geq 3$, with the gap widening as w increases. Recall is lower throughout the evaluated range, consistent with the trade-off between image quality and diversity discussed in Section 5.

Since our method lowers recall at a fixed weight, a comparison at the same weight leaves the two methods at different points on the quality and diversity trade-off. Figure 31 therefore compares the methods at matched recall. For each weight w used with our method, we select the baseline

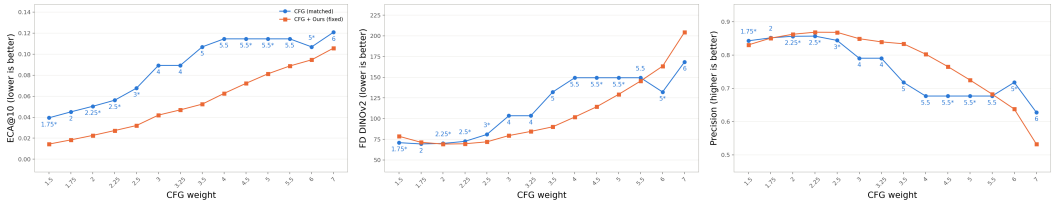


(a) ECA@10 (b) $\mathrm{FD}_{\mathrm{DINOv2}}$ (c) Precision

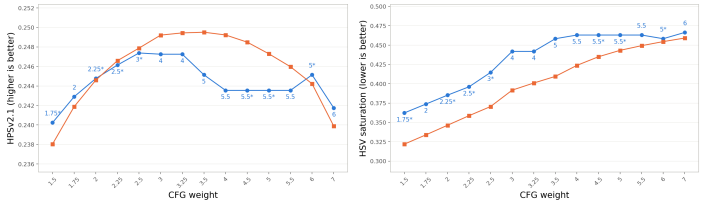


(d) Recall (e) HPSv2.1 (f) HSV saturation

Figure 30: Effect of CFG weight on EDM2-S, evaluated using 20,000 generated images per weight and method. Blue circles indicate CFG, and orange squares indicate CFG with our method ($z_{\max} = 2$, $r = 5$), compared at the same weight. Our method lowers ECA@10 and HSV saturation and raises HPSv2.1 at every weight, raises precision for $w \leq 6$, and lowers $\mathrm{FD}_{\mathrm{DINOv2}}$ for $w \geq 3$. Recall is lower throughout the evaluated range.



(a) ECA@10 (b) $\mathrm{FD}_{\mathrm{DINOv2}}$ (c) Precision

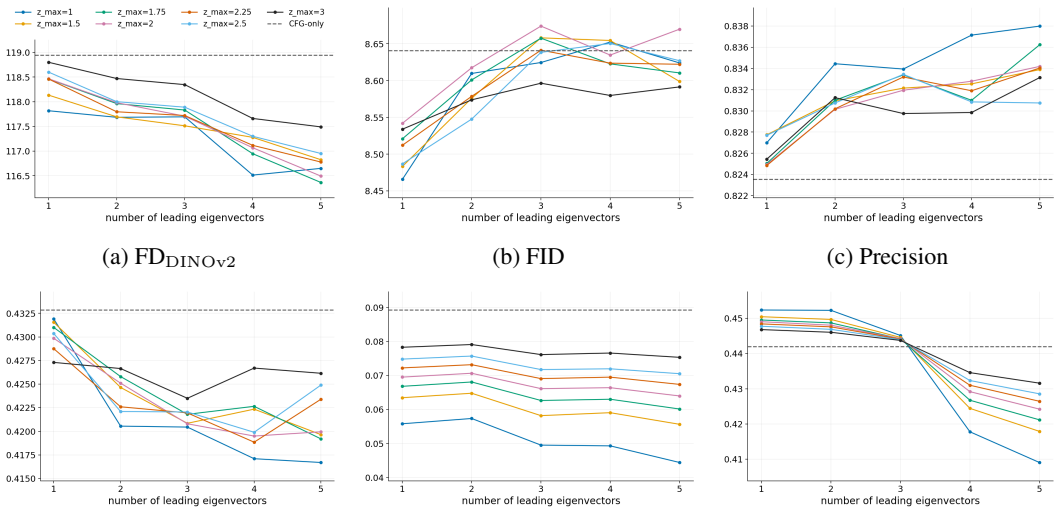


(d) HPSv2.1 (e) HSV saturation

Figure 31: Comparison with baseline CFG at matched recall on EDM2-S. The horizontal axis gives the CFG weight used with our method ($z_{\max} = 2$, $r = 5$; orange squares). Blue circles show baseline CFG at the selected weight, annotated in blue. The selected baseline weight has an absolute recall difference of at most 0.002 from our method; when no evaluated weight meets this tolerance, we use the weight with the closest recall and mark it with $*$. Our method improves all five metrics for $2.25 \leq w \leq 5.5$, and ECA@10 and HSV saturation remain lower at every weight.

CFG weight whose absolute recall difference from our method is at most 0.002. When no evaluated baseline weight meets this tolerance, we select the weight with the closest recall and mark it with an asterisk ($*$). The horizontal axis gives the weight used with our method, and the blue annotations give the selected baseline weights.

At matched recall, our method improves all five metrics for $2.25 \leq w \leq 5.5$: ECA@10, $\mathrm{FD}_{\mathrm{DINOv2}}$, and HSV saturation are lower, and precision and HPSv2.1 are higher. This range covers nine of the fourteen evaluated weights and includes the weights at which HPSv2.1 peaks for both methods. ECA@10 and HSV saturation remain lower at every evaluated weight. At $w = 2$, precision is higher,



(a) $\mathrm{FD}_{\mathrm{DINOv2}}$ (b) FID (c) Precision

(d) Recall (e) ECA@10 (f) HSV saturation

Figure 32: Sweep over the number of leading eigenvectors r used by our method on EDM2-S with CFG weight $w = 4$. Each curve corresponds to a fixed $z_{\max}$. Panels show (a) $\mathrm{FD}_{\mathrm{DINOv2}}$, (b) FID, (c) precision, (d) recall, (e) ECA@10, and (f) HSV saturation, evaluated using 20,000 generated images per method and parameter setting. Dashed lines indicate CFG alone.

while $\mathrm{FD}_{\mathrm{DINOv2}}$ and HPSv2.1 are within noise of the selected baseline. At the two smallest weights ($w = 1.5, 1.75$) and the two largest weights ($w = 6, 7$), $\mathrm{FD}_{\mathrm{DINOv2}}$, precision, and HPSv2.1 favor the selected baseline. These results show that the quality gains of our method persist when recall is controlled, over a range that includes most of the larger weights.

S Effect of Correction Hyperparameters

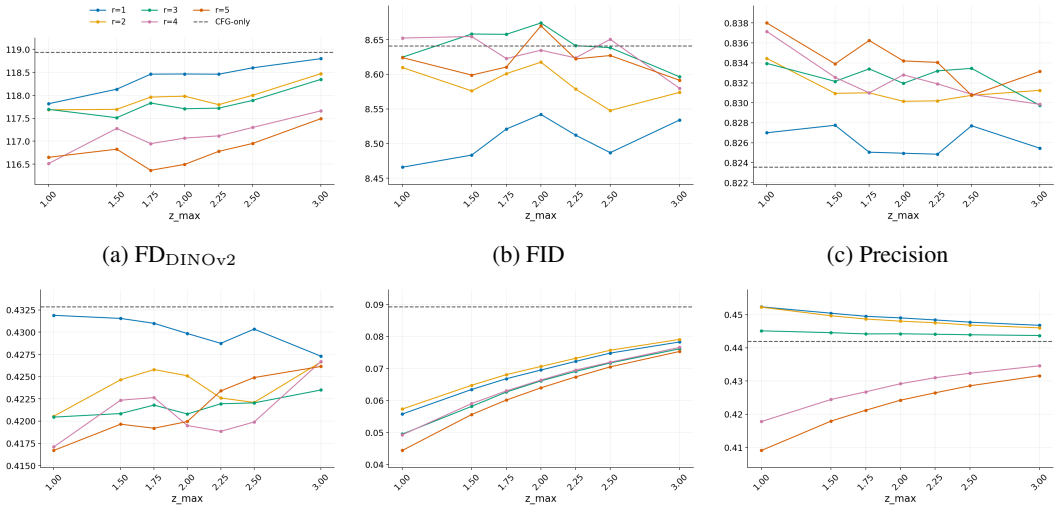
We evaluate the effects of the number of leading eigenvectors r and the correction parameter $z_{\max}$ on EDM2-S with CFG weight $w = 4$. We use 20,000 generated images for each method and parameter setting. Figures 32 and 33 show sweeps over r and $z_{\max}$, respectively. Dashed lines indicate the results obtained using CFG alone.

Smaller $z_{\max}$ and larger r generally yield greater reductions in intensity saturation, as measured by ECA@10. This trend is consistent with the analysis in Section 4.2, which relates reductions in coefficient magnitudes along leading covariance eigenvectors to reduced intensity saturation. HSV saturation falls below the CFG baseline for $r \geq 4$, with smaller $z_{\max}$ producing larger reductions in this range. Across all evaluated parameter settings, our method yields higher precision and lower recall than CFG alone.

T Brightness Changes and Diversity

Applying our method generally increases FID (Heusel et al., 2017) and decreases recall. We examine whether changes in brightness, one of the visual features associated with leading covariance eigenvectors (Appendix E), contribute to these metric changes. To this end, we adjust the brightness of images generated with Ours and reevaluate FID and recall.

Brightness correction. For each model and guidance method, let $\{x_j^{\mathrm{only}}\}_{j=1}^N$ denote the images generated using the original guidance method and $\{x_i^{\mathrm{ours}}\}_{i=1}^N$ those generated with Ours, where $N = 50{,}000$ and pixel values lie in $[0, 1]$. We define the brightness of an image $x \in [0, 1]^{H \times W \times 3}$ as



(a) $\mathrm{FD}_{\mathrm{DINOv2}}$ (b) FID (c) Precision

(d) Recall (e) ECA@10 (f) HSV saturation

Figure 33: Sweep over the correction parameter $z_{\max}$ used by our method on EDM2-S with CFG weight $w = 4$. Each curve corresponds to a fixed number of leading eigenvectors r. Panels show (a) $\mathrm{FD}_{\mathrm{DINOv2}}$, (b) FID, (c) precision, (d) recall, (e) ECA@10, and (f) HSV saturation, evaluated using 20,000 generated images per method and parameter setting. Dashed lines indicate CFG alone.

its mean over all spatial locations and RGB channels:

$$b(x) = \frac{1}{3HW} \sum_{h=1}^{H} \sum_{w=1}^{W} \sum_{c=1}^{3} x_{h,w,c}. \tag{67}$$

We compute a common target brightness from the entire baseline image set,

$$\mu_{\mathrm{only}} = \frac{1}{N} \sum_{j=1}^{N} b\left(x_j^{\mathrm{only}}\right), \tag{68}$$

separately for each model and guidance method. This target aggregates all 50,000 baseline images across classes. We then add the same scalar to every pixel and RGB channel of each Ours image:

$$\widetilde{x}_i = x_i^{\mathrm{ours}} + \left[\mu_{\mathrm{only}} - b(x_i^{\mathrm{ours}})\right] \mathbf{1}, \tag{69}$$

where $\mathbf{1}$ is the all-ones tensor. Before clipping, this operation gives $b(\widetilde{x}_i) = \mu_{\mathrm{only}}$ for every image and preserves differences between pixel values. For evaluation, we clip to the valid range and round to 8-bit values:

$$x_i^{\mathrm{BC}} = \frac{1}{255} \operatorname{round}(255 \operatorname{clip}(\widetilde{x}_i, 0, 1)). \tag{70}$$

Clipping and rounding can shift the final mean away from the target, and clipping can alter pixel differences at affected locations. We recompute FID and recall using $\{x_i^{\mathrm{BC}}\}_{i=1}^{N}$ and retain the original baseline images for comparison.

Results. Table 15 reports the results. Brightness correction reduces the absolute gap to the baseline in 9 of 12 settings for FID and 11 of 12 settings for recall. In the remaining settings for EDM2-S with CFG++ (FID and recall) and DiT-XL/2 with APG (FID), the gap widens in favor of Ours. For DiT-XL/2 with guidance interval, FID increases further after correction, from 3.90 to 4.11, compared with the baseline value of 3.70. The underlining reflects comparisons of unrounded metrics, so improvements can occur even when displayed recall values coincide.

Figure 34 visualizes the relative reduction in the absolute metric gap. For a metric m, we compute

$$R_m = 100 \left(1 - \frac{|m_{\mathrm{BC}} - m_{\mathrm{only}}|}{|m_{\mathrm{ours}} - m_{\mathrm{only}}|}\right), \tag{71}$$

Table 15: FID and recall for each guidance method, with Ours (+ours), and with additional brightness correction (BC). Each condition uses 50,000 images. Ours uses $z_{\max} = 2$ and $r = 5$. Underlined values indicate either a reduced absolute gap to the corresponding baseline compared with +ours, or a wider gap favoring Ours after brightness correction, assessed using unrounded metrics. All values are rounded to two decimal places.

	EDM2-S		EDM2-XXL		DiT-XL/2	
Method	FID ↓	Recall ↑	FID ↓	Recall ↑	FID ↓	Recall ↑
CFG	6.68	0.37	5.48	0.49	16.35	0.20
+ours	6.76	0.36	5.70	0.47	16.61	0.19
+ours with BC	6.65	0.36	5.60	0.47	16.18	0.20
APG	3.68	0.49	1.72	0.60	10.84	0.32
+ours	4.06	0.47	1.98	0.58	10.97	0.31
+ours with BC	3.99	0.47	1.95	0.59	10.68	0.32
CFG++	9.22	0.29	7.93	0.40	13.43	0.26
+ours	9.07	0.29	8.04	0.39	13.76	0.25
+ours with BC	8.89	0.29	7.92	0.39	13.44	0.25
Guidance interval	4.04	0.49	2.30	0.58	3.70	0.64
+ours	4.13	0.48	2.41	0.57	3.90	0.64
+ours with BC	4.06	0.49	2.35	0.58	4.11	0.64

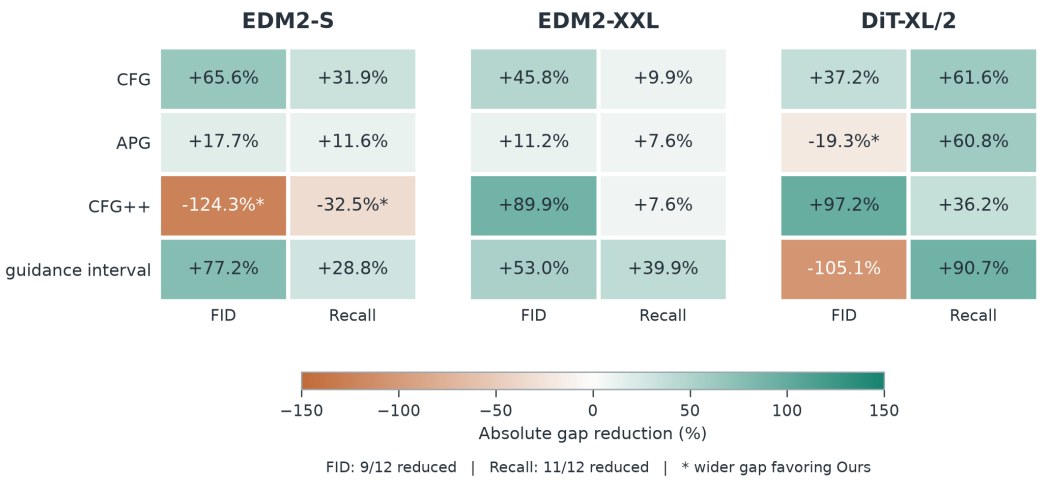


Figure 34: Relative reduction in the absolute gap to the original guidance baseline after brightness correction, computed from unrounded FID and recall values. Positive percentages indicate a smaller absolute gap. Asterisks mark wider gaps favoring Ours. Absolute gaps decrease in 9 of 12 settings for FID and 11 of 12 settings for recall.

using unrounded metric values. Positive values indicate a smaller absolute gap, and negative values indicate a wider gap. Asterisks identify wider gaps favoring Ours.

These results support the interpretation that brightness changes induced by Ours contribute to part of the FID increase and recall decrease. The recall results suggest that some of the measured loss of diversity is associated with these brightness changes. The correction fixes every image's mean brightness to a common target before clipping and thereby also changes brightness variation across images. Its effect therefore reflects this specific intervention, and the remaining causes of the metric degradation warrant further investigation.

U Results on SD 3.5 Medium

We evaluate our method on SD 3.5 Medium by comparing CFG with CFG augmented with our method, using a CFG weight of $w = 4.5$. For our method, we set $r = 5$ and $z_{\max} = 1.5$. We sample 30,000 captions with replacement from the 25,014 captions in the MS-COCO 2017 validation set (Lin et al., 2014), using a caption-sampling seed of 2026. The sampled set contains 17,396 distinct captions: 8,982 are selected once, and 8,414 are selected at least twice. For each method, we generate one image per sampled caption occurrence at a resolution of 1024×1024 using 40 sampling steps. The initial noise seeds range from 0 through 29,999.

Table 16 shows that our method reduces intensity saturation, with ECA@10 decreasing from 0.96 to 0.51. Bold values in Table 16 indicate improvements over the CFG baseline.

Table 16: Comparison of CFG and CFG with our method on SD 3.5 Medium with CFG weight $w = 4.5$, evaluated on 30,000 images per method. Images are generated at 1024×1024 resolution using 40 sampling steps and captions sampled with replacement from the MS-COCO 2017 validation set. Arrows indicate the preferred direction for each metric.

Metric	CFG	CFG with Ours
FID ↓	21.70	**21.56**
$\text{FD}_{\text{DINOv2}}$ ↓	224.4	**223.7**
Precision ↑	0.74	**0.75**
Recall ↑	0.34	**0.35**
HPSv2.1 ↑	28.59	**28.68**
HSV saturation ↓	0.35	**0.33**
ECA@10 ↓	0.96	**0.51**

Figure 35 shows qualitative comparisons for ten prompts. Our method reduces intensity saturation in these examples, with more visible detail in dark regions and less washed-out bright regions. Each pair shows CFG on the left and CFG with our method on the right, with the generation prompt below.

V Intensity Saturation in the Training Dataset

Figure 36 compares intensity saturation in ImageNet, the training dataset used for EDM, EDM2, and DiT, with that in images generated without guidance and at several CFG weights. The generated-image results in panels (a) and (b) are the same as those in Figures 4 and 22(d), respectively; the ImageNet measurements provide the training-data baseline for both comparisons.

The intensity saturation observed under unguided conditional sampling ($w = 1$) closely matches that of the training dataset. The proportion of images with a saturated area of at least 10% is 7.87% in ImageNet and 8.37% under unguided sampling. The corresponding mean saturated areas are also similar: 1.65% and 1.71% for one-sided saturation, and 0.33% and 0.38% for two-sided saturation, respectively. These results support the interpretation that intensity saturation without guidance reflects the characteristics of the training data.

As the CFG weight increases, both the proportion of saturated images and the mean saturated area increase substantially. At $w = 3.5$, 40.45% of generated images have a saturated area of at least 10%, and the mean saturated areas reach 3.62% for one-sided saturation and 8.10% for two-sided saturation. Thus, strong CFG produces intensity saturation that substantially exceeds the level present in the training dataset, with a particularly pronounced increase in two-sided saturation.



basketball players walking across a court with their coach

A crows of people standing outside of a train



A dark skinned woman posses beneath an umbrella

A desk containing a desktop and laptop computers



a male and female a cake and a candle

A man with a silly look playing a game with remote controllers



A person looking at a train going by

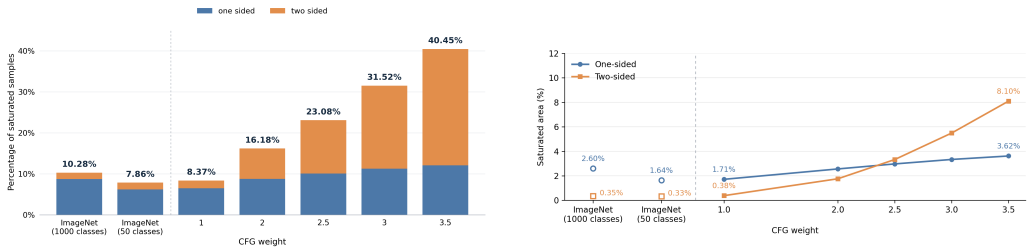
A picture of a cell phone Calculator and a computer



A train with lights on traveling on train tracks

A woman wearing a creepy mask standing next to a child

Figure 35: Qualitative comparison on SD 3.5 Medium. Each pair shows CFG (left) and CFG with our method (right). The prompt used to generate each pair appears below it. These examples illustrate reduced intensity saturation with our method.



(a) Percentage of saturated samples

(b) Mean saturated area

Figure 36: **Intensity saturation in ImageNet and generated images.** ImageNet provides the training-data baseline. (a) Percentage of images with a saturated area of at least 10%, split into one-sided and two-sided saturation. (b) Mean saturated area over all samples for each saturation type. Generated-image results are reproduced from Figures 4 and 22(d), respectively. Unguided sampling closely matches the training-data baseline, while larger CFG weights produce substantially greater intensity saturation.