BOLTZMANN GENERATORS FOR CONDENSED MATTER VIA RIEMANNIAN FLOW MATCHING

Emil Hoffmann, Maximilian Schebek, Leon Klein
Freie Universität Berlin
Berlin, Germany
{emil.hoffmann,m.schebek,leon.klein}@fu-berlin.de

Frank Noé
Microsoft Research & Freie Universität Berlin
Berlin, Germany
frank.noe@microsoft.com

Jutta Rogal
Flatiron Institute
New York, USA
jrogal@flatironinstitute.org

ABSTRACT

Sampling equilibrium distributions is fundamental to statistical mechanics. While flow matching has emerged as scalable state-of-the-art paradigm for generative modeling, its potential for equilibrium sampling in condensed-phase systems remains largely unexplored. We address this by incorporating the periodicity inherent to these systems into continuous normalizing flows using Riemannian flow matching. The high computational cost of exact density estimation intrinsic to continuous normalizing flows is mitigated by using Hutchinson's trace estimator, utilizing a crucial bias-correction step based on cumulant expansion to render the stochastic estimates suitable for rigorous thermodynamic reweighting. Our approach is validated on monatomic ice, demonstrating the ability to train on systems of unprecedented size and obtain highly accurate free energy estimates without the need for traditional multistage estimators.

1 INTRODUCTION

Accurate estimation of thermodynamic observables, such as free energy and heat capacity, in condensed matter systems relies fundamentally on the efficient sampling of the high-dimensional equilibrium distributions of large systems. Traditionally, molecular dynamics (MD) and Monte Carlo (MC) simulations are considered as primary sampling tools, but their sequential nature requires many steps to generate uncorrelated samples, making them computationally expensive. In recent years, machine learning approaches have emerged as a promising alternative. In particular, Boltzmann Generators (BGs) (Noé et al., 2019) leverage invertible neural networks, such as normalizing flows, to generate equilibrium samples directly while enabling tractable likelihood evaluation.

Continuous normalizing flows (CNFs) (Chen et al., 2018) trained with the scalable and efficient flow-matching training paradigm (Lipman et al., 2023; Albergo & Vanden-Eijnden, 2023; Liu et al., 2023) have emerged as the state-of-the-art approach for modeling complex biomolecular systems (Klein & Noe, 2024; Rehman et al., 2025). However, the development of flexible CNFs for condensed-phase systems, commonly modeled using periodic boundary conditions, remains limited (Grenioux et al., 2025). Previous studies have primarily employed architecturally constrained coupling-flow models (Dinh et al., 2017) for crystals (Wirnsberger et al., 2022; Schebek et al., 2024; Ahmad & Cai, 2022) and liquids (Wirnsberger et al., 2020; Coretti et al., 2025), but these approaches were limited by computational cost, restricting training to system sizes of a few hundred particles, which is significantly smaller than what is required to eliminate finite-size effects (Polson et al., 2000).

The goal of this work is to introduce CNFs with periodic boundary conditions in combination with Riemannian flow matching (Chen & Lipman, 2024) and tractable density estimation as a flexible, scalable, and size-transferable framework for equilibrium sampling of condensed-phase systems. We show that this approach overcomes the scalability limitations of prior coupling-flow architectures by training on a monatomic ice system more than four times larger than previous benchmarks,

achieving excellent sample quality while matching the computational budget of the smaller baselines.

2 METHODS

Boltzmann Generators combine exact-likelihood deep generative models, such as normalizing flows (Rezende & Mohamed, 2015; Papamakarios et al., 2021), with importance reweighting to sample from a target Boltzmann distribution $\mu(\mathbf{x}) = Z^{-1}\exp[-\beta U(\mathbf{x})]$. This distribution is defined by the potential $U(\mathbf{x})$, the inverse temperature $\beta = (k_B T)^{-1}$, with Boltzmann constant k_B, temperature T and the partition function Z. The model is trained to learn an invertible mapping f_θ that transforms a prior sample $\mathbf{x}_0 \sim p(\mathbf{x}_0)$ into a generated sample $\mathbf{x}_1 = f_\theta(\mathbf{x}_0)$, whose distribution $\tilde{p}(\mathbf{x}_1)$ approximates $\mu(\mathbf{x}_1)$. Unbiased estimates can be obtained by reweighting the generated samples to the target distribution using the importance weights

$$w(\mathbf{x}_0) = \exp\left(\beta_p U_p(\mathbf{x}_0) - \beta_\mu U_\mu(f_\theta(\mathbf{x}_0)) - \Delta \log p\right), \tag{1}$$

such that $w(\mathbf{x}_0) \propto \mu(f_\theta(\mathbf{x}_0))/\tilde{p}(f_\theta(\mathbf{x}_0))$. The density change $\Delta \log p$ is defined as

$$\Delta \log p := \log \tilde{p}(f_\theta(\mathbf{x}_0)) - \log p(\mathbf{x}_0) = -\log\left|\det J_f(\mathbf{x}_0)\right|, \tag{2}$$

where J_f denotes the Jacobian of the transformation. The importance weights further provide a measure of modeling performance via the effective sample size (ESS) (Kish, 1965) defined by $\mathrm{ESS} = (\sum_{i=1}^{N_S} w(\mathbf{x}_0^{(i)}))^2 / \sum_{i=1}^{N_S} (w(\mathbf{x}_0^{(i)}))^2$, which is typically reported relative to the number of samples N_S. While the free energy F, which governs phase stability, is defined in terms of the partition function Z as $F = -k_B T \log Z$, evaluating the partition function is typically intractable. Within the framework of targeted free energy perturbation (TFEP) (Jarzynski, 2002), a trained BG can be used to directly estimate the free energy difference between prior and target states through $\beta \Delta F_{p\mu} = -\log \mathbb{E}_{\mathbf{x}_0 \sim p(\mathbf{x}_0)}\left[w(\mathbf{x}_0)\right]$. This allows the calculation of free-energy differences without sampling intermediate states via MD simulations, which are required by traditional estimators that rely on sufficient phase-space overlap between distributions (Shirts & Chodera, 2008). Knowing the free energy of the prior state allows estimation of the absolute free energy of the target system and, consequently, a comparison of the thermodynamic stability of different phases.

Normalizing Flows learn an invertible mapping f_θ to transform samples from a prior distribution $p(\mathbf{x}_0)$ to a generated density $\tilde{p}(\mathbf{x}_1)$ that approximates the target distribution $\mu(\mathbf{x}_1)$. CNFs (Chen et al., 2018) model this transformation in continuous time by solving an ordinary differential equation (ODE) over $t \in [0, 1]$,

$$\frac{df_\theta^t(\mathbf{x})}{dt} = v_\theta(t, f_\theta^t(\mathbf{x})), \quad f_\theta^0(\mathbf{x}) = \mathbf{x}_0, \tag{3}$$

where v_θ is a time-dependent vector field with learnable parameters θ. The resulting density evolution follows the continuous change-of-variables formula

$$\Delta \log p = -\int_0^1 \nabla \cdot v_\theta(t, f_\theta^t(\mathbf{x}_0))\, dt, \tag{4}$$

where the divergence of the vector field is equivalent to the trace of the vector field's Jacobian. CNFs can be naturally extended to general Riemannian manifolds $\mathcal{M}$ by formulating and solving the associated ODE directly on $\mathcal{M}$ (Mathieu & Nickel, 2020).

Riemannian Flow Matching. While normalizing flows are commonly trained by maximum-likelihood objective, Lipman et al. (2023) introduced flow matching for efficient and simulation-free training of CNFs, which has been generalized to manifolds by Chen & Lipman (2024). In the case of periodic boundary conditions, corresponding to the flat torus, the Riemannian conditional flow matching (RFM) objective allows to train the vector field through

$$\mathcal{L}_{\mathrm{RFM}}(\theta) = \mathbb{E}_{t\sim[0,1],\, \mathbf{x}\sim p_t(\mathbf{x}|\mathbf{z})}\left[\|v_\theta(t, \mathbf{x}) - u_t(\mathbf{x}|\mathbf{z})\|_2^2\right], \tag{5}$$

where $p_t(\mathbf{x}|\mathbf{z})$ is a probability path conditioned on an arbitrary distribution $\mathbf{z}$. A proven choice for this path and the conditional distribution $\mathbf{z}$ is the direct constant speed interpolation between a single prior and target sample (Chen & Lipman, 2024), which yields the conditional vector field

$$u_t(\mathbf{x}|\mathbf{z}) = \log_{\mathbf{x}_0}(\mathbf{x}_1) \quad \text{with} \quad \mathbf{z} = (\mathbf{x}_0, \mathbf{x}_1),\ \mathbf{x}_0 \sim p,\ \mathbf{x}_1 \sim \mu, \tag{6}$$

where the explicit form of the logarithmic map depends on the shape of the simulation box (see Appendix A.1).

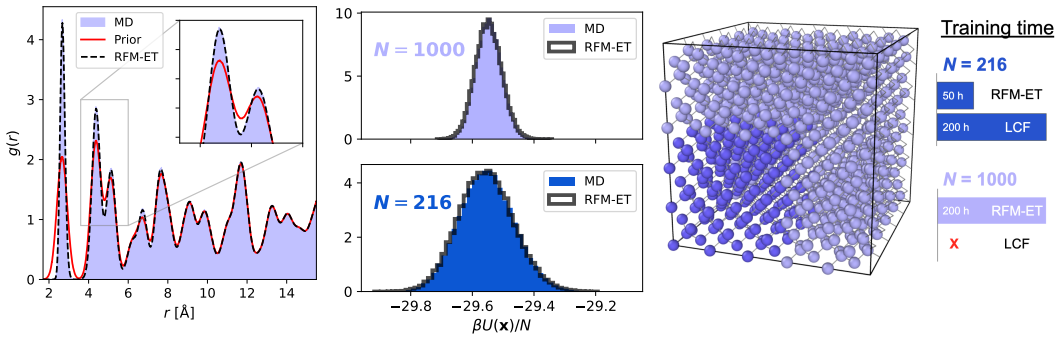

Figure 1: Left: Radial distribution function of the cubic ice system with 1000 particles as obtained from MD and RFM-ET. Center: Energy histograms of the cubic ice system with 1000 (top) and 216 (bottom) particles as obtained from MD and RFM-ET. Right: The 216 particles (blue) and 1000 particles (light blue) cubic ice systems and their training times with our method and local coupling flows (LCF) (Schebek et al. (2025b)).

3 Implementation

RFM as discussed above provides a general framework for modeling condensed-phase systems, while leaving considerable freedom in the precise implementation. In the following section, we specify our particular choices for the vector field and the prior distribution.

Symmetries and Prior Alignment. The equilibrium configurational distribution naturally inherits the symmetries of the potential energy, including invariance wrt. global translations, rotations, and particle permutations. Embedding those symmetries into generative flow models has been demonstrated to substantially improve data efficiency and sample quality (Köhler et al., 2020; Satorras et al., 2021; Klein et al., 2023). In principle, the presented framework allows for straightforward incorporation of such symmetries, by combining a symmetry-invariant prior, such as the uniform distribution on the torus, with an equivariant vector field, as done in Grenioux et al. (2025). However, for the specific case of crystalline solids, we exploit the underlying lattice structure to adopt a distinct, physics-informed strategy: A $3N$-dimensional wrapped Gaussian centered at equilibrium lattice sites (Wirnsberger et al., 2022) (see Appendix A.2.1) known as Einstein crystal (Frenkel & Ladd, 1984).

As the chosen lattice prior is not translation invariant, we restrict the system to a mean-free subspace, by fixing the center of mass of both the prior and target data. The learned dynamics are constrained to preserve the center of mass by subtracting the mean particle velocity from both the learned vector field and the conditional target field. The continuous $\mathrm{SO}(3)$ symmetry of Euclidean systems is reduced to a discrete subgroup for periodic boundary conditions. Our lattice-based prior implicitly enforces a global alignment with the simulation box axes and a trivial assignment between prior and target samples. While we perform mini-batch optimal transport reordering (Tong et al., 2023; Pooladian et al., 2023), we omit further searches over discrete rotations and only adapt the transport cost calculation to torus distances.

Vector Field Parameterization. To scale effectively to larger systems, we adopt a local modeling approach for the vector field, which ensures that the learned dynamics are transferable across different system sizes (Schebek et al., 2025b). We leverage the architectural flexibility of CNFs to adapt a powerful, size-transferable representation of atomic environments commonly used in machine learning interatomic potential architectures — specifically the *Equivariant Transformer* (ET) (Pelaez et al., 2024) as proposed in Hassan et al. (2024) — and consequently term our method RFM-ET in the following. Details on adaptations and hyperparameters are in Appendix A.5.

Density Estimation. As the chosen vector field parametrization naturally incorporates the toroidal topology and its outputs are directly in the Euclidean tangent space $\mathcal{T}_{\mathbf{x}}\mathcal{M} = \mathbb{R}^{3N}$, the Riemannian divergence simplifies to the standard Euclidean trace of the Jacobian, and we can use off-the-shelf ODE solvers, simply by projecting the final state back onto the torus. Due to the GNN's local cutoff, the computational cost of evaluating the CNF vector field scales linearly with system size,

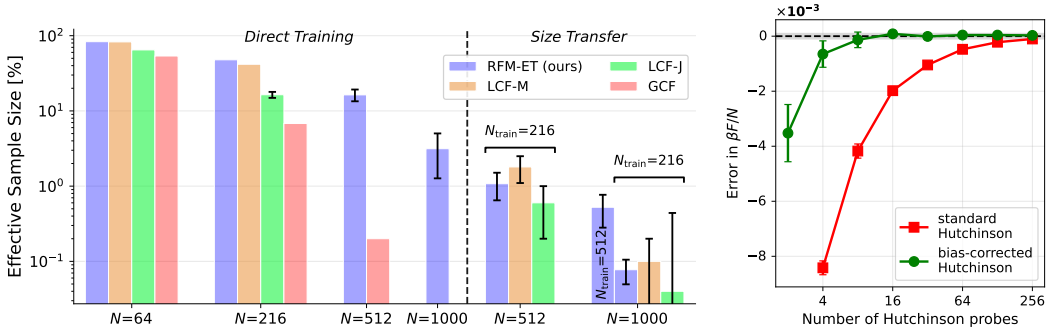

Figure 2: Left: Effective sample size computed for the models described in the text, including results for models evaluated on the same system size as used during training, as well as for models transferred to larger system sizes than those seen during training. Error bars are shown where visible. Right: The deviation to the reference free energy of a 512 particles mW system in the cubic ice phase, computed with a varying number of Hutchinson probes, with and without bias correction. The gray shaded area indicates a tolerance window of 10^{-4}. Errors for our method in all plots were estimated over 10k samples obtained from three models trained with different seeds.

$\mathcal{O}(N)$. However, calculating the exact divergence requires computing the trace of the full Jacobian. Because the input dimension D is proportional to N, obtaining this trace requires $\mathcal{O}(N)$ evaluations of the vector field. This results in an overall $\mathcal{O}(N^2)$ complexity for both computation and memory, making the exact calculation infeasible for large systems.

Hutchinson's trace estimator (Hutchinson, 1989) offers an $\mathcal{O}(N)$ alternative but is often considered unsuitable for Boltzmann sampling applications due to noise and statistical bias. Although Hutchinson's method provides an unbiased estimate of the vector field's divergence and thus the log-density change $\Delta \log p$, thermodynamic reweighting relies on the exponentiated log-density change. Due to Jensen's inequality (Jensen, 1906), the nonlinearity of the exponential function transforms the estimator's variance into a systematic bias in the importance weights. This bias propagates to all downstream equilibrium estimates, such as free energy differences. To address this, we propose a bias correction based on the second-order cumulant expansion (Zwanzig, 1954). By treating the stochastic error from the trace estimator as a fluctuating work term, we can expand the expectation of the likelihood estimate. Because the total accumulated noise arises from the sum of independent stochastic evaluations over multiple integration steps, its distribution rapidly converges to a Gaussian via the central limit theorem. Consequently, cumulants higher than the second order vanish, justifying the truncation of the series. For each sample, we estimate the variance of the stochastic log-density change, $\hat{\sigma}^2$, and apply a correction term directly to the log-weights:

$$\widehat{\log w_{\text{corr}}}(\mathbf{x}_0) \coloneqq \widehat{\log w}(\mathbf{x}_0) - \frac{1}{2}\hat{\sigma}^2(\mathbf{x}_0). \tag{7}$$

This effectively removes the bias introduced by the estimator noise, allowing for accurate thermodynamic estimates. To efficiently evaluate the correction term, we augment the ODE state with the empirical variance of the trace estimator, integrating it alongside the positions and log density. A more detailed discussion is provided in Appendix A.3 and Appendix A.4.

4 Experimental Results

We evaluate the performance of our method regarding both the quality of the generated samples and its accuracy as a free energy estimator. As a test system, we consider the common benchmark system cubic ice in the monoatomic water (mW) (Molinero & Moore, 2009; Wirnsberger et al., 2022) parametrization. Figure 1 (left and center) shows radial distribution functions and histograms of the potential energy as obtained from MD and RFM-ET for system sizes of 216 and 1000 particles. Importantly, even without any reweighting, excellent agreement is achieved for both observables, demonstrating that RFM-ET can accurately approximate the target Boltzmann distribution.

The strong performance of RFM-ET is further reflected in the high ESS values shown in the left panel of Fig. 2. We first evaluate the performance of our model in the non-transferable setting,

with all models assessed on the system size they were trained on. Where available, we compare to literature results obtained using a global transformer-based coupling flow (GCF) (Wirnsberger et al., 2022) and a local augmented coupling flow (LCF) (Schebek et al., 2025b); for the LCF, estimates based on joint (LCF-J) and marginal (LCF-M) densities are included. RFM-ET consistently matches or outperforms the others, demonstrating its high sampling quality. Notably, training on a system size of $N = 512$ yields a sampling efficiency two orders of magnitude higher than that of GCF, and, unlike the other models for which $N = 1000$ is computationally infeasible, RFM-ET still achieves an ESS above three percent on this system size. Importantly, RFM-ET significantly reduces training times and enables training on large system sizes, such as $N = 1000$, using the same computational budget that LCF requires to train on only $N = 216$ (Fig. 1, right).

A crucial feature of our method is its transferability to larger system sizes, allowing models trained on small systems to be evaluated on larger systems without requiring additional data or training, similarly to the LCF architecture. As shown in the center of Fig. 2, RFM-ET matches the sampling efficiency of LCF-M when transferring models trained on $N = 216$ to $N = 512$ and $N = 1000$, but only RFM-ET can be trained efficiently on $N = 512$, which allows increasing the sampling efficiency for $N = 1000$ by more than fivefold.

The right panel of Fig. 2 demonstrates the efficiency of the proposed bias correction by showing the relationship between the number of Hutchinson probes used per integration step and the free energy estimate obtained via TFEP. While the variance of the free energy estimates decreases rapidly with increasing number of probes, systematic bias remains the dominant source of error when no correction is applied. Even with 256 probes, the free energy estimated using biased weights is underestimated. In contrast, with as few as 16 probes, the bias-corrected weights allow the free energy estimates to converge to an accuracy of around 10^{-4}, enabling resolution of free energy differences on the order of 10^{-3} between hexagonal and cubic ice under the given settings (Schebek et al., 2025b). With fewer probes, the variance of the log-density change is underestimated, resulting in an insufficient bias correction.

5 Discussion

We showed that Riemannian Flow matching provides a powerful tool for equilibrium sampling of condensed matter, exhibiting superior scaling during training compared to coupling flows, enabling training on system sizes that were previously infeasible. Our method improves sampling on directly trained systems and enhances performance when models are transferred to larger systems. While the results presented here focus on cubic ice, we recently showed that the methodology is also applicable to other crystal structures within the mW potential and to systems described by the Lennard-Jones potential (Schebek et al., 2025a). As also noted in the same study, two primary challenges remain. First, inference cost is high due to the computationally expensive ODE integration and stochastic divergence estimation required for density evaluation, making it one to two orders of magnitude slower than sampling with coupling flows. Second, although flow-matching formulations are generally robust even with limited training data, the method still relies on high-quality configurations of the target distribution, preventing the amortization of training costs through generalization over external conditions (Schebek et al., 2024). A promising direction is data-free training of continuous-time models, which is currently being actively explored (Havens et al., 2025; Blessing et al., 2026). Future work could further explore the integration of consistency distillation (Geng et al., 2025) to enable few-step sampling without sacrificing the benefits of the continuous-time formulation, as well as the extension to the NPT ensemble and physically more realistic potentials.

Acknowledgments

MS acknowledges financial support from Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 “Scaling Cascades in Complex Systems”, Project Number 235221301, Project B08 “Multiscale Boltzmann Generators”. The Flatiron Institute is a division of the Simons Foundation. We thank Michael Plainer for fruitful discussions and for reviewing the manuscript.

REFERENCES

Rasool Ahmad and Wei Cai. Free energy calculation of crystalline solids using normalizing flows. *Modell. Simul. Mater. Sci. Eng.*, 30(6):065007, 2022. doi: 10.1088/1361-651X/ac7f4b. URL https://doi.org/10.1088/1361-651X/ac7f4b.

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.

Haim Avron and Sivan Toledo. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. *J. ACM*, 58(2), April 2011. doi: 10.1145/1944345.1944349. URL https://doi.org/10.1145/1944345.1944349.

Denis Blessing, Lorenz Richter, Julius Berner, Egor Malitskiy, and Gerhard Neumann. Bridge matching sampler: Scalable sampling via generalized fixed-point diffusion matching, 2026. URL https://arxiv.org/abs/2603.00530.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=g7ohDlTITL.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31, 2018.

Alessandro Coretti, Sebastian Falkner, Phillip L. Geissler, and Christoph Dellago. Learning mappings between equilibrium states of liquid systems using normalizing flows. *J. Chem. Phys.*, 162 (18):184102, 2025. doi: 10.1063/5.0253034.

Aaron Defazio, Xingyu Yang, Harsh Mehta, Konstantin Mishchenko, Ahmed Khaled, and Ashok Cutkosky. The Road Less Scheduled. *Advances in Neural Information Processing Systems*, 37: 9974–10007, 2024.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real NVP. In *International Conference on Learning Representations (ICLR)*, 2017.

Peter Eastman, Raimondas Galvelis, Raúl P. Peláez, Charlles R. A. Abreu, Stephen E. Farr, Emilio Gallicchio, Anton Gorenko, Michael M. Henry, Frank Hu, Jing Huang, Andreas Krämer, Julien Michel, Joshua A. Mitchell, Vijay S. Pande, João PGLM Rodrigues, Jaime Rodriguez-Guerra, Andrew C. Simmonett, Sukrit Singh, Jason Swails, Philip Turner, Yuanqing Wang, Ivy Zhang, John D. Chodera, Gianni De Fabritiis, and Thomas E. Markland. Openmm 8: Molecular dynamics simulation with machine learning potentials. *The Journal of Physical Chemistry B*, 128(1): 109–116, December 2023. doi: 10.1021/acs.jpcb.3c06662. URL http://dx.doi.org/10.1021/acs.jpcb.3c06662.

William Falcon and The PyTorch Lightning team. PyTorch Lightning, 2019. URL https://github.com/Lightning-AI/lightning.

Matthias Fey, Jinu Sunil, Akihiro Nitta, Rishi Puri, Manan Shah, Blaz Stojanovic, Ramona Bendias, Alexandria Barghi, Vid Kocijan, Zecheng Zhang, Xinwei He, Jan Eric Lenssen, and Jure Leskovec. PyG 2.0: Scalable Learning on Real World Graphs. *Temporal Graph Learning Workshop @ KDD*, 2025.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T. H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. POT: Python Optimal Transport. *J. Mach. Learn. Res*, 22(78):1–8, 2021.

Daan Frenkel and Anthony J. C. Ladd. New Monte Carlo method to compute the free energy of arbitrary solids. Application to the fcc and hcp phases of hard spheres. *J. Chem. Phys.*, 81(7): 3188–3193, 1984. doi: 10.1063/1.448024.

Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 3rd edition, 2023.

Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian geometry*, volume 2. Springer, 1990.

Zhengyang Geng, Yiyang Lu, Zongze Wu, Eli Shechtman, J. Zico Kolter, and Kaiming He. Improved mean flows: On the challenges of fastforward generative models, 2025. URL https://arxiv.org/abs/2512.02012.

Louis Grenioux, Leonardo Galliano, Ludovic Berthier, Giulio Biroli, and Marylou Gabrié. Riemannian stochastic interpolants for amorphous particle systems, 2025. URL https://arxiv.org/abs/2512.16607.

Majdi Hassan, Nikhil Shenoy, Jungyoon Lee, Hannes Stärk, Stephan Thaler, and Dominique Beaini. ET-Flow: Equivariant Flow-Matching for Molecular Conformer Generation. *Advances in Neural Information Processing Systems*, 37:128798–128824, 2024.

Aaron J Havens, Benjamin Kurt Miller, Bing Yan, Carles Domingo-Enrich, Anuroop Sriram, Daniel S. Levine, Brandon M Wood, Bin Hu, Brandon Amos, Brian Karrer, Xiang Fu, Guan-Horng Liu, and Ricky T. Q. Chen. Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. 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. 22204–22237. PMLR, 2025. URL https://proceedings.mlr.press/v267/havens25a.html.

M. F. Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. 1989.

C. Jarzynski. Targeted free energy perturbation. *Physical Review E*, 65(4):046122, 2002. doi: 10.1103/PhysRevE.65.046122.

J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30:175–193, 1906. doi: 10.1007/BF02418571.

L. Kish. Sampling organizations and groups of unequal sizes. *American Sociological Review*, 30: 564–572, August 1965. ISSN 0003-1224.

Leon Klein and Frank Noe. Transferable Boltzmann generators. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Leon Klein, Andreas Krämer, and Frank Noé. Equivariant flow matching. In *Advances in Neural Information Processing Systems*, 2023.

Jonas Köhler, Leon Klein, and Frank Noé. Equivariant flows: Exact likelihood generative learning for symmetric densities. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *ICML'20*, pp. 5361–5370. JMLR.org, 2020.

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.

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*, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

Emile Mathieu and Maximilian Nickel. Riemannian continuous normalizing flows. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

Valeria Molinero and Emily B. Moore. Water Modeled As an Intermediate Element between Carbon and Silicon. *The Journal of Physical Chemistry B*, 113(13):4008–4016, 2009. doi: 10.1021/jp805227c.

Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457), 2019. doi: 10.1126/science.aaw1147.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.*, 22(1):57:2617–57:2680, January 2021.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, number 721, pp. 8026–8037. 2019.

Raul P. Pelaez, Guillem Simeon, Raimondas Galvelis, Antonio Mirarchi, Peter Eastman, Stefan Doerr, Philipp Thölke, Thomas E. Markland, and Gianni De Fabritiis. TorchMD-Net 2.0: Fast Neural Network Potentials for Molecular Simulations. *J. Chem. Theory Comput.*, 20(10):4076–4087, 2024. doi: 10.1021/acs.jctc.4c00253.

J. M. Polson, E. Trizac, S. Pronk, and D. Frenkel. Finite-size corrections to the free energies of crystalline solids. *J. Chem. Phys.*, 112(12):5339–5342, 2000. doi: 10.1063/1.481102.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample flow matching: straightening flows with minibatch couplings. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Danyal Rehman, Tara Akhound-Sadegh, Artem Gazizov, Yoshua Bengio, and Alexander Tong. Falcon: Few-step accurate likelihoods for continuous flows, 2025. URL `https://arxiv.org/abs/2512.09914`.

Danilo Rezende and Shakir Mohamed. Variational Inference with Normalizing Flows. In *International conference on machine learning*, pp. 1530–1538. PMLR, 2015.

Victor Garcia Satorras, Emiel Hoogeboom, Fabian Bernd Fuchs, Ingmar Posner, and Max Welling. E(n) equivariant normalizing flows. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021.

Maximilian Schebek, Michele Invernizzi, Frank Noé, and Jutta Rogal. Efficient mapping of phase diagrams with conditional Boltzmann Generators. *Mach. Learn.: Sci. Technol.*, 5(4):045045, 2024. doi: 10.1088/2632-2153/ad849d.

Maximilian Schebek, Jiajun He, Emil Hoffmann, Yuanqi Du, Frank Noé, and Jutta Rogal. Assessing generative modeling approaches for free energy estimates in condensed matter. arXiv preprint arXiv:2512.23930, 2025a. `https://arxiv.org/abs/2512.23930`.

Maximilian Schebek, Frank Noé, and Jutta Rogal. Scalable Boltzmann Generators for equilibrium sampling of large-scale materials. arXiv preprint arXiv:2509.25486, 2025b. `https://arxiv.org/abs/2509.25486`.

K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.*, 148(24):241722, 2018. doi: 10.1063/1.5019779.

Michael R. Shirts and John D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, 129(12):124105, September 2008. ISSN 0021-9606. doi: 10.1063/1.2978177.

Frank H. Stillinger and Thomas A. Weber. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B*, 31(8):5262–5271, 1985. doi: 10.1103/PhysRevB.31.5262.

Alexander Tong, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Kilian FATRAS, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023. URL https://openreview.net/forum?id=HgDwiZrpVq.

Peter Wirnsberger, Andrew J. Ballard, George Papamakarios, Stuart Abercrombie, Sébastien Racanière, Alexander Pritzel, Danilo Jimenez Rezende, and Charles Blundell. Targeted free energy estimation via learned mappings. *J. Chem. Phys.*, 153(14):144112, 2020. doi: 10.1063/5.0018903.

Peter Wirnsberger, George Papamakarios, Borja Ibarz, Sébastien Racanière, Andrew J Ballard, Alexander Pritzel, and Charles Blundell. Normalizing flows for atomic solids. *Mach. Learn.: Sci. Technol.*, 3(2):025009, 2022. doi: 10.1088/2632-2153/ac6b16.

Robert W. Zwanzig. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.*, 22(8):1420–1426, 1954. doi: 10.1063/1.1740409.

A APPENDIX

A.1 GEODESIC ON THE FLAT TORUS

Let $\mathbf{x}_0, \mathbf{x}_1 \in \mathbb{R}^3$ be two points inside a periodic simulation box. The *shortest vector* connecting them under periodic boundary conditions corresponds to the geodesic, which can defined by the log map $\log_{\mathbf{x}_0}(\mathbf{x}_1) \in \mathbb{R}^3$ with corresponding distance $d = |\log_{\mathbf{x}_0}(\mathbf{x}_1)| \in \mathbb{R}$ (Gallot et al., 1990) The concrete form of this log map depends on the shape of the simulation box (Frenkel & Smit, 2023):

General triclinic box Let $\mathbf{A} \in \mathbb{R}^{3\times 3}$ be the matrix of lattice vectors,

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3], \quad \mathbf{a}_i \in \mathbb{R}^3. \tag{8}$$

The log map is then

$$\mathbf{s} = \mathbf{A}^{-1}(\mathbf{x}_1 - \mathbf{x}_0) \in \mathbb{R}^3, \tag{9}$$

$$\mathbf{s}_{\min} = \mathbf{s} - \left\lfloor \mathbf{s} + \frac{1}{2} \right\rfloor \in \mathbb{R}^3, \tag{10}$$

$$\log_{\mathbf{x}_0}(\mathbf{x}_1) = \mathbf{A}\,\mathbf{s}_{\min} \in \mathbb{R}^3. \tag{11}$$

Throughout, $\lfloor \cdot \rfloor$ denotes the floor function applied component-wise.

Orthorhombic box Let $\mathbf{L} = (L_x, L_y, L_z) \in \mathbb{R}^3$ contain the box lengths along each axis. Then the log map reads

$$\log_{\mathbf{x}_0}(\mathbf{x}_1) = (\mathbf{x}_1 - \mathbf{x}_0) - \mathbf{L} \cdot \left\lfloor \frac{\mathbf{x}_1 - \mathbf{x}_0}{\mathbf{L}} + \frac{1}{2} \right\rfloor \in \mathbb{R}^3, \tag{12}$$

where division and multiplication are understood component-wise.

Cubic box For a cubic box of side $L \in \mathbb{R}$, the formula simplifies to

$$\log_{\mathbf{x}_0}(\mathbf{x}_1) = (\mathbf{x}_1 - \mathbf{x}_0) - L \left\lfloor \frac{\mathbf{x}_1 - \mathbf{x}_0}{L} + \frac{1}{2} \right\rfloor \in \mathbb{R}^3. \tag{13}$$

A.2 SYSTEMS

A.2.1 $3N$-DIM GAUSSIAN PRIOR: EINSTEIN-CRYSTAL

The $3N$-dimensional Gaussian prior employed in this work corresponds to the configurational distribution of the so-called *Einstein crystal*, i.e., a system of N independent harmonic oscillators with spring constant Λ_E, centered on equilibrium lattice positions $\{\mathbf{X}^i\}_{i=1}^N$ and confined to a box of volume V. The corresponding potential energy is given by

$$U_{\mathrm{id}}(\mathbf{x}) = \Lambda_E \sum_{i=1}^{N} \left|\mathbf{x}^i - \mathbf{X}^i\right|^2. \tag{14}$$

The partition function for this system can be analytically calculated and thus also its absolute free energy. Assuming a fixed center of mass, one obtains

$$\frac{\beta F_0}{N} = \frac{1}{N} \ln\left(\frac{N\Lambda^3}{V}\right) + \frac{3}{2}\left(1 - \frac{1}{N}\right) \ln\left(\frac{\beta \Lambda_E \Lambda^2}{\pi}\right) - \frac{3}{2N} \ln N, \tag{15}$$

where Λ is the de Broglie wavelength, which is commonly set to a characteristic length scale of the interaction potential in model systems. The second term accounts for the constraint of fixing the center of mass (Frenkel & Ladd, 1984).

By computing the free energy difference between the Einstein Crystal and any arbitrary system, one can obtain the absolute free energy of the target system, which allows ranking different phases of a given system by stability. We follow Wirnsberger et al. (2022) by setting the width of the Gaussians, which is related to Λ_E, similar to the width of the displacements of the atoms around their equilibrium positions, with the exact widths provided in Tab. 1. For us, this ensures that the vector field always acts on a similar environment for each time t. The widths for the different systems are collected in Tab. 1.

A.2.2 MONOATOMIC-WATER

The monoatomic water is a special case of the Stillinger-Weber (SW) potential (Stillinger & Weber, 1985). It incorporates two-body terms (ϕ_2) as well as three-body (ϕ_3) interaction terms, where the latter promote tetrahedral coordination environments. The total potential energy is expressed as:

$$U_{\mathrm{SW}}(\mathbf{x}) = \sum_i \sum_{j>i} \phi_2(d_{ij}) + \lambda_3 \sum_i \sum_{j\neq i} \sum_{k>j} \phi_3(d_{ij}, d_{ik}, \theta_{ijk}), \tag{16}$$

where the individual interaction terms are defined as:

$$\phi_2(r) = A\varepsilon \left[B \left(\frac{\sigma}{r}\right)^4 - 1 \right] \exp\left(\frac{\sigma}{r - a\sigma}\right), \tag{17}$$

$$\phi_3(r, s, \theta) = \lambda\varepsilon(\cos\theta - \cos\theta_0)^2 \cdot \exp\left(\frac{\gamma\sigma}{r - a\sigma}\right) \cdot \exp\left(\frac{\gamma\sigma}{s - a\sigma}\right). \tag{18}$$

Here, d_{ij} represents the distance between particles i and j, and θ_{ijk} is the bond angle formed by the triplet of atoms i, j, and k. Most parameters in the SW potential are held fixed, with only ε, σ, and λ_3 being adjustable. The parameter λ_3 controls the strength of the three-body interactions, while ε and σ define the characteristic energy and length scales, respectively. The specific parameter values employed in this work follow those established in Molinero & Moore (2009), effectively modeling a coarse-grained representation of water.

Simulations were performed with the same settings as in Wirnsberger et al. (2022) and Schebek et al. (2025b) in the NVT ensemble using the openMM package (Eastman et al., 2023). The mW ice was modeled at $T = 200\,\mathrm{K}$ at a density of $\rho = 1.004\,\mathrm{g\,cm^{-3}}$.

A.3 SECOND-ORDER CUMULANT EXPANSION FOR HUTCHINSON'S ESTIMATOR

To scale to large systems, we circumvent the effective $\mathcal{O}(N^2)$ complexity of the exact divergence calculation by employing Hutchinson's trace estimator. At each integration step t, we approximate the trace using K stochastic noise vectors $\{\boldsymbol{\epsilon}_k\}_{k=1}^K$ with Rademacher entries, i.e., $\epsilon_{k,i} \in \{-1, +1\}$ i.i.d. with equal probability 1/2:

$$\mathrm{Tr}(\nabla v_\theta) \approx \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{\epsilon}_k^T \nabla v_\theta \boldsymbol{\epsilon}_k. \tag{19}$$

We denote the stochastic estimate of the log-density change obtained via this estimator as $\widehat{\Delta \log p}$. This estimate is *unbiased*, meaning $\mathbb{E}[\widehat{\Delta \log p}] = \Delta \log p$. This results in unbiased *log* importance weights:

$$\widehat{\log w}(\mathbf{x}_0) = \beta_p U_p(\mathbf{x}_0) - \beta_\mu U_\mu(f_\theta(\mathbf{x}_0)) - \widehat{\Delta \log p}. \tag{20}$$

However, the computation of thermodynamic observables, such as the free energy difference, depends on the *exponentiated* log weights. By applying Jensen's inequality to the expectation over the Hutchinson probes $\mathbb{E}_{\boldsymbol{\epsilon}\sim p(\boldsymbol{\epsilon})}$, we see that:

$$\mathbb{E}_{\boldsymbol{\epsilon}\sim p(\boldsymbol{\epsilon})}[\hat{w}(\mathbf{x}_0)] = C(\mathbf{x}_0) \cdot \mathbb{E}_{\boldsymbol{\epsilon}\sim p(\boldsymbol{\epsilon})}[\exp(\widehat{\Delta \log p})] \geq C(\mathbf{x}_0) \cdot \exp(\mathbb{E}_{\boldsymbol{\epsilon}\sim p(\boldsymbol{\epsilon})}[\widehat{\Delta \log p}]) = w(\mathbf{x}_0), \tag{21}$$

where $C(\mathbf{x}_0)$ contains the deterministic potential energy terms. This inequality implies that for any given sample $\mathbf{x}_0$, the stochastic weight is systematically overestimated on average. This bias propagates to the free energy estimator, which involves a second expectation over the samples $\mathbf{x}_0$ drawn from the prior $p(\mathbf{x}_0)$. The estimator for the exponential of the free energy difference behaves as follows:

$$\mathbb{E}_{\mathbf{x}_0\sim p(\mathbf{x}_0)} \left[\mathbb{E}_{\boldsymbol{\epsilon}\sim p(\boldsymbol{\epsilon})}[\hat{w}(\mathbf{x}_0)] \right] \geq \mathbb{E}_{\mathbf{x}_0\sim p(\mathbf{x}_0)}[w(\mathbf{x}_0)] = e^{-\beta\Delta F}. \tag{22}$$

Consequently, taking the negative logarithm reverses the inequality, showing that unbiased stochastic trace estimation leads to a systematic underestimation of the free energy difference:

$$\beta\Delta\hat{F} := -\log \mathbb{E}_{\mathbf{x}_0\sim p(\mathbf{x}_0), \boldsymbol{\epsilon}\sim p(\boldsymbol{\epsilon})}[\hat{w}(\mathbf{x}_0)] \leq -\log \mathbb{E}_{\mathbf{x}_0\sim p(\mathbf{x}_0)}[w(\mathbf{x}_0)] = \beta\Delta F. \tag{23}$$

To address this, we treat the stochastic error from the trace estimator as a fluctuating work term and apply the *second-order cumulant expansion*, a method rooted in thermodynamic perturbation

theory (Zwanzig, 1954). The expected value of the likelihood estimate can be expressed via the cumulant generating function expansion:

$$\log \mathbb{E}[e^{\widehat{\Delta \log p}}] = \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} \tag{24}$$

$$= \mathbb{E}_{\boldsymbol{\epsilon}\sim p(\boldsymbol{\epsilon})}[\widehat{\Delta \log p}] + \frac{1}{2}\left(\mathbb{E}_{\boldsymbol{\epsilon}\sim p(\boldsymbol{\epsilon})}[(\widehat{\Delta \log p})^2] - \mathbb{E}_{\boldsymbol{\epsilon}\sim p(\boldsymbol{\epsilon})}[\widehat{\Delta \log p}]^2\right) + \ldots \tag{25}$$

We truncate this series after the second order. This approximation is justified by the Central Limit Theorem: because the total accumulated noise arises from the sum of independent stochastic evaluations over many integration steps, its distribution converges rapidly to a Gaussian, even if the individual noise terms are non-Gaussian. Since all cumulants κ_n with $n > 2$ vanish for a Gaussian distribution, we correct the bias in the log-weights using the accumulated variance $\hat{\sigma}^2$ of the trace estimator along the trajectory. Since each sample $\mathbf{x}_0$ experiences a unique realization of the stochastic noise, both the log-density estimate $\widehat{\Delta \log p}$ and the variance correction $\hat{\sigma}^2$ are sample-specific:

$$\widehat{\log w_{\text{corr}}}(\mathbf{x}_0) \coloneqq \widehat{\log w}(\mathbf{x}_0) - \frac{1}{2}\hat{\sigma}^2(\mathbf{x}_0). \tag{26}$$

The total variance $\hat{\sigma}^2$ is estimated by accumulating the local variance of the K probes during the integration, as detailed in Appendix A.4.

It is worth noting that for very few integration steps (e.g., $N < 5$), the Gaussian assumption may not hold. In such regimes, the noise follows the specific distribution of the Hutchinson estimator. When using Rademacher probes ($\epsilon \in \{-1, 1\}$), the estimator possesses desirable sub-Gaussian concentration properties and minimizes the variance for specific matrix classes compared to Gaussian probes (Avron & Toledo, 2011). Importantly, the Rademacher distribution is symmetric, implying the skewness (κ_3) is zero. Therefore, if the Gaussian approximation fails due to a low step count, the leading error term in our expansion would be determined by the fourth cumulant κ_4 (kurtosis), rather than the third.

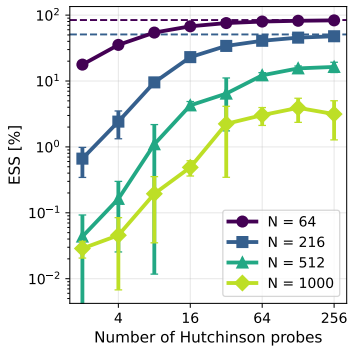

Figure 3: ESS against the number of Hutchinson probes used at each integration step. Dotted lines correspond to values obtained with exact divergence calculations.

Finally, the stochasticity of the trace estimator directly impacts sampling efficiency. Since the importance weights inherit the noise from the $\Delta \log p$ estimate, the ESS inevitably degrades when fewer probes are used. Figure 3 demonstrates that with a sufficient number of probes, the ESS converges to the baseline value obtained from exact divergence calculations. It is crucial to note, however, that the validity of our correction relies on an accurate estimate of the noise variance $\hat{\sigma}^2$. Using too few probes may yield an unreliable variance estimate, which can distort the correction term itself and result in residual bias. We note that the ESS is negligibly affected by the bias-correction.

A.4 ODE Integration

We perform the integration of the CNF dynamics using a fixed-step 4th-order Runge-Kutta (RK4) solver. Adaptive step-size solvers are ill-suited for this application, as the stochastic nature of the Hutchinson trace estimator introduces high-frequency noise into the derivative of the log-density. Adaptive solvers typically interpret this statistical variance as local truncation error, leading to excessive and inefficient reduction of the step size. One could, however, use them for the positions only, but we found 10 RK4 steps to be sufficient in eliminating numerical integration errors, both in the positions and exact divergence calculations.

To estimate the accumulated variance σ^2 required for the cumulant correction without the computational cost of repeated trajectory realizations, we further augment the ODE state. We compute the empirical variance of the trace estimator, $\widehat{\text{Var}}[\widehat{\text{Tr}}]$, efficiently at each evaluation using K probes and

integrate it alongside the density:

$$\frac{d}{dt}\begin{bmatrix}\mathbf{x}\\ \Delta\log p\\ \mathcal{V}\end{bmatrix}=\begin{bmatrix}v_\theta(\mathbf{x},t)\\ -\widehat{\mathrm{Tr}}(\nabla v_\theta)\\ \widehat{\mathrm{Var}}[\widehat{\mathrm{Tr}}(\nabla v_\theta)]\end{bmatrix}. \tag{27}$$

By reusing the vector-Jacobian products computed for the trace, the overhead of computing the variance derivative is negligible.

However, the raw accumulated variance output $\mathcal{V}(T)$ must be corrected to account for the noise reduction inherent in the RK4 scheme. An RK4 step updates the log-density using a weighted average of derivatives k_i evaluated at four stages:

$$\log p_{t+h}=\log p_t+\frac{h}{6}(k_1+2k_2+2k_3+k_4), \tag{28}$$

where h is the step size. Since the Hutchinson noise vectors are resampled at every stage, the noise terms in k_i are independent random variables with local variance σ_t^2. The variance of the update step is therefore the sum of the variances of the weighted terms:

$$\mathrm{Var}(\Delta\log p_{\text{step}})=\left(\frac{h}{6}\right)^2\sum_{i=1}^{4}a_i^2\mathrm{Var}(k_i) \tag{29}$$

$$\approx\frac{h^2}{36}(1^2+2^2+2^2+1^2)\sigma_t^2 \tag{30}$$

$$=\frac{10}{36}h^2\sigma_t^2=\frac{5}{18}h^2\sigma_t^2. \tag{31}$$

The augmented ODE integrator accumulates the scalar rate, returning $\mathcal{V}(T)\approx\sum_{steps}\sigma_t^2 h$. To obtain the true variance of the path integral, we apply the correction:

$$\sigma^2=\mathcal{V}(T)\cdot h\cdot\frac{5}{18}. \tag{32}$$

A.5 Implementation details

We used the equivariant transformer from Pelaez et al. (2024) and followed Hassan et al. (2024) by replacing the scalar output head with the equivariant output head from SchNet (Schütt et al., 2018) and using layer norms. Additionally, we added a bigger time dimension, removed the molecular features such as the bond specification and replaced the SMILES labeling by a sinusoidal encoding, based on the equilibrium position of an atom in its unit cell, as proposed in Schebek et al. (2025b). The hyperparameteres are collected in Tab. 3. For optimization the schedule free version of the AdamW optimizer (Defazio et al., 2024), with the hyperparameters collected in Tab. 2.

The work was implemented in `python`, using `pytorch` (Paszke et al., 2019), `lightning` (Falcon & The PyTorch Lightning team, 2019), `pytorch-geometric` (Fey et al., 2025), and `POT` (Flamary et al., 2021). To avoid waiting for mini-batch optimal transport reordering during training, it was reordered on the CPU simultaneously with training on the GPU. The code is available under `https://github.com/emil-ho/bg-cnf-fm-condensed-matter`.

Table 1: System-dependent Hyperparameters

Parameter	$N = 64$	$N = 216$	$N = 512$	$N = 1000$
Batch size	128	128	128	128
Mini-Batch OT size	128	128	64	32
Train steps	800 K	1 M	1 M	1 M
Prior width [Å]	0.1631	0.174	0.179	0.183

Table 2: Optimizer Hyperparameters

Parameter	**Value**
Gradient clipping	0.05
β_1	0.9
β_2	0.999
ϵ	10^{-8}
Learning rate	0.0003
Warmup steps	500
Weight decay	0

Table 3: Vector Field Architecture

Parameter	**Value**
Trainable params	130 K
Hidden channels	24
Attention heads	3
Layers	12
Unit cell embedding	1
Time embedding	10
RBF features	16
Activation	SiLU