Squeezed Diffusion Models

Jyotirmai Singh
Stanford University
joesingh@stanford.edu

Samar Khanna
Stanford University
samarkhanna@cs.stanford.edu

James Burgess
Stanford University
jmhb@stanford.edu

Abstract

Diffusion models typically inject *isotropic* Gaussian noise, disregarding structure in the data. Motivated by the way *quantum squeezed states* redistribute uncertainty according to the Heisenberg uncertainty principle, we introduce **Squeezed Diffusion Models (SDM)**, which scale noise anisotropically along the principal component of the training distribution. As squeezing enhances the signal-to-noise ratio in physics, we hypothesize that scaling noise in a data-dependent manner can better assist diffusion models in learning important data features. We study two configurations: (i) a *Heisenberg diffusion model* that compensates the scaling on the principal axis with inverse scaling on orthogonal directions and (ii) a *standard* SDM variant that scales only the principal axis. Counterintuitively, on CIFAR-10/100 and CelebA-64, mild *antisqueezing* – i.e. increasing variance on the principal axis – consistently improves FID by up to 15% and shifts the precision–recall frontier toward higher recall. Our results demonstrate that simple, data-aware noise shaping can deliver robust generative gains without architectural changes.

1 Introduction

Score-based diffusion models have become a standard tool for high-fidelity image generation, achieving state-of-the-art performance on datasets from CIFAR-10 to ImageNet [1, 2]. While diffusion models typically rely on adding standard isotropic Gaussian noise during the training process, there is evidence that this standard "one-size-fits-all" approach can degrade generative quality in certain contexts like in natural images which have an anisotropic power distribution in frequency space [3]. This has motivated the investigation of methods to engineer the noise used during the diffusion process in a data-dependent way to overcome these shortcomings [4, 5].

The field of quantum metrology [6], has developed an extensive array of techniques to manipulate noise in the context of precision measurement. One such technique is quantum squeezing [7], which redistributes noise so that variance is reduced along a "measurement" axis of interest, boosting signal-to-noise ratio (SNR) and enabling weak-signal extraction. Inspired by this, we introduce Squeezed Diffusion Models (SDM), which anisotropically scale noise along the dataset's principal component. We hypothesize this encourages the model to learn more semantically meaningful patterns during training, improving sample quality in generation.

To test this, we investigated two model variants: (1) a *Heisenberg diffusion model* that scales the principal direction and counter-scales the orthogonal ones and (2) a "standard" SDM that only scales the principal direction. Surprisingly, we observed that best performance came by mildly *antisqueezing* along the principal axis – i.e. injecting noise and degrading SNR – across multiple datasets.

Background Quantum squeezing exploits the Heisenberg uncertainty principle:

$$\Delta A \cdot \Delta B \geq \frac{1}{2} |\langle [A, B] \rangle| \tag{1}$$
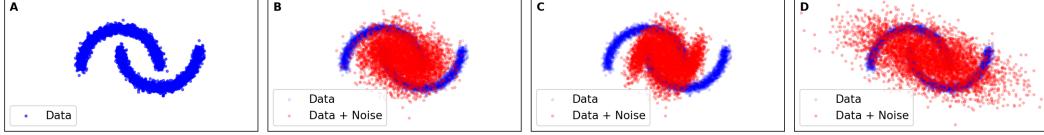
Figure 1: Diffusion with squeezed noise. (a) the raw data distribution. (b) standard isotropic noising process. (c)–(d) anisotropic noising process with (c) squeezing and (d) antisqueezing along the principal direction.

by reducing variance in one physical variable ΔA while enlarging it in the conjugate one ΔB but keeping the product constant. Squeezed states boost signal-to-noise along a chosen axis and underpin precision measurements such as gravitational-wave [8] and dark matter [9] detectors.

Figure 1 illustrates how this effect can be applied to diffusion models. In fig. 1(c) the noise (red) is squeezed along the first principal component and the original distribution (blue) is still somewhat intact due to higher SNR relative to isotropic noise (fig. 1(b)). In fig. 1(d), the noise is instead *antisqueezed* or enhanced along the same direction, reducing SNR compared with the isotropic case.

Related Work   Recent work has showed that alternative noise schedules can impact generative performance [10]. Improvements have been demonstrated with for example the cosine schedule [2] and variationally optimized schedules [11]. Other work has focused on changing noise structure in the frequency domain by e.g. equalizing or biasing certain frequencies [3, 4, 12]. Closer to our aim, other work has demonstrated benefits by learning a multivariate data-dependent noise process [5]. Our approach differs by using a simple, physics-inspired PCA alignment to apply a single-parameter anisotropic scaling.

2   Method: Squeezed Diffusion

As in a standard DDPM, an SDM corrupts a clean datapoint x_0 with additive Gaussian noise over a schedule $\{\beta_t\}_{t=1}^T \subset (0, 1)$. In an SDM however, at each step we *anisotropically* scale the noise with a *squeeze matrix* $S_t(s)$ whose strength is controlled by a single hyper-parameter s and which may also vary with t:

$$x_t^{\mathrm{sq}} = \sqrt{\alpha_t}\, x_{t-1}^{\mathrm{sq}} + \sqrt{1-\alpha_t}\, S_t(s)\, \varepsilon_t, \qquad \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I), \qquad \alpha_t = 1 - \beta_t. \tag{2}$$

Marginal distribution   By unrolling eq. (2) we obtain

$$x_t^{\mathrm{sq}} = \sqrt{\bar{\alpha}_t}\, x_0 + \sum_{i=1}^{t}\Big(\sqrt{1-\alpha_i}\, \prod_{j=i+1}^{t} \sqrt{\alpha_j}\Big) S_i(s)\, \varepsilon_i, \qquad \bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i, \tag{3}$$

As in a standard DDPM, we have that $q(x_t|x_0) \sim \mathcal{N}\left(\sqrt{\bar{\alpha}_t} x_0, \Sigma_t\right)$ but instead now the covariance is modified to be $\Sigma_t = \sum_{i=1}^{t}\left((1-\alpha_i) \prod_{j=i+1}^{t} \alpha_j\right) S_i S_i^\top$.

Training objective   The network is trained to predict the *squeezed* noise $\varepsilon_t^{\mathrm{sq}} = S_t(s)\varepsilon_t$ via the standard MSE:

$$\mathcal{L}_{\mathrm{SDM}} = \mathbb{E}_{x_0, t, \varepsilon_t}\left\| \varepsilon_t^{\mathrm{sq}} - \hat{\varepsilon}_\theta^{\mathrm{sq}}(x_t^{\mathrm{sq}}, t)\right\|_2^2. \tag{4}$$

Reverse step (whiten–denoise–resqueeze)   For the reverse step we first *whiten* each state:

$$\tilde{x}_t = S_t^{-1} x_t^{\mathrm{sq}}, \qquad \tilde{\varepsilon}_\theta = S_t^{-1} \hat{\varepsilon}_\theta^{\mathrm{sq}}. \tag{5}$$

With this change of variable, eq. (2) becomes *approximately* equal to a standard DDPM in $\tilde{x}_t$, which allows us to reuse the Ho et al. [1] posterior:

$$\tilde{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\Big(\tilde{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\, \tilde{\varepsilon}_\theta(\tilde{x}_t, t)\Big) + \sqrt{\tilde{\beta}_t}\, z, \qquad z \sim \mathcal{N}(0, I), \tag{6}$$

where $\tilde{\beta}_t = (1 - \bar{\alpha}_{t-1})/(1 - \bar{\alpha}_t)\, \beta_t$. Finally we *re-squeeze* with the *next* matrix S_{t-1} to stay in squeezed coordinates:

$$x_{t-1}^{\mathrm{sq}} = S_{t-1}(s)\, \tilde{x}_{t-1}. \tag{7}$$

2

Note that with the whitening, eq. (2) becomes

$$\tilde{x}_t = \sqrt{\alpha_t} S_t^{-1} S_{t-1} \tilde{x}_{t-1} + \sqrt{1 - \alpha_t} \varepsilon_t \tag{8}$$

When S_t is time-independent, equation eq. (8) is an exact DDPM because $S_t = S_{t-1}$ and eq. (6) is the exact posterior. For time-dependent squeezing this is only an approximate DDPM with some drift $R_t = S_t^{-1} S_{t-1}$. However, for the form of squeezing matrices we define and our linear noise schedule, this drift ends up being negligible so the posterior approximation does not impact generative performance. More details are in appendix B.

Squeeze matrix $S_t(s)$ To choose the squeezing direction, we do a PCA decomposition of the data in RGB space at the individual pixel level. For large natural image sets, this aligns with the opponent-colour axes luminance, red–green, and blue–yellow [13, 14]. Let $\hat{v}$ be the unit vector in the direction of the first principal component, which for RGB images is the luminance. We can define the squeeze matrices S_t^{HDM} for the Heisenberg diffusion model and S_t^{SDM} for the standard squeezed diffusion model:

$$S_t^{\mathrm{HDM}}(s) = e^{-s}\, \hat{v}\hat{v}^\top + e^{\frac{-s}{n-1}}\big(I - \hat{v}\hat{v}^\top\big), \qquad S_t^{\mathrm{SDM}}(s) \quad = I + \hat{v}\hat{v}^\top\big(e^{-s} - 1\big). \tag{9}$$

Here s is a hyperparameter corresponding to the squeezing strength and n is the dimension of the vector space on which squeezing acts. For the case of RGB pixels, $n = 3$. Note that S_t^{SDM} is derived from S_t^{HDM} by just discarding the antisqueezing in the subspace orthogonal to $\hat{v}\hat{v}^\top$. These matrices mirror the exponential form of electromagnetic squeezing operators in quantum mechanics [15]:

$$S(z) = \exp\left(\frac{z^* \hat{a}^2 - z(\hat{a}^\dagger)^2}{2}\right) \tag{10}$$

For $z = re^{i\theta}$, $S(z)$ scales the noise by e^{-r} in the θ direction in the phase space determined by the electromagnetic creation and annihilation operators $\hat{a}^\dagger$, $\hat{a}$, just as the S_t do so by e^{-s} in the direction determined by $\hat{v}$. Note also that with this parametrisation, there is a smooth, continuous path between an SDM and an isotropic DDPM (achieved at $s = 0$). This mirrors the way squeezing of vacuum noise is inherently continuous, and it situates SDM as a principled middle ground between isotropic, data-unaware noise and fully data-aware approaches such as covariance-matching diffusion [5].

Finally, to better align the squeezing process with the noise schedule of the diffusion model, which in our experiments was linear, we made the squeezing strength s vary linearly as a function of time $s(t) = s_0 \frac{\beta_t}{\beta_{\max}}$ which makes the squeezing matrix S_t time dependent.

3   Results

Our experiments consist of ablation studies to evaluate squeezed diffusion over a variety of standard image datasets. All experiments were run on a single NVIDIA A100 (40GB) GPU with mixed precision (fp16) via accelerate. The results are shown in fig. 2. On CIFAR-10 [16], FID-10K shows a consistent U-shaped dependence on squeeze strength (fig. 2a) across multiple random seeds. The average baseline DDPM FID was 39.8 while the SDM reached an FID of 33.6 at $s \approx -0.4$ – a $\sim$15% improvement. Notably, this is in the *antisqueezing* regime with degraded SNR.

Next, with a fixed seed, we compare the standard squeezed diffusion model against the uncertainty preserving Heisenberg variant on CIFAR-10. Both the FID-10K (fig. 2b) and Inception Score (fig. 2c) follow the same U-shaped trend in s, with optima near $s \approx -0.4$ (FID) and $s \approx -0.3$ (IS). SDM outperforms HDM in the mild antisqueezing regime.

To check the robustness of squeezing, we evaluated SDM generative performance on CIFAR-100 [16] in fig. 2d and CelebA-64 [17] in fig. 2e. The parabolic shape is reproduced across datasets, with the optimum achieved in the mild antisqueezing regime again. The exact location of the optimum varies across datasets, with best performance on CIFAR-100 at $s \sim -0.3$ and on CelebA-64 at $s \sim -0.2$.

To gain further insight into the impact of squeezing vs antisqueezing, we analyze the precision and recall metrics of our models on CIFAR-10 [18–20]. Figure 2f shows that squeezing the SDM noise corresponds to reducing recall while getting marginal gains in precision. By contrast antisqueezing the noise produces large gains in recall at minor cost in precision up to a point. Overall generative quality, as measured by F-score, once again improves up to a point showing parabolic behavior.

(a) SDM FID vs. s on CIFAR-10 (b) SDM vs. HDM FID CIFAR-10 (c) SDM vs. HDM IS CIFAR-10

(d) SDM FID vs. s on CIFAR-100 (e) SDM FID vs. s on CelebA-64 (f) SDM P-R on CIFAR-10
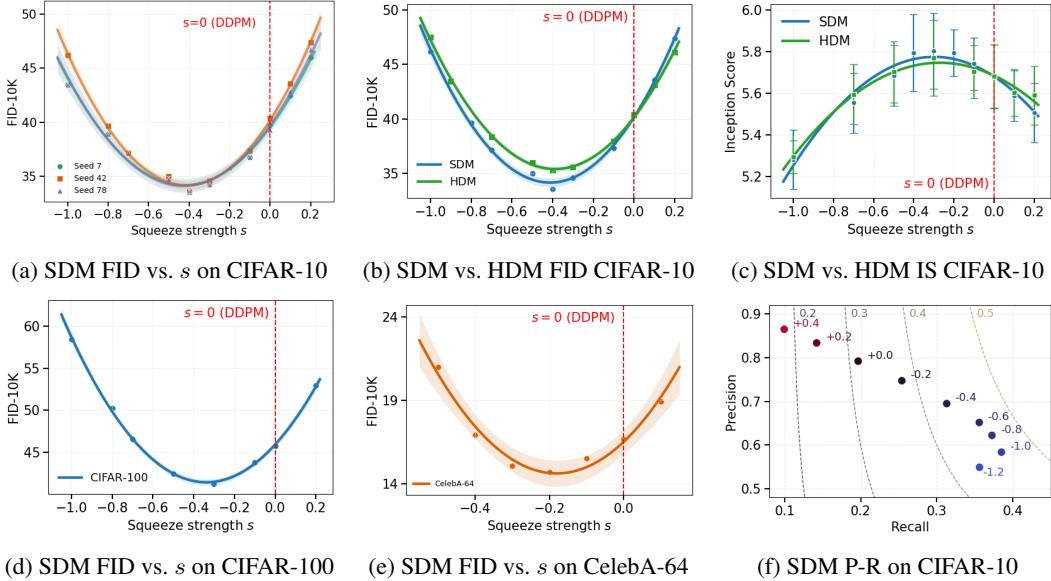
Figure 2: Overview of results across datasets and metrics. (a)–(c) CIFAR-10 FID and IS comparisons for SDM vs. HDM; (d)–(e) cross-dataset FID ablations on CIFAR-100 and CelebA-64; (f) precision–recall analysis on CIFAR-10, with squeeze strength annotated on points and contours of constant F-score.

(a) $s = 0.0$ (baseline) (b) $s = -0.4$ (optimal antisqueezing)
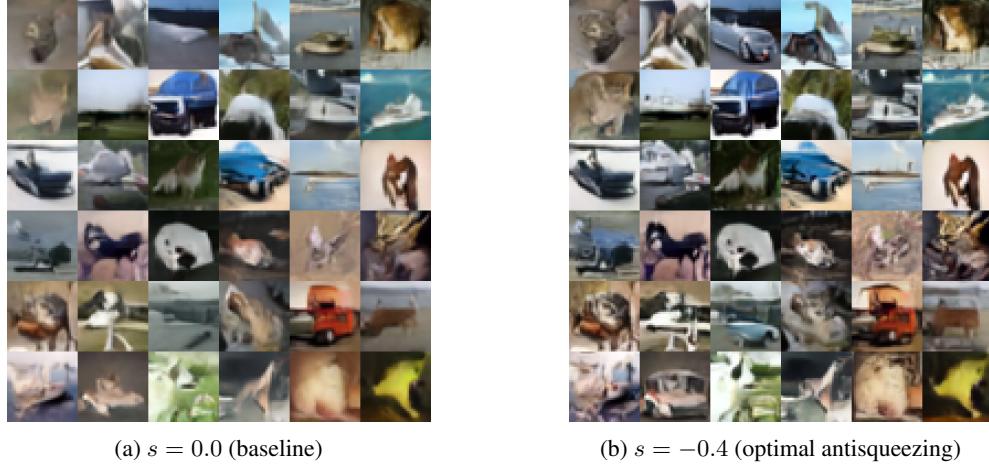
Figure 3: **Generated CIFAR-10 samples for a standard DDPM and an SDM with fixed seed.** (a) standard DDPM; (b) antisqueezed SDM at $s = -0.4$.

Qualitative Samples In fig. 3 we compare 36 samples generated with the same noise seed for the baseline DDPM and the SDM at the optimal antisqueezing $s = -0.4$. The SDM images show crisper object contours: row 1 column 3 becomes a clear car instead of a grey blur, and row 2 column 6 reveals sharper boat edges and clouds. These sharper textures are consistent with the $\sim 15\%$ FID drop at $s = -0.4$ seen in 2a. Occasional higher frequency ringing artifacts remain (e.g. car in row 2 column 3), echoing the precision loss seen in the PR analysis as antisqueezing is increased.

4 Conclusion and Future Work

Across three image datasets we observe a stable U-shaped dependence of sample quality on the squeeze strength: modest antisqueezing enlarges variance along luminance, boosting recall at little precision cost, whereas strong squeezing or antisqueezing harms both. The standard SDM variant matches or outperforms the Heisenberg counterpart, underscoring that diffusion models need not strictly conserve uncertainty to improve generative performance. Future work could include scaling

to higher resolution datasets and exploring frequency-dependent or modality-specific squeezing such as audio. Overall, SDM offers a principled method to manipulate noise that delivers robust gains in generative quality.

References

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

[2] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*, 2021.

[3] Fabian Falck, Teodora Pandeva, Kiarash Zahirnia, Rachel Lawrence, Richard Turner, Edward Meeds, Javier Zazo, and Sushrut Karmalkar. A fourier space perspective on diffusion models. *arXiv preprint arXiv:2505.11278*, 2025.

[4] Xingchang Huang, Corentin Salaun, Cristina Vasconcelos, Christian Theobalt, Cengiz Oztireli, and Gurprit Singh. Blue noise for diffusion models. In *SIGGRAPH: Computer Graphics and Interactive Techniques*, 2024.

[5] Subham Sekhar Sahoo, Aaron Gokaslan, Chris De Sa, and Volodymyr Kuleshov. Diffusion models with learned adaptive noise. In *Advances in Neural Information Processing Systems*, 2024.

[6] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf. Introduction to quantum noise, measurement and amplification. *Review of Modern Physics*, 2008.

[7] D. F. Walls. Squeezed states of light. *Nature*, 1983.

[8] D. Ganapathy et al. Broadband quantum enhancement of the ligo detectors with frequency-dependent squeezing. *Phys. Rev. X*, 13:041021, Oct 2023. doi: 10.1103/PhysRevX.13.041021. URL https://link.aps.org/doi/10.1103/PhysRevX.13.041021.

[9] K.M. Backes et al. A quantum enhanced search for dark matter axions. *Nature 590, 238–242*, 2021.

[10] Ting Chen. On the importance of noise scheduling for diffusion models, 2023. URL https://arxiv.org/abs/2301.10972.

[11] Diederik P. Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, 2021.

[12] Anh Tran Hao Phung, Quan Dao. Wavelet diffusion models are fast and scalable image generators. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[13] Gershon Buchsbaum and Allan Gottschalk. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal society of London. Series B. Biological sciences*, 220(1218):89–113, 1983.

[14] Daniel L. Ruderman, Thomas W. Cronin, and Chuan-Chin Chiao. Statistics of cone responses to natural images: implications for visual coding. *J. Opt. Soc. Am. A*, 15(8):2036–2045, Aug 1998. doi: 10.1364/JOSAA.15.002036. URL https://opg.optica.org/josaa/abstract.cfm?URI=josaa-15-8-2036.

[15] Christopher Gerry and Peter L. Knight. *Introductory Quantum Optics*. Cambridge University Press, 2005.

[16] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[17] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3730–3738, 2015.

[18] Mehdi S. M. Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing generative models via precision and recall. In *International Conference on Learning Representations (ICLR)*, 2018.

[19] Teemu Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[20] Anton Obukhov, Maximilian Seitzer, Po-Wei Wu, Semen Zhydenko, Jonathan Kyl, and Elvis Yu-Jing Lin. High-fidelity performance metrics for generative models in pytorch, 2020. URL `https://github.com/toshas/torch-fidelity`. Version: 0.3.0, DOI: 10.5281/zenodo.4957738.

[21] Patrick von Platen, Suraj Patil, Anton Lozhkov, Pedro Cuenca, Nathan Lambert, Kashif Rasul, Mishig Davaadorj, Dhruv Nair, Sayak Paul, Steven Liu, William Berman, Yiyi Xu, and Thomas Wolf. Diffusers: State-of-the-art diffusion models. URL `https://github.com/huggingface/diffusers`.

A   Details of Model Training

A.1   Implementation

We use the HuggingFace `diffusers` library [21] with a `UNet2DModel` backbone. To apply squeezing, we keep the UNet architecture intact and subclass the `DDPMScheduler` class overriding the `add_noise` and `step` methods. All other training and sampling code remains standard. The model uses a linear noise schedule. Training is done with 1000 timesteps while sampling is done using 50. We also used EMA with a value of EMA decay set to 0.9999. Code for these experiments is available on Github at `https://github.com/joe-singh/squeezing`.

B   Details of Forward and Reverse Process

B.1   Forward Process

Proof. We prove eq. (3) by induction. The statement is true for x_1 by eq. (2). Now let us assume it is true for x_t. Then:

$$
\begin{aligned}
x_{t+1} &= \sqrt{\alpha_{t+1}}x_t + \sqrt{1-\alpha_{t+1}}S_{t+1}\varepsilon_{t+1} \\
&= \sqrt{\alpha_{t+1}}\left[\sqrt{\alpha_t\ldots\alpha_1}x_0 + \sum_{i=1}^{t}S_i\varepsilon_i\sqrt{1-\alpha_i}\prod_{j=i+1}^{t}\sqrt{\alpha_j}\right] + \sqrt{1-\alpha_{t+1}}S_{t+1}\varepsilon_{t+1} \\
&= \sqrt{\alpha_{t+1}\alpha_t\ldots\alpha_1}x_0 + \sum_{i=1}^{t}S_i\varepsilon_i\sqrt{1-\alpha_i}\prod_{j=i+1}^{t+1}\sqrt{\alpha_j} + \sqrt{1-\alpha_{t+1}}S_{t+1}\varepsilon_{t+1} \\
&= \sqrt{\bar{\alpha}_{t+1}}x_0 + \sum_{i=1}^{t+1}S_i\varepsilon_i\sqrt{1-\alpha_i}\prod_{j=i+1}^{t+1}\sqrt{\alpha_j}
\end{aligned}
$$

$\square$

B.2   Validity of Posterior Approximation

In the text we noted that for time dependent squeezing, the Ho *et al.* posterior is an approximation. To see why, we compute the drift factor $S_t^{-1}S_{t-1}$ for the standard SDM case. We first compute the inverse of S_t^{SDM} using the Sherman-Morrison formula:

$$
S_t^{-1} = I + vv^\top\left(e^{s(t)} - 1\right) \tag{11}
$$

In our work $s(t)$ follows a linear schedule given by $s(t) = s_0\frac{\beta_t}{\beta_{\max}}$. Computing $S_t^{-1}S_{t-1}$ with this linear schedule yields

$$
S_t^{-1}S_{t-1} = I + vv^\top\left(\exp\left[\frac{s_0\Delta\beta}{\beta_{\max}}\right] - 1\right) \tag{12}
$$

Where $\Delta\beta = \beta_t - \beta_{t-1}$. The `diffusers` default linear schedule begins at $\beta = 1\times 10^{-4}$ and ends at $\beta = 0.02$. With 1000 timesteps (training), $\Delta\beta \sim 2\times 10^{-5}$ while with 50 timesteps (inference) $\Delta\beta \sim 4\times 10^{-4}$. Both numbers are small enough such that for small squeezing strengths $|s_0|$ the exponential is close to 1 so the drift term is approximately the identity and the posterior approximation in equation eq. (6) holds.