Where to Inject Noise

Spatially Adaptive Sampling for Diffusion Models

Summer ML Workshop, 17 June 2026

Outline

Motivation

Diffusion Models: Two Sampler Families

The generalized diffusion update rule bridges deterministic and stochastic regimes via global scalar schedules [1]:

$$\mathrm{x}_{t-1} = \underbrace{\sqrt{\bar{\alpha}_{t-1}}\,\hat{\mathrm{x}}_0}_{\text{denoised mean}} + \underbrace{\sqrt{1-\bar{\alpha}_{t-1}-\eta\sigma_t^2}\;\boldsymbol{\epsilon}_\theta}_{\text{direction}} + \underbrace{\eta\sigma_t\,\mathrm{z}}_{\text{noise}} \tag{1}$$

DDIM ($\eta = 0,\ \sigma_t^2 = 0$)

- Deterministic trajectory
- Fast (few steps), but no error correction

DDPM ($\eta = 1,\ \sigma_t^2 = \tilde{\beta}_t$)

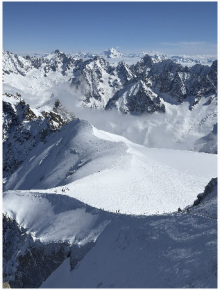
- Stochastic dynamics
- Inject noise every step $\Rightarrow$ corrects errors, but slow.

Key limitation: σ_t^2 is a **scalar** $\Rightarrow$ identical at every pixel.

Central question

At each **pixel** and **timestep**, do we need stochastic correction?

The Spatial Uniformity Problem



Natural images have heterogeneous geometric structures.

The Conflict:

- Smooth regions are "easy" for the model (high confidence) $\rightarrow$ unnecessary noise hurts fidelity.
- Edges/Textures are "hard" for the model (high uncertainty) $\rightarrow$ need stochastic exploration to prevent errors.

Takeaway: Applying the exact same noise level to both regions is suboptimal.

Before going there: Why smooth regions are confident and edges uncertain?

Formalisation: Per-Pixel Geometry

Let p_{data} be the distribution of natural images over $\mathbb{R}^D$.

Sample an image $\mathrm{x} \in \mathbb{R}^D$. Perturb a single pixel i by $\delta \in \mathbb{R}$:

$$\log p_{\mathrm{data}}(\mathrm{x} + \delta\, \mathrm{e}_i) \;=\; \log p_{\mathrm{data}}(\mathrm{x}) \;+\; \delta \underbrace{[\nabla_{\mathrm{x}} \log p_{\mathrm{data}}]_i}_{\text{score at pixel } i} + \frac{\delta^2}{2} \underbrace{[\nabla^2_{\mathrm{x}} \log p_{\mathrm{data}}]_{ii}}_{[\mathrm{H}]_{ii}} + O(\delta^3)$$

$[\mathrm{H}]_{ii} = \mathrm{e}_i^\top \nabla^2_{\mathrm{x}} \log p_{\mathrm{data}}(\mathrm{x})\, \mathrm{e}_i$: curvature of $\log p_{\mathrm{data}}$ in direction $\mathrm{e}_i \in \mathbb{R}^D$.

Confident pixel

$[\mathrm{H}]_{ii} < 0 \rightarrow \log p$ drops

Value tightly constrained by surrounding structure.

Adding noise hurts.

Indifferent pixel (sky, skin)

$[\mathrm{H}]_{ii} \approx 0 \rightarrow \log p$ nearly unchanged

Many nearby values equally plausible.

Noise has little effect.

Ambiguous pixel

(uncertain boundary)

$[\mathrm{H}]_{ii} > 0$ locally convex

Current value unstable $\rightarrow$ alternatives more likely.

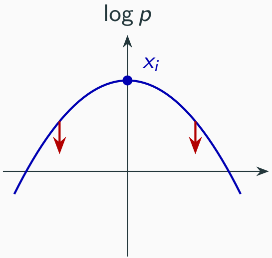
Adding noise helps.

What $[H]_{ii}$ Tells Us: Three Regimes

Cross-section of $\log p_{\text{data}}$ along pixel i, all other pixels fixed (quadratic approximation from previous slide):

$[H]_{ii} \ll 0$

concave

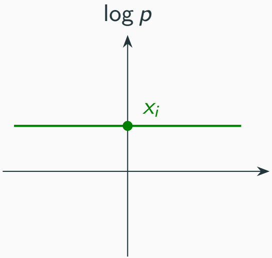


At a **peak**.
Any perturbation decreases $\log p$.

Adding noise hurts.

$[H]_{ii} \approx 0$

flat

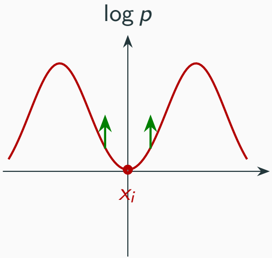


On a **plateau**.
Perturbation barely affects $\log p$.

Noise has little effect.

$[H]_{ii} > 0$

convex



At a **local minimum**.
Any perturbation increases $\log p$.

Adding noise helps.

From Data Geometry to Denoiser distribution

Key insight: The denoiser inherits the heterogeneous geometry of p_{data}

Concept: Instead of choosing between DDIM and DDPM globally, what if we could interpolate them spatially?

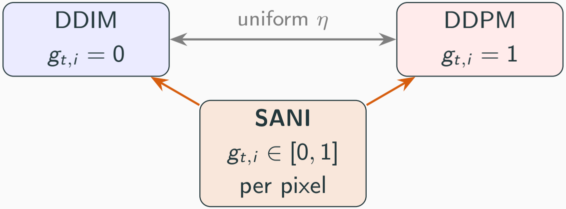
- Confident Pixels $\rightarrow$ Use Deterministic ODE (DDIM-like).
- Uncertain Pixels $\rightarrow$ Use Stochastic SDE (DDPM-like).
- semi-confident Pixels $\rightarrow$ Modulate the injecting noise.

The Challenge: How can we characterize the local geometry and uncertainty of the diffusion process?

To answer this, we turn to information theory to analyze the denoising distribution $p_\theta(\mathrm{x}_0|\mathrm{x}_t)$.

Spatially Adaptive Noise Injection (SANI):

1. Theory: We use the Fisher Information Matrix to quantify per-pixel uncertainty.
2. Method: We design a probabilistic gating function ($g_{t,i} \in [0, 1]$).
3. Result: An inference-time sampling framework that modulates stochasticity pixel-by-pixel.



Theoretical Analysis

Step 1: Exponential Family Structure

Proposition 1. The denoising distribution $p(\mathrm{x}_0 \mid \mathrm{x}_t)$ is an exponential family indexed by x_t:

$$p(\mathrm{x}_0 \mid \mathrm{x}_t) = h(\mathrm{x}_0) \exp\left(\boldsymbol{\eta}(\mathrm{x}_t, t)^\top \mathrm{T}(\mathrm{x}_0) - \psi(\mathrm{x}_t, t)\right)$$

with natural parameters $\boldsymbol{\eta} = \left(\frac{\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}\mathrm{x}_t,\ -\frac{\bar{\alpha}_t}{2(1-\bar{\alpha}_t)}\right)$ and base measure $h(\mathrm{x}_0) = q(\mathrm{x}_0)$. (details: app 4)

Why it matters:

- Posterior mean & covariance are derivatives of log-partition ψ
- Enables Fisher Information Matrix analysis in closed form

Step 2: FIM-Jacobian-Post. Covariance Correspondence

Proposition 2. Let $\hat{x}_0 = \mathbb{E}[x_0 \mid x_t]$ (optimal MMSE denoiser). Then:

$$\boxed{\mathcal{I}(x_t) = \frac{\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}\, J_{\hat{x}_0}(x_t) = \frac{\bar{\alpha}_t}{(1-\bar{\alpha}_t)^2}\, \mathrm{Cov}[x_0 \mid x_t]}$$

where $J_{\hat{x}_0} = \frac{\partial \hat{x}_0}{\partial x_t} \in \mathbb{R}^{D \times D}$. (proof app 5)

Reminder1: FIM is exactly the covariance matrix of the sufficient statistics.
Reminder2: FIM quantifies how sensitive a distribution is to perturbations in its parameters.

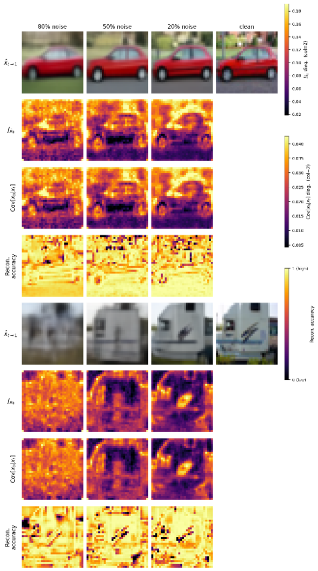
Connection to reconstruction uncertainty:

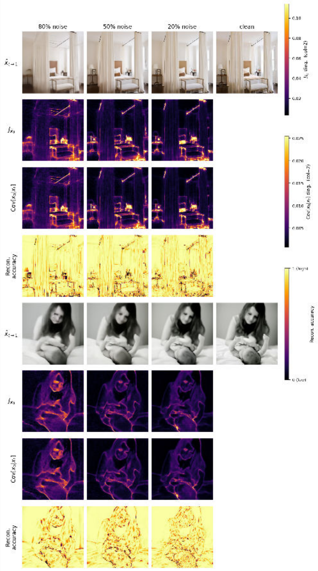
$$\mathrm{Tr}(\mathcal{I}) \;\propto\; \mathrm{MSE}(x_t),$$

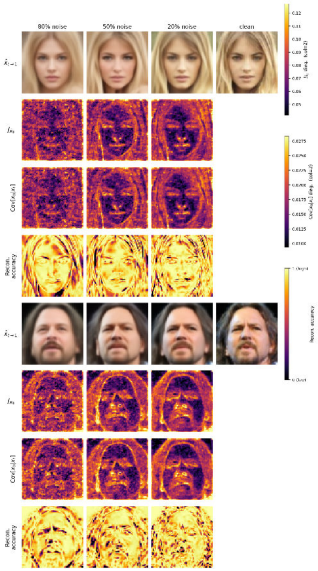
where $\mathrm{MSE}(x_t) := \mathbb{E}[\|x_0 - \hat{x}_0\|^2 \mid x_t] = \mathrm{Tr}(\mathrm{Cov}[x_0 \mid x_t])$ is the minimum mean squared error of the optimal denoiser.

Takeaway: Regions with large Fisher Information (high sensitivity) = regions where the model is **uncertain** $\Rightarrow$ stochastic correction is most valuable there.

Proof of concept







Step 3: Per-Pixel Uncertainty via Score Hessian

Re-express the posterior covariance using the score $s_\theta \approx \nabla_{x_t} \log p_t(x_t)$:

$$\mathrm{Cov}[x_0 \mid x_t] = \frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}\left(I_D + (1-\bar{\alpha}_t)H_t(x_t)\right)$$

where $H_t = \nabla_{x_t} s_\theta$ is the score Hessian.

Extract the diagonal $\Rightarrow$ **per-pixel posterior variance**:

$$\boxed{v_i(x_t, t) = \frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}\left(1 + (1-\bar{\alpha}_t)[H_t]_{ii}\right)}$$

- $[H_t]_{ii} \ll 0$ (log p_t concave in direction i) $\Rightarrow$ confident, v_i small
- $[H_t]_{ii} > 0$ (log p_t convex in direction i) $\Rightarrow$ uncertain, v_i large

Method: SANI

Probabilistic Gating Function

How can we determine, at each pixel and timestep, whether stochastic correction is needed?

We map coordinate variance v_i into a localized gating map $\mathrm{g}_t \in [0,1]^D$, defined as the probability that the local denoising error exceeds a tolerance threshold τ.

Definition. The gate at pixel i, timestep t:

$$g_{t,i} := \mathbb{P}\Big((X_{0,i} - \hat{x}_{0,i})^2 > \tau \;\Big|\; \mathrm{X}_t = \mathrm{x}_t\Big)$$

Proposition 3 (Closed form). Under Gaussian posterior assumption:

$$\boxed{g_{t,i} = 2\left(1 - \Phi\left(\sqrt{\frac{\tau}{v_i(\mathrm{x}_t, t)}}\right)\right)}$$

where Φ is the standard normal CDF. (proof app 6)

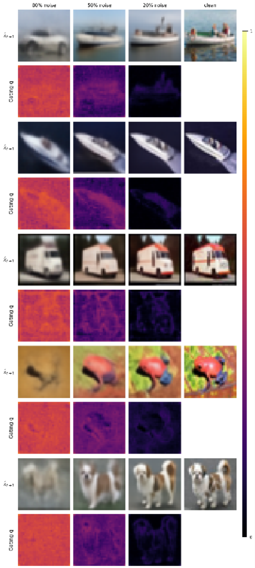
$$v_i \gg \tau \;\Rightarrow\; g_{t,i} \to 1$$

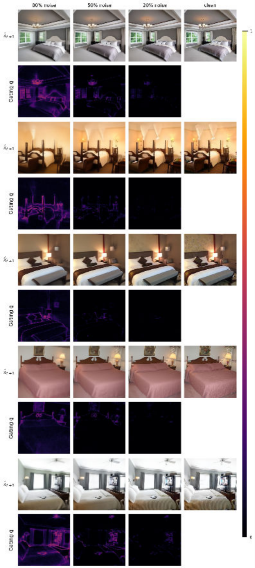
High uncertainty → stochastic exploration

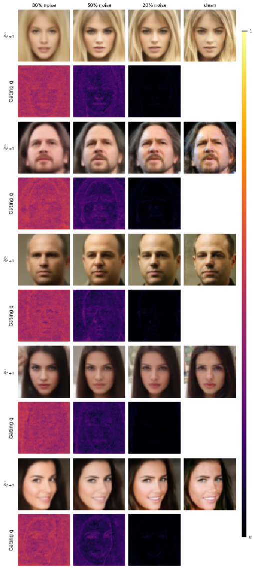
$$v_i \ll \tau \;\Rightarrow\; g_{t,i} \to 0$$

High confidence → deterministic updates

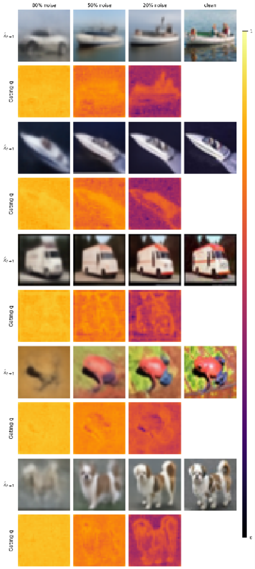
Proof of concept $\tau = 0.1$

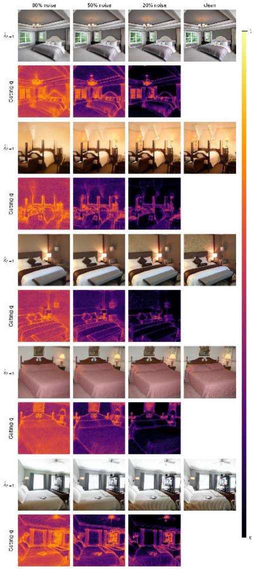


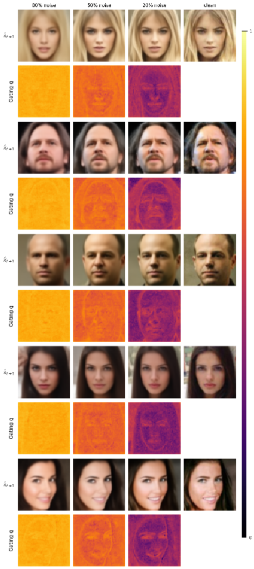




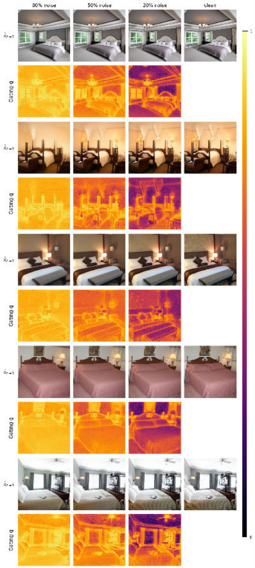
Proof of concept $\tau = 0.01$

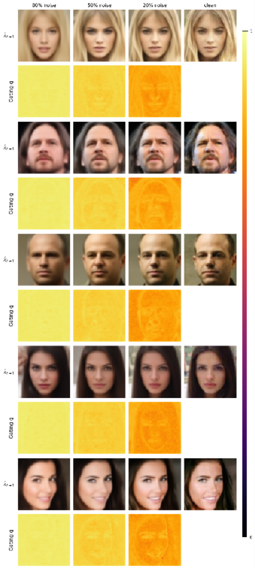




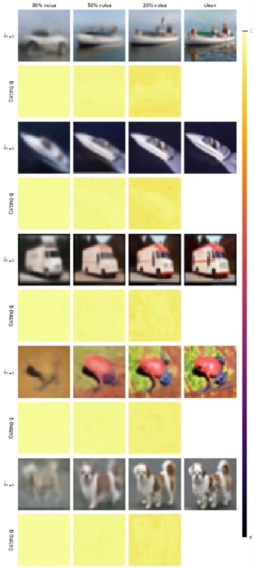


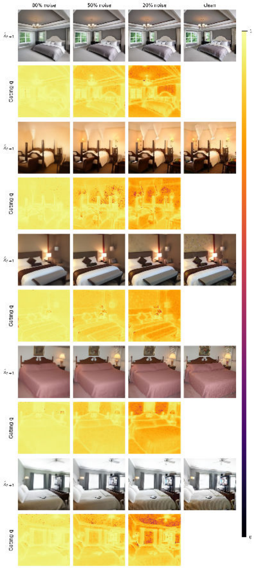
Proof of concept $\tau = 0.001$

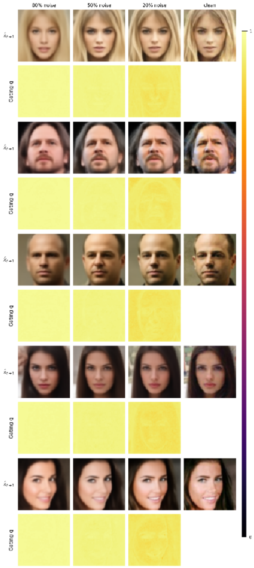




Proof of concept $\tau = 0.0001$







KL-Optimal Per-Pixel Variance Ceiling

Instead of using spatially uniform heuristic, ($\gamma_t \in \{\tilde{\beta}_t, \beta_t\}$), we derive a **per-pixel KL-optimal variance ceiling** $\gamma^*_{t,i}$ by minimizing the expected KL divergence to the true reverse posterior:

Proposition 4 The KL-optimal per-pixel reverse variance is (following [2]):

$$\boxed{\gamma^{*2}_{t,i} = \underbrace{\tilde{\beta}_t}_{\text{irreducible}} + \underbrace{c_t^2\, v_i}_{\text{mean-error}}} \qquad c_t = \frac{\sqrt{\bar{\alpha}_{t-1}}\,\beta_t}{1-\bar{\alpha}_t}$$

Interpretation:

- $\tilde{\beta}_t$: stochasticity present even with a perfect denoiser
- $c_t^2 v_i$: inflates variance where the model's mean estimate is unreliable

Bounded: $\tilde{\beta}_t \le \gamma^*_{t,i} \le \beta_t/\alpha_t$

The SANI update: gate only the exploration

Apply per-pixel KL-optimal variance everywhere but gate **only** the exploratory term:

$$\Sigma_{t,i}^2 = \underbrace{c_t^2\, v_i}_{\text{mean-error correction}} + \underbrace{g_{t,i}\, \tilde{\beta}_t}_{\text{gated exploration}}$$

$$x_{t-1,i} = \sqrt{\bar{\alpha}_{t-1}}\, \hat{x}_{0,i} + \sqrt{1 - \bar{\alpha}_{t-1} - \Sigma_{t,i}^2}\, \epsilon_{\theta,i} + \Sigma_{t,i}\, z_i.$$

Coupled by construction. As $\Sigma_{t,i}$ grows, noise $\uparrow$ *and* the direction coef. $\downarrow$ (less reliance on unreliable trajectory)

Regime	Condition	Reduces to
Confident pixel	$g_{t,i}, v_i \to 0$	**DDIM** (deterministic)
Uncertain pixel	$g_{t,i} \to 1$	KL-optimal stochastic ($\Sigma^2 = \gamma_{t,i}^*$)
Uniform gate $g_{t,i} = \eta,\ v_i = 0$	—	DDIM's η-interpolation, data-aware
Uniform gate $g_{t,i} = 1,\ v_i = 0$	—	DDPM

The SANI sampler

Algorithm 1: SANI Sampling

Input: $x_T \sim \mathcal{N}(0, I)$, model ϵ_θ, gating $\mathcal{G}$, schedule $\{\bar{\alpha}_t, \tilde{\beta}_t\}$, tolerance τ

for $t = T, \ldots, 1$ **do**

$z \sim \mathcal{N}(0, I)$ if $t > 1$, else $z = 0$

$\hat{\epsilon} \leftarrow \epsilon_\theta(x_t, t)$; $\hat{x}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1 - \bar{\alpha}_t}\,\hat{\epsilon})$

$v \leftarrow \text{diag}(\mathrm{Cov}[x_0 \mid x_t])$; $v_i \leftarrow \max(v_i, \epsilon)$

$g_t \leftarrow \mathcal{G}(v, \tau)$

$\boldsymbol{\Sigma}_t^2 = c_t^2\, v + g_t\, \tilde{\beta}_t$

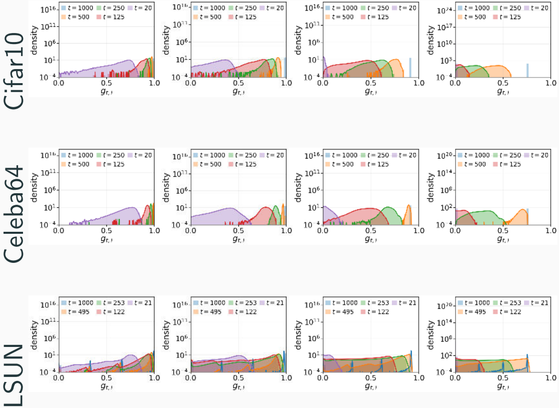
$x_{t-1} \leftarrow \sqrt{\bar{\alpha}_{t-1}}\hat{x}_0 + \sqrt{1 - \bar{\alpha}_{t-1} - \boldsymbol{\Sigma}_t^2} \odot \hat{\epsilon} + \boldsymbol{\Sigma}_t \odot z$

end for

return x_0

Experiments

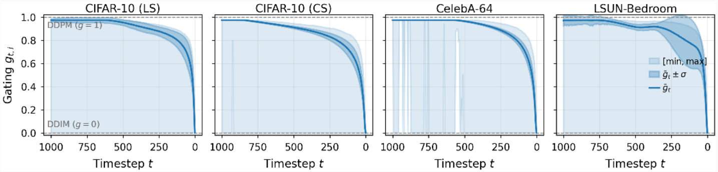
Tolerance Scale τ



(a) $\tau = 10^{-4}$ **(b)** $\tau = 10^{-3}$ **(c)** $\tau = 10^{-2}$ **(d)** $\tau = 10^{-1}$

- Near-point mass at t=1000 (pure noise $\Rightarrow$ uniform uncertainty)
- $g_{t,i} \approx 0$ at the end
- Mid-trajectory: broad support over $[0, 1]$
- Increasing τ slides the stochastic$\to$deterministic transition earlier

Gating Dynamics Over Time



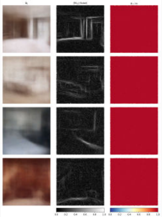
(a) $\tau = 10^{-3}$

1. **Early** (t large): $\bar{g} \approx 1$
   global stochastic exploration
2. **Mid**: $\bar{g}$ decreases
3. **Late** (t small): $\bar{g} \to 0$
   (depending on τ)
   deterministic refinement

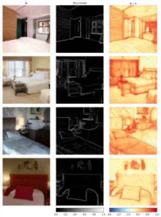
- $\bar{g}_t$ decays **monotonically** $1 \to 0$: stochastic early, deterministic late.
- The band is **widest at intermediate** t: spatial heterogeneity peaks where the gate discriminates.

Gating Maps Reflect Image Geometry

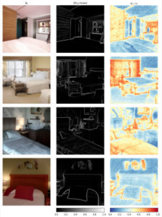
LSUN Bedroom: gating maps $g_{t,i}$ across timesteps:



(a) $t = 800$



(b) $t = 200$



(c) $t = 100$

(d) $t = 25$

$t = 800$
$g \approx 1$ everywhere
(global uncertainty)

↓

$t = 200 \rightarrow 100$
spatially structured
flat regions $\rightarrow$ blue ($g \approx 0$)
edges $\rightarrow$ red ($g \approx 1$)

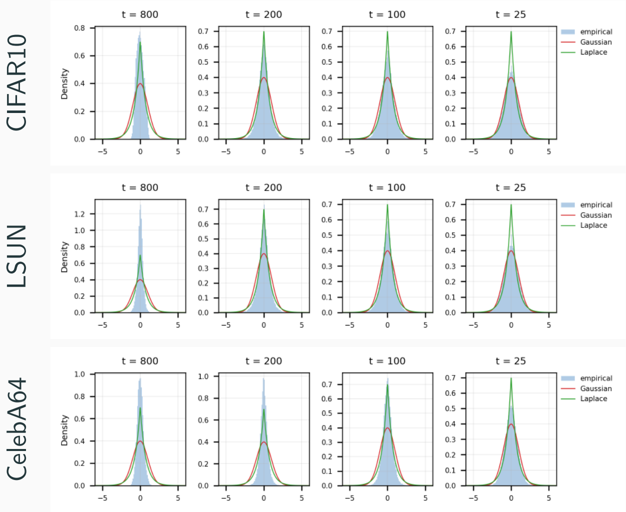
↓

$t = 25$
$g \approx 0$ (edges only)
stochasticity only at boundaries

$K = 50$ $\tau = 0.001$. *Each subplot: predicted image at timestep t (left), Sobel magnitude* $|\nabla \hat{x}_0|$ *(center), gating map (right; blue=0, red=1)*

Emerges purely from the posterior covariance of the pre-trained score network.

Gaussian vs Laplace



- Standardizing the residuals: $Z_i = R_i / \sqrt{v_i}$
- Empirical distribution of Z_i Vs Gaussian, Laplace
- Empirical standardized residuals:
    - Shapely picked at high noise
    - Concentrate toward $\mathcal{N}(0, 1)$ as t decreases

FID scores (↓) across datasets

CELEBA	DDPM 10	DDPM 25	DDPM 50	DDPM 100	DDPM 200	DDIM 10	DDIM 25	DDIM 50	DDIM 100	DDIM 200
Base, $\tilde{\beta}$ [3]	36.69	24.46	18.96	14.31	10.48	20.54	13.54	9.33	6.60	4.96
A- [2]	28.99	16.01	11.23	8.08	6.51	15.62	9.22	6.13	4.29	3.46
NPR- [4]	28.37	15.74	10.89	8.23	7.03	14.98	8.93	6.04	4.27	3.59
SN- [4]	20.60	12.00	7.88	5.89	5.02	10.20	5.48	3.83	3.04	2.85
OCM- [5]	21.55	12.71	9.24	6.97	5.92	10.28	5.72	4.42	3.54	3.17

SANI^G (ours)	K=10: 17.09	K=25: **8.87**	K=50: **5.56**	K=100: 4.64	K=200: **3.04**
SANI^L (ours)	K=10: **16.07**	K=25: 9.84	K=50: 6.25	K=100: **3.93**	K=200: 4.52

CIFAR10 (CS)	DDPM 10	DDPM 25	DDPM 50	DDPM 100	DDPM 200	DDIM 10	DDIM 25	DDIM 50	DDIM 100	DDIM 200
Base, $\tilde{\beta}$ [3]	34.76	16.18	11.11	8.38	6.66	34.34	16.68	10.48	7.94	6.69
A- [2]	22.94	8.50	5.50	4.45	4.04	26.43	9.96	6.02	4.88	4.92
NPR- [4]	19.94	7.99	5.31	4.52	4.10	22.81	9.47	6.04	5.02	5.06
SN- [4]	16.33	6.05	4.17	**3.83**	3.72	17.90	7.36	5.16	4.63	4.63
OCM- [5]	**14.32**	**5.54**	**4.10**	3.84	3.75	16.70	6.71	4.72	4.30	4.54

SANI^G (ours)	K=10: 17.01	K=25: 7.35	K=50: 4.95	K=100: **3.83**	K=200: 4.17
SANI^L (ours)	K=10: 16.37	K=25: 6.55	K=50: 4.88	K=100: 4.50	K=200: **3.52**

Conclusion & Future Work

Summary

1. **Theory:** FIM $\propto$ denoiser Jacobian $\propto$ posterior covariance $\Rightarrow$ per-pixel uncertainty from the score Hessian diagonal
2. **Gating:** closed-form $g_{t,i} = 2(1 - \Phi(\sqrt{\tau / v_i}))$: tail probability of denoising error
3. **SANI update:** coupled modulation of noise variance *and* direction coefficient per pixel
   - Stochastic correction where the model is uncertain
   - Deterministic efficiency where it is confident
4. **Results:** Consistent FID gains over DDIM and DDPM-family samplers in the low-step regime

Limitations & future directions:

- Fixed τ becomes suboptimal (?) at high $K \overset{?}{\to}$ dynamic tolerance scheduling $\tau(K)$
- Diagonal Hessian approximation ignores inter-pixel correlations $\to$ frequency-domain block-diagonal covariance [6]
- Potential integration with early-stopping for further acceleration [7]

Thank you! Questions?

References

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

[2] Fan Bao, Chongxuan Li, Jun Zhu, and Bo Zhang. Analytic-dpm: an analytic estimate of the optimal reverse variance in diffusion probabilistic models. In *International Conference on Learning Representations (ICLR)*, 2022. URL `https://openreview.net/forum?id=1y5M8mE4v49`.

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

[4] Fan Bao, Chongxuan Li, Jun Sun, Jiachen Shao, Ruixuan Yan, Binyuan Chen, and Hang Liu. Estimating the optimal covariance with imperfect mean in diffusion probabilistic models. In *International Conference on Machine Learning*, pages 1555–1584. PMLR, 2022. URL `https://proceedings.mlr.press/v162/bao22d.html`.

[5] Zijing Ou, Mingtian Zhang, Andi Zhang, Tim Z. Xiao, Yingzhen Li, and David Barber. Improving probabilistic diffusion models with optimal diagonal covariance matching. In *The Thirteenth International Conference on Learning Representations*, 2025. URL `https://openreview.net/forum?id=fV0t65OBUu`.

[6] Andrea Schioppa and Tim Salimans. Covariance-aware sampling for diffusion models. *arXiv preprint arXiv:2605.13910*, 2026.

[7] Johannes Schusterbauer, Ming Gui, Yusong Li, Pingchuan Ma, Felix Krause, and Björn Ommer. Denoising, fast and slow: Difficulty-aware adaptive sampling for image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Appendix / Backup Slides

Baselines-Datasets

Baselines:

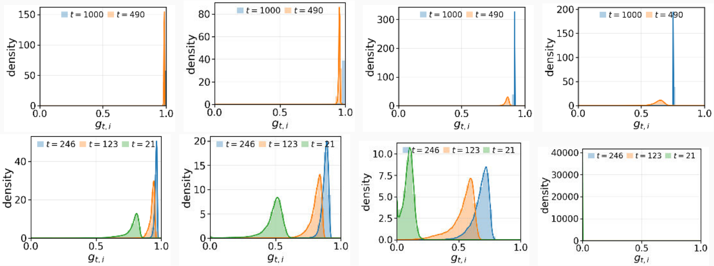
1. DDPM with $\sigma_t^2 = \tilde{\beta}_t$ and $\sigma_t^2 = \beta_t$ [3]
2. DDIM [1]
3. Analytic-DPM (A-DDPM/DDIM) [2]
4. NPR DDPM/DDIM [4] (models the noise prediction residual)
5. SN-DDPM/DDIM [4] (models the second moment of the noise)
6. OCM-DDPM/DDIM [5]

Datasets tested:

Dataset	Resolution
CIFAR10 (LS)	32×32
CIFAR10 (CS)	32×32
CelebA	64×64
InageNet	64×64
LSUN Bedroom	256×256

Tolerance Scale τ no log scale

τ controls gating 'aggressiveness' via $\sqrt{\tau / v_i}$ in the CDF argument.



(a) $\tau = 10^{-4}$ (b) $\tau = 10^{-3}$ (c) $\tau = 10^{-2}$ (d) $\tau = 10^{-1}$

- $\tau \leq 10^{-4}$: **Degenerate stochastic** $g_{t,i} \approx 1$, collapses to DDPM
- $\tau \geq 10^{-1}$: **Degenerate deterministic** $g_{t,i} \approx 0$, collapses to DDIM
- $\tau \in [10^{-3}, 10^{-2}]$: Spatially adaptive meaningful discrimination between confident and uncertain pixels

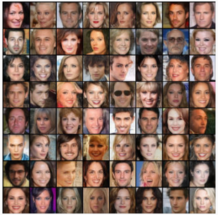
FID scores (↓) across datasets

	DDPM					DDIM				
CELEBA	10	25	50	100	200	10	25	50	100	200
Base, $\tilde{\beta}$	36.69	24.46	18.96	14.31	10.48	20.54	13.54	9.33	6.60	4.96
A-	28.99	16.01	11.23	8.08	6.51	15.62	9.22	6.13	4.29	3.46
NPR-	28.37	15.74	10.89	8.23	7.03	14.98	8.93	6.04	4.27	3.59
SN-	20.60	12.00	7.88	5.89	5.02	10.20	5.48	3.83	3.04	2.85
OCM-	21.55	12.71	9.24	6.97	5.92	10.28	5.72	4.42	3.54	3.17
SANI^G (ours)	K=10: 17.09		K=25: 8.87		K=50: 5.56		K=100: 4.64		K=200: 3.04	
SANI^L (ours)	K=10: 16.07		K=25: 9.84		K=50: 6.25		K=100: 3.93		K=200: 4.52	
CIFAR10 (LS)	10	25	50	100	200	10	25	50	100	200
Base, $\tilde{\beta}$	44.45	21.83	15.21	10.94	8.23	21.31	10.70	7.74	6.08	5.07
A-	34.26	11.60	7.25	5.40	4.01	14.00	5.81	4.04	3.55	3.39
13.34	5.38	3.95	3.53	3.42						
SN-	24.06	6.91	4.63	3.67	3.11	12.19	4.28	3.39	3.23	3.22
OCM-	24.94	9.19	5.95	4.36	3.48	10.66	4.35	3.48	3.27	3.29
SANI^G (ours)	K=10: 15.27		K=25: 6.64		K=50: 6.50		K=100: 3.82		K=200: 3.99	
SANI^L (ours)	K=10: 17.96		K=25: 8.13		K=50: 5.65		K=100: 4.85		K=200: 5.85	
CIFAR10 (CS)	10	25	50	100	200	10	25	50	100	200
Base, $\tilde{\beta}$	34.76	16.18	11.11	8.38	6.66	34.34	16.68	10.48	7.94	6.69
A-	22.94	8.50	5.50	4.45	4.04	26.43	9.96	6.02	4.88	4.92
NPR-	19.94	7.99	5.31	4.52	4.10	22.81	9.47	6.04	5.02	5.06
SN-	16.33	6.05	4.17	3.83	3.72	17.90	7.36	5.16	4.63	4.63
OCM-	14.32	5.54	4.10	3.84	3.75	16.70	6.71	4.72	4.30	4.54
SANI^G (ours)	K=10: 17.01		K=25: 7.35		K=50: 4.95		K=100: 3.83		K=200: 4.17	
SANI^L (ours)	K=10: 16.37		K=25: 6.55		K=50: 4.88		K=100: 4.50		K=200: 3.52	

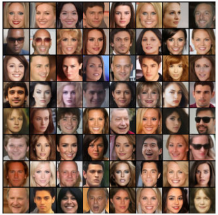
NLL (↓) across datasets

	CIFAR10 (LS)					CelebA 64x64					CIFAR10 (CS)				
# timesteps K	10	25	50	100	200	10	25	50	100	200	10	25	50	100	200
DDPM, $\tilde{\beta}$	74.95	24.98	12.01	7.08	5.03	33.42	13.09	7.14	4.60	3.45	75.96	24.94	11.96	7.04	4.95
DDPM, β	6.99	6.11	5.44	4.86	4.39	6.67	5.72	4.98	4.31	3.74	6.51	5.55	4.92	4.41	4.03
A-DDPM	5.47	4.79	4.38	4.07	3.84	4.54	3.89	3.48	3.16	2.92	5.08	4.45	4.09	3.83	3.64
NPR-DDPM	5.40	4.64	**4.25**	3.98	3.79	**4.46**	**3.78**	**3.40**	**3.11**	**2.89**	5.03	**4.33**	**3.99**	**3.76**	**3.59**
SN-DDPM	30.79	11.83	7.13	5.24	4.39	18.09	8.05	5.29	4.05	3.40	90.85	19.81	9.72	6.72	5.58
OCM-DDPM	5.32	**4.63**	**4.25**	**3.97**	**3.78**	4.69	3.86	3.43	3.13	2.90	**4.99**	4.34	**3.99**	**3.76**	**3.59**
SANI^G	**5.31**	4.67	4.31	4.04	3.85	4.67	3.88	3.45	3.15	2.95	**3.67**	4.39	4.07	3.83	3.66
SANI^L	5.34	4.68	4.34	4.06	3.87	4.71	3.87	3.48	3.15	2.97	5.04	4.40	4.08	3.82	3.68

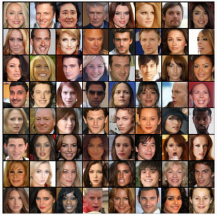
Generated Samples: CelebA



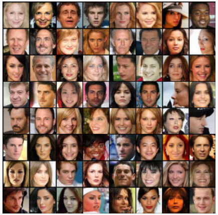
(a) $K = 10$



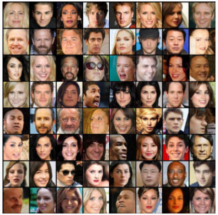
(b) $K = 25$



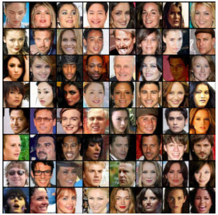
(c) $K = 50$



(d) $K = 100$



(e) $K = 200$

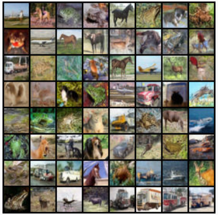


(f) $K = 1000$

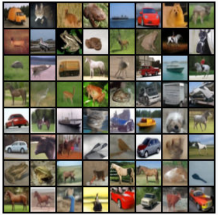
Generated Samples: LSUN bedroom

(a) $K = 25$ **(b)** $K = 50$ **(c)** $K = 100$ **(d)** $K = 200$ **(e)** $K = 1000$

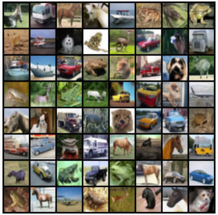
Generated samples: CIfar10 (LS)



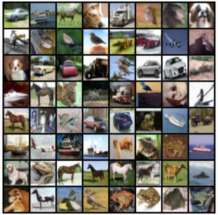
(a) $K = 10$



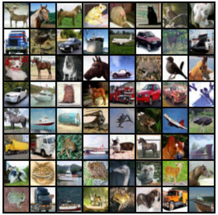
(b) $K = 25$



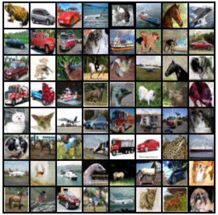
(c) $K = 50$



(d) $K = 100$

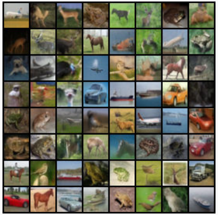


(e) $K = 200$

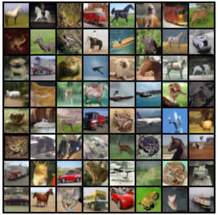


(f) $K = 1000$

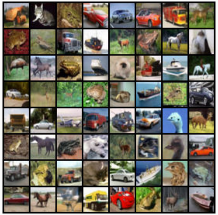
Generated samples: CIfar10 (CS)



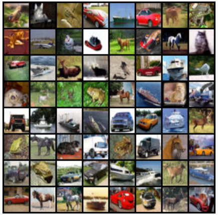
(a) $K = 10$



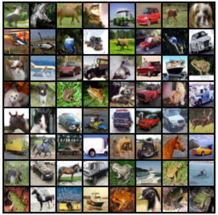
(b) $K = 25$



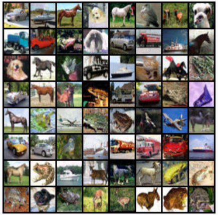
(c) $K = 50$



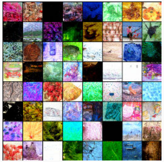
(d) $K = 100$



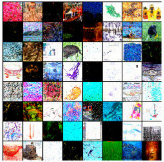
(e) $K = 200$



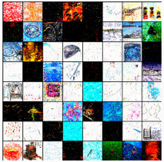
(f) $K = 1000$



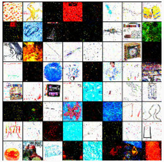
(a) $K = 10$ $\tau = 0.001$



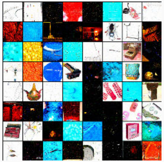
(b) $K = 25$ $\tau = 0.001$



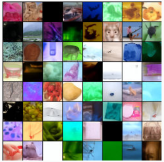
(c) $K = 50$ $\tau = 0.001$



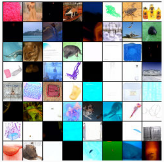
(d) $K = 100$ $\tau = 0.001$



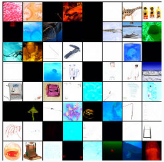
(e) $K = 1000$ $\tau = 0.001$



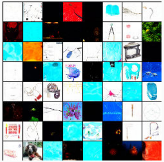
(f) $K = 10$ $\tau = 1e-05$



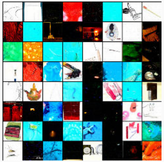
(g) $K = 25$ $\tau = 1e-05$



(h) $K = 50$ $\tau = 1e-05$



(i) $K = 100$ $\tau = 1e-05$



(j) $K = 1000$ $\tau = 1e-05$