LOW-PASS FLOW MATCHING

Francesco M. Ruscio
ELLIS Institute Tübingen &
Max Planck Institute for Intelligent Systems &
AITHYRA
fruscio@aithyra.at

T. Konstantin Rusch
ELLIS Institute Tübingen &
Max Planck Institute for Intelligent Systems &
Tübingen AI Center &
Liquid AI
tkrusch@tue.ellis.eu

ABSTRACT

Flow Matching typically relies on white noise sources, a choice often misaligned with the power spectra of natural data, which tend to decay with frequency. To address this, we introduce **Low-Pass Flow Matching**, a variant of Flow Matching based on an operator-modulated interpolant. This formulation induces a time-varying spectral bias that transitions from the source spectrum to a frequency-decaying bias as the path approaches the data. We validate our method on unconditional image generation tasks, including the scientific Galaxy10 dataset. Empirically, we show that our method is particularly effective when paired with adaptive ODE solvers, where it improves or preserves sample quality while substantially reducing sampling cost compared to standard baselines.

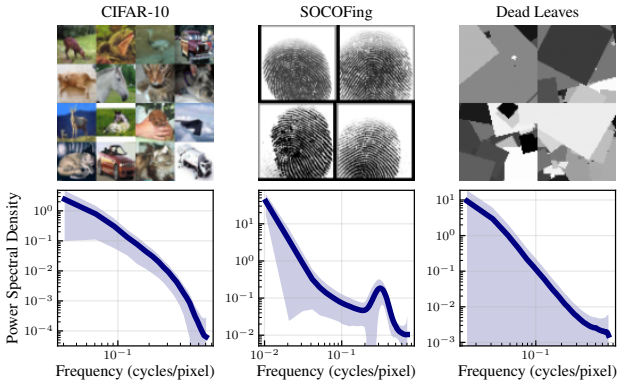

Figure 1: Radially averaged power spectral density of the training set across datasets. *(Left)*CIFAR-10. *(Center)*Sokoto Coventry Fingerprint dataset (Shehu et al., 2018). *(Right)*Dead Leaves synthetic dataset (Baradad et al., 2021).

1 INTRODUCTION

Dynamical systems have emerged as a powerful framework for generative modeling (Lai et al., 2025; Lipman et al., 2024). Modern approaches rely on SDEs (Ho et al., 2020; Song et al., 2021a;b), ODEs (Albergo & Vanden-Eijnden, 2023; Lipman et al., 2023; Liu et al., 2023b), and flow maps (Boffi et al., 2025), achieving substantial progress in vision (Gupta et al., 2024; Rombach et al., 2022), audio (Liu et al., 2023a), and scientific domains (Bose et al., 2024; Zeni et al., 2025). A wide range of inductive biases has been explored for diffusion models (Hoogeboom et al., 2022; Falck et al., 2025; Shariatian et al., 2025; Xu et al., 2022; 2023; Yoon et al., 2023), but comparable biases have been less studied for flow-based models. One example is spectral bias (Falck et al., 2025).

Many natural datasets exhibit power spectra that approximately follow $1/f^\gamma$, with $0 < \gamma \leq 2$ (Dutta & Horn, 1981; Ivanov et al., 1999; Ruderman, 1997; Srivastava et al., 2003; Torralba & Oliva, 2003; Van der Schaaf & van Hateren, 1996; Voss & Clarke, 1978). More generally, Figure 1 shows

that even when spectra do not follow a power law, they typically decay with increasing frequency. In contrast, diffusion models and flow matching (FM) commonly inject white noise (Dieleman, 2024), which can be misaligned with the spectral structure of the data (Falck et al., 2025; Randono, 2025). To better match the data's spectral content, we introduce **Low-Pass Flow Matching** (LP-FM), a variant of FM based on an operator-modulated interpolant. LP-FM induces a time-dependent spectral bias; for small t the path preserves the source spectrum, while as $t \to 1$ it transitions to a frequency-decaying bias. Empirically, we show that the proposed bias preserves or improves sample quality, can accelerate training, and reduces sampling cost across datasets, including a scientific imaging benchmark.

2 Related Work

Recent works analyze generative dynamics through a spectral lens. Liu et al. (2023c) frame generative processes as physical evolutions satisfying precise spectral properties, while Hoogeboom & Salimans (2023); Rissanen et al. (2023) recast the noising process as heat-equation-based blurring. A growing line of research focuses on spectral data properties (Dieleman, 2024; Jiralerspong et al., 2025) to optimize noise schedules (Benita et al., 2025) or to favor specific Fourier components by introducing autocorrelated noise (Falck et al., 2025; Huang et al., 2024; Randono, 2025). Others have introduced generative models that operate directly in general functional spaces (Gerdes et al., 2024; Phillips et al., 2022; Crabbé et al., 2024). Complementary to this, architectural interventions (Si et al., 2024; Yang et al., 2023; Wang et al., 2024; 2025) counteract the inherent spectral bias of the underlying network to better synthesize high- and low-frequency components during generation.

We defer a detailed comparison with Falck et al. (2025) and Randono (2025) to Appendix A.1.

3 Background

3.1 Flow Matching

Let p and q be probability distributions on $\mathbb{R}^d$, where p is a simple source distribution and q is the (unknown) data distribution. FM learns a time-varying vector field $u : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ that defines the ODE

$$\frac{d}{dt}\phi_t(x) = u_t\left(\phi_t(x)\right), \qquad \phi_0(x) = x. \tag{1}$$

Setting $p_0 = p$, the ODE induces the probability path $p_t \coloneqq [\phi_t]_\#(p_0)$. The goal is to learn u_t such that the terminal distribution p_1 matches the data distribution q.

In general, $u_t(x)$ is intractable. Instead, Lipman et al. (2023); Tong et al. (2024) show it is possible to learn $u_t(x)$ by introducing a conditioning variable z and minimizing the conditional flow matching (CFM) objective

$$\mathcal{L}_{\text{CFM}}(\theta) \coloneqq \mathbb{E}_{t,\rho(z),p_t(x|z)} \left\| v_\theta(t;x) - u_t(x|z) \right\|^2,$$

where ρ is the distribution over z, and $p_t(x|z)$, $u_t(x|z)$ denote the conditional probability path and its associated vector field. Setting $z = x_1$, with $\rho(z) = q(x_1)$, recovers FM (Lipman et al., 2023), while setting $z = (x_0, x_1)$, with $\rho(z) = p(x_0)q(x_1)$, yields CFM (Tong et al., 2024).

3.2 Power Spectra of Diffusion Models

Consider a random field $x \in \mathbb{R}^{\mathcal{G}}$ on a grid $\mathcal{G}$ with $|\mathcal{G}| = d$, e.g., images with $d = H \times W$. Let $\hat{x}(\boldsymbol{k}) = \mathcal{F}\{x\}(\boldsymbol{k})$ denote its discrete Fourier transform (DFT). The power spectral density (PSD) is

$$S_x(\boldsymbol{k}) \coloneqq \mathbb{E}\left[|\hat{x}(\boldsymbol{k})|^2\right].$$

Assuming x_0 and x_1 independent, a standard interpolant for diffusion models and FM is

$$x_t = \alpha_t\, x_1 + \sigma_t\, x_0, \qquad x_0 \sim \mathcal{N}(0, I_d), \;\; x_1 \sim q. \tag{2}$$

Consequently, the interpolant's PSD is (Dieleman, 2024; Falck et al., 2025)

$$S_{x_t}(\boldsymbol{k}) = \alpha_t^2 S_{x_1}(\boldsymbol{k}) + \sigma_t^2,$$

i.e., the injected noise is additive and constant across frequencies. In the following sections we assume that x_0 and x_1 are spectrally uncorrelated, i.e. $\mathbb{E}[\hat{x}_0(\boldsymbol{k})\,\overline{\hat{x}_1(\boldsymbol{k})}] = 0$ for all $\boldsymbol{k}$, where $\overline{(\cdot)}$ denotes complex conjugation.

4 FREQUENCY BIASED INTERPOLANT FOR FLOW MATCHING

We generalize Equation 2 by replacing the scalar noise scaling (σ_t) with a time-dependent linear shift-invariant (LSI) operator. Let $\mathcal{L}_t$ be a bounded LSI on $\ell^2(\mathcal{G})$. We define the interpolant

$$x_t \coloneqq tx_1 + \mathcal{L}_t x_0. \tag{3}$$

We design $\mathcal{L}_t = (1-t)\mathcal{L}'_t$ such that, informally, $\mathcal{L}'_t \to \mathcal{I}$ as $t \to 0$ (where $\mathcal{I}$ is the identity operator), preserving the source bias early in the path, and $\mathcal{L}'_t \to \mathcal{L}$ as $t \to 1$ for a target bias $\mathcal{L}$.

Proposition 4.1. *Let $\mathcal{L}_t$ be a bounded LSI operator with frequency response L_t and consider the interpolant in Equation 3. The PSD is*

$$S_{x_t}(\boldsymbol{k}) = t^2\, S_{x_1}(\boldsymbol{k}) + |L_t(\boldsymbol{k})|^2\, S_{x_0}(\boldsymbol{k}).$$

Proofs of this proposition and the subsequent one are given in Appendix B. Following Lipman et al. (2023), we recover the conditional vector field $u_t(x|x_1)$ from the interpolant in Equation 3.

Proposition 4.2. *Let $\mathcal{L}_t$ be differentiable in t and invertible for $t \in [0,1)$. Then, defining $\dot{\mathcal{L}}_t \coloneqq \partial_t \mathcal{L}_t$, the conditional vector field is*

$$u_t(x|x_1) = x_1 + \dot{\mathcal{L}}_t\, \mathcal{L}_t^{-1}(x - tx_1)\,.$$

Moreover, if we assume $x_0 \sim \mathcal{N}(\cdot|\mathbf{0}, \boldsymbol{I})$, the conditional path is still a Gaussian path

$$x_t|x_1 \sim \mathcal{N}\left(\cdot\middle|tx_1, \mathcal{L}_t\mathcal{L}_t^*\right)$$

where $\mathcal{L}_t^*$ is the adjoint of $\mathcal{L}_t$.

Finally, following Tong et al. (2024), we can condition on $z = (x_0, x_1)$ and obtain the CFM form of the conditional vector field (which we always use in our experiments):

$$u_t(x|z) = x_1 + \dot{\mathcal{L}}_t x_0 = x_1 + \mathcal{F}^{-1}\left\{(\partial_t L_t(\boldsymbol{k}))\, \hat{x}_0(\boldsymbol{k})\right\}.$$

Rotation-Invariant Parameterization. For image-like data we use isotropic filters, i.e. $L_t(\boldsymbol{k}) = L_t(k)$, with radial frequency $k \coloneqq \sqrt{\|\boldsymbol{k}\|^2}$. We consider two low-pass parameterizations: rational low-pass (RLP),

$$|L_t(\boldsymbol{k})|^2 = \frac{(1-t)^2}{\left(1 + \frac{k^2}{k_0^2}\right)^{\gamma_1 t}},$$

and Gaussian low-pass (GLP),

$$|L_t(\boldsymbol{k})|^2 = (1-t)^2 e^{-\gamma_1 t \|\boldsymbol{k}\|^2},$$

where $k_0 > 0$ sets the cutoff scale and $\gamma_1 > 0$ controls the strength of the spectral bias. When used within CFM, we refer to the resulting methods as RLP-CFM and GLP-CFM, respectively.

5 EXPERIMENTS

We evaluate RLP-CFM and GLP-CFM against FM, CFM, and variance-preserving flow matching (VP-FM; Lipman et al. (2023)) for unconditional image generation. We consider three datasets: Dead Leaves (Baradad et al., 2021), CIFAR-10, and Galaxy10 DECaLS (Leung & Bovy). Dead Leaves provides a controlled setting with an approximately $1/f^\gamma$ spectrum (cf. Figure 1 and the dashed reference in Figure 2), CIFAR-10 is a standard natural-image benchmark, and Galaxy10 represents a real-world scientific imaging task. We report the Fréchet Inception Distance (FID) (Heusel et al., 2017) and the number of function evaluations (NFE) required for sampling. FID is computed with CLEAN-FID (Parmar et al., 2022) using `legacy_tensorflow` settings and training-set Inception statistics. For sampling, we compare several ODE solvers: EULER and adaptive DOPRI5 (Dormand & Prince, 1980) and TSIT5 (Tsitouras, 2011). Our implementation follows the training protocol of Tong et al. (2024). We use the U-Net backbone of Dhariwal & Nichol (2021) with exponential moving average (EMA), dropout, and gradient-norm clipping, and optimize with either Adam (Kingma & Ba, 2014) or AdamW (Loshchilov & Hutter, 2019). Full hyperparameters are

reported in Appendix C. Qualitative comparisons between real and generated samples are deferred to Appendix D.

Dead Leaves. We generate 50k $1\times64\times64$ images by placing randomly oriented squares uniformly at random, following Baradad et al. (2021) to obtain an approximately $1/f^{\gamma}$ spectrum. We compare CFM and RLP-CFM; results are in Figure 2. RLP-CFM achieves substantially lower FID and reaches its best performance earlier: it achieves roughly half the FID of CFM and converges in $\sim$1350 epochs, while CFM continues to improve up to the end of training. The PSD evolution (right panel) suggests that RLP-CFM recovers higher-frequency content earlier along the path and matches the target spectral shape more closely around $t \approx 0.75$.

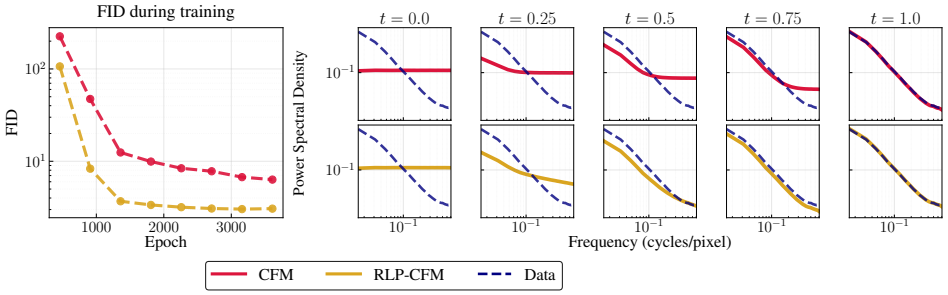

Figure 2: Dead Leaves results for CFM (crimson) and RLP-FM (goldenrod). *(Left)*FID (12.5k samples), evaluated every 450 training epochs. *(Right)*Radially averaged power spectral density of generated samples at different generation times: top row CFM, bottom row RLP-FM.

CIFAR-10 & Galaxy10 DECaLS. Table 1 summarizes the results (Galaxy10: 16k RGB images at 256×256). With EULER solver, baselines are strongest, suggesting that a spectral bias may be less beneficial under a fixed-step discretization[1]. Under adaptive solvers, however, RLP-CFM becomes competitive: on CIFAR-10 it preserves sample quality while reducing NFE, and on Galaxy10 it improves FID while substantially reducing NFE (up to $1.65\times$ vs. CFM and $2.16\times$ vs. VP-FM). In contrast, GLP-CFM underperforms RLP-CFM on CIFAR-10 and we were unable to obtain stable training on Galaxy10. Overall, these results indicate that a spectral bias can be most useful when paired with adaptive integration.

Dataset →	CIFAR-10			Galaxy10 DECaLS		
Solver →	Euler (NFE=100)	Adaptive		Euler (NFE=100)	Adaptive	
Method ↓	FID	FID	NFE	FID	FID	NFE
CFM/FM	4.66	**3.96**	80.7	**11.94**	11.23	92.0
VP-FM	**4.44**	4.29	101.2	12.42	12.84	120.75
RLP-CFM	5.27	4.09	67.5	13.78	**10.03**	**56.0**
GLP-CFM	6.71	5.35	**59.1**	–	–	–

Table 1: FID and NFE with EULER and adaptive solvers (DOPRI5/TSIT5; best over adaptive). Baselines: CFM, VP-FM. This work: RLP-CFM, GLP-CFM.

6 CONCLUSION

We introduced Low-Pass Flow Matching, a simple operator-modulated interpolant that induces a time-varying spectral bias in Flow Matching. Experiments suggest the bias is most useful with adaptive ODE solvers, improving or preserving sample quality while reducing sampling cost. Future work will evaluate RLP-CFM on additional scientific datasets.

[1]Analogously, this may apply to DDIM (Song et al., 2021a); we do not evaluate this here.

REFERENCES

Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=li7qeBbCR1t.

Manel Baradad, Jonas Wulff, Tongzhou Wang, Phillip Isola, and Antonio Torralba. Learning to see by looking at noise. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=RQUl8gZnN7O.

Roi Benita, Michael Elad, and Joseph Keshet. Spectral analysis of diffusion models with application to schedule design. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=ymmY3rrD1t.

Nicholas Matthew Boffi, Michael Samuel Albergo, and Eric Vanden-Eijnden. How to build a consistency model: Learning flow maps via self-distillation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=Di5apl8HSH.

Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian FATRAS, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael M. Bronstein, and Alexander Tong. SE(3)-stochastic flow matching for protein backbone generation. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=kJFIH23hXb.

Jonathan Crabbé, Nicolas Huynh, Jan Stanczuk, and Mihaela Van Der Schaar. Time series diffusion in the frequency domain. *arXiv preprint arXiv:2402.05933*, 2024.

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

Sander Dieleman. Diffusion is spectral autoregression, 2024. URL https://sander.ai/2024/09/02/spectral-autoregression.html.

J. R. Dormand and P. J. Prince. A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math*, 6:19–26, 1980.

Pulak Dutta and PM Horn. Low-frequency fluctuations in solids: 1 f noise. *Reviews of Modern physics*, 53(3):497, 1981.

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.

Ruiqi Gao, Emiel Hoogeboom, Jonathan Heek, Valentin De Bortoli, Kevin P. Murphy, and Tim Salimans. Diffusion meets flow matching: Two sides of the same coin. 2024. URL https://diffusionflow.github.io/.

Victor Garcia Satorras, Emiel Hoogeboom, Fabian Fuchs, Ingmar Posner, and Max Welling. E (n) equivariant normalizing flows. *Advances in Neural Information Processing Systems*, 34:4181–4192, 2021.

Mathis Gerdes, Max Welling, and Miranda CN Cheng. Gud: Generation with unified diffusion. *arXiv preprint arXiv:2410.02667*, 2024.

Agrim Gupta, Lijun Yu, Kihyuk Sohn, Xiuye Gu, Meera Hahn, Fei-Fei Li, Irfan Essa, Lu Jiang, and José Lezama. Photorealistic video generation with diffusion models. In *European Conference on Computer Vision*, pp. 393–411. Springer, 2024.

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

Emiel Hoogeboom and Tim Salimans. Blurring diffusion models. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OjDkC57x5sz.

Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pp. 8867–8887. PMLR, 2022.

Xingchang Huang, Corentin Salaun, Cristina Vasconcelos, Christian Theobalt, Cengiz Oztireli, and Gurprit Singh. Blue noise for diffusion models. In *ACM SIGGRAPH 2024 conference papers*, pp. 1–11, 2024.

Plamen Ch Ivanov, Luis A Nunes Amaral, Ary L Goldberger, Shlomo Havlin, Michael G Rosenblum, Zbigniew R Struzik, and H Eugene Stanley. Multifractality in human heartbeat dynamics. *Nature*, 399(6735):461–465, 1999.

Thomas Jiralerspong, Berton Earnshaw, Jason Hartford, Yoshua Bengio, and Luca Scimeca. Shaping inductive bias in diffusion models through frequency-based noise control. In *ICLR 2025 Workshop on Deep Generative Model in Machine Learning: Theory, Principle and Efficacy*, 2025.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL https://api.semanticscholar.org/CorpusID:6628106.

Jonas Köhler, Leon Klein, and Frank Noé. Equivariant flows: exact likelihood generative learning for symmetric densities. In *International conference on machine learning*, pp. 5361–5370. PMLR, 2020.

Chieh-Hsin Lai, Yang Song, Dongjun Kim, Yuki Mitsufuji, and Stefano Ermon. The principles of diffusion models, 2025. URL https://arxiv.org/abs/2510.21890.

Henry Leung and Jo Bovy. Galaxy10 decals dataset. https://github.com/henrysky/Galaxy10.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky T. Q. Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code, 2024. URL https://arxiv.org/abs/2412.06264.

Haohe Liu, Zehua Chen, Yi Yuan, Xinhao Mei, Xubo Liu, Danilo Mandic, Wenwu Wang, and Mark D. Plumbley. Audioldm: text-to-audio generation with latent diffusion models. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023a.

Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023b. URL https://openreview.net/forum?id=XVjTT1nw5z.

Ziming Liu, Di Luo, Yilun Xu, Tommi Jaakkola, and Max Tegmark. Genphys: From physical processes to generative models. *arXiv preprint arXiv:2304.02637*, 2023c.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Gaurav Parmar, Richard Zhang, and Jun-Yan Zhu. On aliased resizing and surprising subtleties in gan evaluation. In *CVPR*, 2022.

Angus Phillips, Thomas Seror, Michael Hutchinson, Valentin De Bortoli, Arnaud Doucet, and Emile Mathieu. Spectral diffusion processes. *arXiv preprint arXiv:2209.14125*, 2022.

Andrew Randono. Cloud diffusion part 1: Theory and motivation, 2025. URL https://arxiv.org/abs/2507.05496.

Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=4PJUBT9f2Ol.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Daniel L Ruderman. Origins of scaling in natural images. *Vision research*, 37(23):3385–3398, 1997.

Dario Shariatian, Umut Simsekli, and Alain Oliviero Durmus. Denoising levy probabilistic models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=SYmUS6qRub.

Yahaya Isah Shehu, Ariel Ruiz-Garcia, Vasile Palade, and Anne James. Sokoto coventry fingerprint dataset, 2018. URL https://arxiv.org/abs/1807.10609.

Chenyang Si, Ziqi Huang, Yuming Jiang, and Ziwei Liu. Freeu: Free lunch in diffusion u-net. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4733–4743, 2024.

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

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Anuj Srivastava, Ann B Lee, Eero P Simoncelli, and S-C Zhu. On advances in statistical modeling of natural images. *Journal of mathematical imaging and vision*, 18(1):17–33, 2003.

Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=CD9Snc73AW. Expert Certification.

Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: computation in neural systems*, 14(3):391, 2003.

Ch Tsitouras. Runge–kutta pairs of order 5 (4) satisfying only the first column simplifying assumption. *Computers & Mathematics with Applications*, 62(2):770–775, 2011.

van A Van der Schaaf and JH van van Hateren. Modelling the power spectra of natural images: statistics and information. *Vision research*, 36(17):2759–2770, 1996.

Richard F Voss and John Clarke. "1/f noise"in music: Music from 1/f noise. *The Journal of the Acoustical Society of America*, 63(1):258–263, 1978.

Haixin Wang, Jiashu Pan, Hao Wu, Fan Zhang, and Tailin Wu. Fourierflow: Frequency-aware flow matching for generative turbulence modeling. *arXiv preprint arXiv:2506.00862*, 2025.

Jing Wang, Songtao Wu, Zhiqiang Yuan, Qiang Tong, and Kuanhong Xu. Frequency compensated diffusion model for real-scene dehazing. *Neural Networks*, 175:106281, 2024.

Yilun Xu, Ziming Liu, Max Tegmark, and Tommi S. Jaakkola. Poisson flow generative models. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=voV_TRqcWh.

Yilun Xu, Ziming Liu, Yonglong Tian, Shangyuan Tong, Max Tegmark, and Tommi Jaakkola. Pfgm++: Unlocking the potential of physics-inspired generative models. In *International Conference on Machine Learning*, pp. 38566–38591. PMLR, 2023.

Xingyi Yang, Daquan Zhou, Jiashi Feng, and Xinchao Wang. Diffusion probabilistic model made slim. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pp. 22552–22562, 2023.

Eunbi Yoon, Keehun Park, Sungwoong Kim, and Sungbin Lim. Score-based generative models with lévy processes. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=0Wp3VHX0Gm.

Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, et al. A generative model for inorganic materials design. *Nature*, 639(8055):624–632, 2025.

A EXTENDEND RELATED WORK

Inductive biases in generative modeling. Generative models often rely on inductive biases to tailor their sampling dynamics to specific data structures. Notable examples include physics-inspired biases, such as Lévy processes for heavy-tailed data (Yoon et al., 2023; Shariatian et al., 2025) and Poissonian flows (Xu et al., 2022; 2023), that have been found to be a valid alternative to diffusion models. Another deeply explored area is geometric bias; Köhler et al. (2020); Garcia Satorras et al. (2021) developed equivariant continuous normalizing flows, while Hoogeboom et al. (2022); Bose et al. (2024) demonstrate the necessity of symmetry-preserving models for biological data. Beyond symmetry, a broad body of work covers alternative inductive biases in generative dynamics, including methods that design probability path or flows with low-curvature trajectories (Lipman et al., 2023; Tong et al., 2024; Liu et al., 2023b).

A.1 COMPARISON WITH EQUALSNR AND CLOUD DIFFUSION

We compare LP-FM with EqualSNR and Cloud Diffusion, two methods that modify diffusion models by using autocorrelated noise.

Cloud Diffusion (Randono, 2025) considers noise spectra of the form $S_{\text{noise}}(k) \propto 1/k$, which diverge as $k \to 0$ and may therefore lead to numerical instability near the DC component; since no experiments are reported, we limit our discussion to methodological considerations.

EqualSNR (Falck et al., 2025) instead sets $S_{\text{noise}}(\boldsymbol{k}) \propto S_1(\boldsymbol{k})$ to align the forward noising spectrum with the data spectrum, motivated by Gaussian assumptions in the DDPM reverse process (Ho et al., 2020). In contrast, FM does not require the same reverse-time Gaussianity; consequently, it is a priori unclear whether this rationale transfers. This motivates studying spectral bias in FM in its own right. Although EqualSNR is motivated by DDPM assumptions, it is evaluated with DDIM sampling (Song et al., 2021a), which is closely related to FM with EULER solver (Gao et al., 2024). Empirically, we do not observe benefits from introducing a spectral bias under EULER (except on the toy Dead Leaves), whereas gains emerge when FM is paired with adaptive ODE solvers. Therefore, the benefits we report arise under sampling conditions substantially different from DDIM. There are also methodological differences: EqualSNR operates in Fourier space, whereas we work in pixel space.

Finally, our approach induces a time-varying bias: $\mathcal{L}_t$ is designed to preserve the source spectral structure early along the path while progressively imposing a low-pass bias near the data distribution. In contrast, EqualSNR and Cloud Diffusion prescribe a noise spectrum intended to match the data, but do not explicitly preserve the spectral bias of the source distribution. Overall, our results highlight that (i) the solver choice can be critical for realizing the benefits of spectral bias in FM, and (ii) time-varying operator modulation provides a way to incorporate spectral structure while remaining compatible with nontrivial source distributions.

B PROOFS

This appendix collects proofs of the propositions in Section 4. For convenience, we restate each proposition before its proof. We also recall a few operator-theoretic and Fourier-analytic facts used throughout.

B.1 PRELIMINARIES: LSI OPERATORS ON A FINITE GRID

Let $\mathcal{G}$ be a finite periodic grid, and equip $\mathbb{C}^{\mathcal{G}}$ with the standard ℓ^2 inner product $\langle x, y\rangle := \sum_{g\in\mathcal{G}} x(g)\,\overline{y(g)}$ and norm $\|x\|_2^2 = \langle x, x\rangle$. We denote by $\hat{x}(\boldsymbol{k}) = \mathcal{F}\{x\}(\boldsymbol{k})$ the DFT of x. We use the convention that the DFT is unitary so that Parseval's identity holds:

$$\|x\|_2^2 = \sum_{\boldsymbol{k}} |\hat{x}(\boldsymbol{k})|^2.$$

Bounded linear operators. A linear map $\mathcal{L} : \ell^2(\mathcal{G}) \to \ell^2(\mathcal{G})$ is bounded if

$$\|\mathcal{L}\|_{\text{op}} := \sup_{\|x\|_2=1} \|\mathcal{L}x\|_2 < \infty.$$

On a finite-dimensional space every linear operator is bounded.

Shift-invariant (LSI) operators and frequency response. An operator $\mathcal{L}$ is shift-invariant if it commutes with grid shifts. Equivalently, $\mathcal{L}$ is a circular convolution with some kernel $h \in \mathbb{R}^{\mathcal{G}}$:

$$(\mathcal{L}x)(g) = (h * x)(g).$$

Such operators diagonalize in the Fourier domain: there exists a complex-valued function $L(\boldsymbol{k})$ (the frequency response) such that for all $\boldsymbol{k}$,

$$\mathcal{F}\left\{\mathcal{L}x\right\}(\boldsymbol{k}) = L(\boldsymbol{k})\,\hat{x}(\boldsymbol{k}) \tag{4}$$

For a time-indexed family $\{\mathcal{L}_t\}_{t\in[0,1]}$ we write $L_t(\boldsymbol{k})$ for its response.

Invertibility. An LSI operator $\mathcal{L}$ is invertible on $\ell^2(\mathcal{G})$ if and only if $L(\boldsymbol{k}) \neq 0$ for all $\boldsymbol{k}$. In that case the inverse is also LSI and satisfies $\mathcal{F}\left\{\mathcal{L}^{-1}x\right\}(\boldsymbol{k}) = \frac{1}{L(\boldsymbol{k})}\,\hat{x}(\boldsymbol{k})$.

Spectral (un)correlatedness. For random fields $x_0, x_1 \in \mathbb{R}^{\mathcal{G}}$ we say they are spectrally uncorrelated if

$$\mathbb{E}\Big[\hat{x}_0(\boldsymbol{k})\,\overline{\hat{x}_1(\boldsymbol{k})}\Big] = 0 \qquad \text{for all } \boldsymbol{k}.$$

B.2 Proof of Proposition 4.1

Proposition B.1. *Let $\mathcal{L}_t$ be a bounded LSI operator with frequency response $L_t(\boldsymbol{k})$ and consider*

$$x_t = t\,x_1 + \mathcal{L}_t x_0.$$

Assume x_0 and x_1 are spectrally uncorrelated. Then the PSD of x_t satisfies

$$S_{x_t}(\boldsymbol{k}) = t^2\,S_{x_1}(\boldsymbol{k}) + |L_t(\boldsymbol{k})|^2\,S_{x_0}(\boldsymbol{k}).$$

Proof of Proposition B.1. Taking the DFT of the interpolant and using linearity yields, for each $\boldsymbol{k}$,

$$\hat{x}_t(\boldsymbol{k}) = t\,\hat{x}_1(\boldsymbol{k}) + \mathcal{F}\left\{\mathcal{L}_t x_0\right\}(\boldsymbol{k}).$$

Since $\mathcal{L}_t$ is LSI with response $L_t(\boldsymbol{k})$, Equation 4 yields $\mathcal{F}\{\mathcal{L}_t x_0\}(\boldsymbol{k}) = L_t(\boldsymbol{k})\,\hat{x}_0(\boldsymbol{k})$, hence

$$\hat{x}_t(\boldsymbol{k}) = t\,\hat{x}_1(\boldsymbol{k}) + L_t(\boldsymbol{k})\,\hat{x}_0(\boldsymbol{k}).$$

Therefore,

$$\begin{aligned} S_{x_t}(\boldsymbol{k}) &= \mathbb{E}\big[|\hat{x}_t(\boldsymbol{k})|^2\big] = \mathbb{E}\Big[\big|t\,\hat{x}_1(\boldsymbol{k}) + L_t(\boldsymbol{k})\,\hat{x}_0(\boldsymbol{k})\big|^2\Big] \\ &= t^2\,\mathbb{E}\big[|\hat{x}_1(\boldsymbol{k})|^2\big] + |L_t(\boldsymbol{k})|^2\,\mathbb{E}\big[|\hat{x}_0(\boldsymbol{k})|^2\big] + 2\,\Re\Big(t\,L_t(\boldsymbol{k})\,\mathbb{E}\Big[\hat{x}_0(\boldsymbol{k})\,\overline{\hat{x}_1(\boldsymbol{k})}\Big]\Big). \end{aligned}$$

The cross term vanishes by spectral uncorrelatedness. Recognizing $\mathbb{E}[|\hat{x}_i(\boldsymbol{k})|^2] = S_{x_i}(\boldsymbol{k})$ for $i \in \{0,1\}$ yields

$$S_{x_t}(\boldsymbol{k}) = t^2\,S_{x_1}(\boldsymbol{k}) + |L_t(\boldsymbol{k})|^2\,S_{x_0}(\boldsymbol{k}).$$

□

B.3 Proof of Proposition 4.2

Proposition B.2. *Let $\{\mathcal{L}_t\}_{t\in[0,1)}$ be a family of bounded, invertible LSI operators, and assume that $t \mapsto \mathcal{L}_t$ is differentiable with respect to the operator norm. Define $\dot{\mathcal{L}}_t := \partial_t \mathcal{L}_t$ and consider the interpolant*

$$x_t = t\,x_1 + \mathcal{L}_t x_0.$$

Then the conditional vector field $u_t(x|x_1)$ associated with the conditional path $p_t(\cdot|x_1)$ is

$$u_t(x|x_1) = x_1 + \dot{\mathcal{L}}_t\,\mathcal{L}_t^{-1}(x - tx_1)\,. \tag{a}$$

Moreover, if $x_0 \sim \mathcal{N}(\cdot|\mathbf{0}, \boldsymbol{I})$, then

$$x_t|x_1 \sim \mathcal{N}\left(\cdot\middle|tx_1, \mathcal{L}_t\mathcal{L}_t^*\right) \tag{b}$$

where $\mathcal{L}_t^*$ is the adjoint of $\mathcal{L}_t$.

Proof of Proposition B.2. We prove Equation a and Equation b:

(a) Our derivation proceeds along the the lines of Lipman et al. (2023), who derive the conditional vector field for a Gaussian path with scalar noise scaling.

Fix x_1. Define the family of conditional flow maps $\phi_t(\cdot|x_1) : \mathbb{R}^{\mathcal{G}} \to \mathbb{R}^{\mathcal{G}}$ by

$$\phi_t(x_0|x_1) \coloneqq t\, x_1 + \mathcal{L}_t x_0.$$

For each fixed x_1, the conditional path $p_t(\cdot|x_1)$ is the pushforward of the source law of x_0 through $\phi_t(\cdot|x_1)$, i.e. $x_t|x_1 = \phi_t(x_0|x_1)$.

Let $y = \phi_t(x_0|x_1)$. Since $\mathcal{L}_t$ is invertible for all $t \in [0, 1)$, the map $\phi_t(\cdot|x_1)$ is invertible and

$$x_0 = \phi_t^{-1}(y|x_1) = \mathcal{L}_t^{-1}(y - tx_1).$$

Differentiating $\phi_t(x_0|x_1)$ with respect to t yields

$$\partial_t \phi_t(x_0|x_1) = x_1 + \dot{\mathcal{L}}_t x_0.$$

By definition of the velocity field associated with the conditional flow,

$$u_t(y|x_1) = \partial_t \phi_t(x_0|x_1)\Big|_{x_0=\phi_t^{-1}(y|x_1)} = x_1 + \dot{\mathcal{L}}_t\, \mathcal{L}_t^{-1}(y - tx_1).$$

Renaming y as x gives Equation a.

(b) Assume $x_0 \sim \mathcal{N}(\cdot \mid \mathbf{0}, \boldsymbol{I})$ and fix x_1. Since

$$x_t = t\, x_1 + \mathcal{L}_t x_0$$

is an affine transformation of the Gaussian random vector x_0, it is Gaussian. Its conditional mean is

$$\mathbb{E}[x_t|x_1] = t\, x_1 + \mathcal{L}_t\, \mathbb{E}[x_0] = t\, x_1,$$

and its conditional covariance is

$$\begin{aligned}
\mathrm{Cov}(x_t|x_1) &= \mathbb{E}\Big[\big(x_t - \mathbb{E}[x_t|x_1]\big)\big(x_t - \mathbb{E}[x_t|x_1]\big)^* \mid x_1\Big] \\
&= \mathbb{E}[x_t x_t^*|x_1] - \mathbb{E}[x_t|x_1]\, \mathbb{E}[x_t|x_1]^* \\
&= \mathbb{E}\Big[\big(\mathcal{L}_t x_0\big)\big(\mathcal{L}_t x_0\big)^* \mid x_1\Big] \\
&= \mathcal{L}_t\, \mathbb{E}[x_0 x_0^*]\, \mathcal{L}_t^* \\
&= \mathcal{L}_t\, \mathrm{Cov}(x_0)\, \mathcal{L}_t^* \\
&= \mathcal{L}_t\, I_d\, \mathcal{L}_t^* \\
&= \mathcal{L}_t \mathcal{L}_t^*.
\end{aligned}$$

Therefore,

$$x_t \mid x_1 \sim \mathcal{N}(\cdot \mid tx_1,\ \mathcal{L}_t \mathcal{L}_t^*)\,,$$

which is Equation b

□

C EXPERIMENTAL SETTINGS

Dataset ↓	Channels	Depth	Channel multipliers	Attention resolutions	Head channels
Dead Leaves	128	2	[1,2,2,2]	16	64
CIFAR-10	128	2	[1,2,2,2]	16	64
Galaxy10	128	2	[1,1,2,2,2,4]	32,16,8	64

Table 2: UNET settings for each dataset. For more information, please refer to Dhariwal & Nichol (2021).

Dataset ↓	Epochs	EMA	Dropout	Clip grad norm	Optimizer	LR	LR Warmup
Dead Leaves	3750	0.9999	0.1	1.0	AdamW	1×10^{-4}	✗
CIFAR-10	1900	0.9999	0.1	1.0	Adam	2×10^{-4}	✗
Galaxy10	3750	0.9999	0.05	0.75	Adam	1×10^{-4}	✓

Table 3: Training hyperparameters used for each dataset (training epochs, EMA decay, dropout, gradient clipping norm, optimizer, learning rate, learning rate warmup).

D GENERATED IMAGES

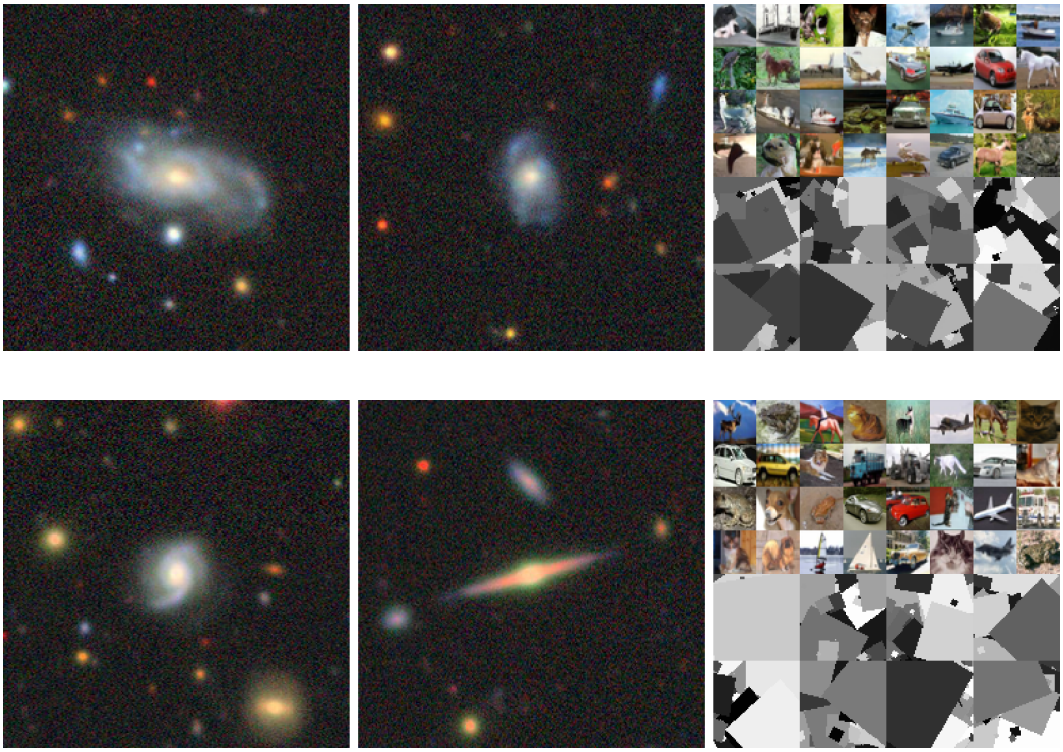

Figure 3: Galaxy10 DECaLS, CIFAR-10, and Dead Leaves images. *(Top)*Generated images with RLP-CFM. *(Bottom)*Real images.