Continuous-Time Generative Models for Sampling Unnormalized Densities

Weilong Chen[1]
[1]Department of Mathematics, Chalmers University of Technology, Sweden
weilong@chalmers.se

Abstract

This research proposal is submitted as part of the application for the Adaptive and Neural Computation PhD program at the University of Edinburgh. It centers on using and improving continuous-time generative models to address the challenge of sampling problem in statistical physics. This proposal incorporates recent advancements in this domain, the author's relevant experience, and outlines a preliminary plan for the PhD study.

1 Introduction

In statistical mechanics, computing the average behavior of microscopic states is crucial. For instance, in molecular systems, the computation of statistics [1], known as observables, for equilibrium states is essential. The distribution of equilibrium states follows the Boltzmann distribution, which can be utilized, for example, to calculate the probability of protein folding under certain conditions.

The challenge lies in the difficulty of sampling, as the unnormalized density is known but hard to sample from. Typically, sampling is performed using molecular dynamics [2] (MD), which faces challenges in obtaining independent and identically distributed (i.i.d) samples due to rare events. This makes MD computationally expensive and impractical for large systems with real-world applications. Extensive research has been conducted to address this issue, known as the enhanced sampling methods [3] within the MD community.

Machine learning offers new perspectives for tackling this problem. One notable approach is the Boltzmann generator [4], which employs generative models to generate independent samples in one-shot. Additionally, the Boltzmann generator can be further applied together with reweighting methods, where a rich set of tools is provided.

The rapidly growing field of deep learning, particularly deep generative models, presents new opportunities to design efficient sampling methods for many-body systems [5]. This provides a powerful tool for advancing molecular simulation, thereby benefiting various applications such as drug discovery, material science, catalysis, and green chemistry.

2 Background and Preliminaries

Molecular dynamcis and observables MD stands as a widely used computational technique in chemistry and physics. It involves modeling the time evolution of N-particle systems, where states are denoted by $\mathbf{x} \in \Omega \subset \mathbb{R}^d$ with dimensionality $d \in \mathbb{N}$. These systems follow stochastic differential equations [6] (SDE), wherein the drift term is determined by the potential energy function $E(\mathbf{x}) : \Omega \to \mathbb{R}$. The Boltzmann distribution associated with the energy function E is expressed as follows:

$$\rho(\mathbf{x}) = \frac{1}{Z} \exp(-\beta E(\mathbf{x})), \quad Z = \int_{\Omega} \exp(-\beta E(\mathbf{x})) d\mathbf{x}, \tag{1}$$

where $\beta^{-1} = kT$, with k representing the Boltzmann constant and T indicating the system's temperature. The constant Z is commonly known as the *partition function* and is computed by integrating over all states $\mathbf{x}$ of dimensionality.

After obtaining MD results, insights into the properties of a molecular system can be gained through the calculation of observables [7]. These observables can be categorized into two types:

1. **Stationary observables:** $O_f = \mathbb{E}_\mu [f(\mathbf{x})] = \int_\Omega f(\mathbf{x})\rho(\mathbf{x})d\mathbf{x}$

Here, μ denotes the normalized Gibbs measure, $f : \Omega \to \mathbb{R}$ represents a function describing a microscopic observation process corresponding to a specific quantity of interest, e.g. distance or angle, and ρ is defined through Equation 1.

2. **Dynamic observables:** $O_{f(t)h(t+\Delta t)} = \mathbb{E}_{\mathbf{x}_t \sim \mu} \left[\mathbb{E}_{\mathbf{x}_{t+\Delta t} \sim \rho_\tau(\mathbf{x}_{t+\Delta t}|\mathbf{x}_t)} [f(\mathbf{x}_t)h(\mathbf{x}_{t+\Delta t})] \right]$

$f, h : \Omega \to \mathbb{R}$ are functions that describe processes corresponding to the dynamic observable of interest, such as binding affinities and binding rates of a drug. $\rho_\tau(\mathbf{x}_{t+\Delta t}|\mathbf{x}_t)$ represents the conditional probability density function for the time-discrete evolution of the state vector $\mathbf{x}$ at $\Delta t = N\tau$ steps into the future, given an initial state $\mathbf{x}_t$. Here, $N \in \mathbb{N}$ and $\tau, t \in \mathbb{R}$.

Conventionally, these observables are estimated from simulation trajectories using simple Monte Carlo estimators [8]. Typical method involves iteratively solving the SDE to obtain a trajectory of the states the system has visited over time. However, these samples will not be i.i.d because each new sample will be correlated with the previous one. Additionally, many molecules have complex energy landscapes with barriers [9] that have very low probabilities of transition between modes of their corresponding probability distributions, known as *rare events*. These factors collectively make MD simulations extremely challenging and costly.

Continuous-time generative models Generative models are being widely discussed today because of applications such as autoregressive models like LLM [10], image-generation GAN [11] models, and latent variables models like VAE [12]. Each model has its own advantages and drawbacks across various application domains. In this proposal, we will specifically focus on continuous-time generative models. These models can be defined using SDEs or, equivalently, corresponding Probability Flow Ordinary Differential Equations (PF-ODEs).

A possible way of modelling continuous density transformations is to exploit NeuralODE [13]. By defining an ODE that describes the flow of a time-dependent probability density, such that it corresponds to the prior distribution at t_0 and to the data distribution at t_1, samples can be transformed in a continuous manner, called *continuous normalizing flows* (CNFs) [13]. Assume initial states $\mathbf{x}_0 \in \Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, which follow an initial distribution ρ_0 and time-dependent states $\mathbf{x}_t \in \Omega$, which follow as time-dependent distribution ρ_t. Further, assume that each $\mathbf{x}_t$ obeys the ordinary differential equation

$$\frac{d}{dt}\mathbf{x}_t = \mathbf{b}(\mathbf{x}_t), \quad \mathbf{x}_{t=0} = \mathbf{x}_0, \tag{2}$$

where the vector field $\mathbf{b} : [0,\, 1] \times \Omega' \to \mathbb{R}^d,\ d \in \mathbb{R}$, describes the velocity of $\mathbf{x}_t$. Then, the continuity equation describes the evolution of the time-dependent particle density $\rho_t(\mathbf{x})$ where

$$\frac{\partial}{\partial t}\rho_t + \nabla \cdot (\rho_t \mathbf{b}) = 0. \tag{3}$$

By replacing $\mathbf{b}$ with a learned velocity field $\mathbf{b}_\theta$, samples can be generated in a continuous manner by integrating Equation (2) from initial conditions $\mathbf{x}_0 \sim \rho_{t=0}$ using a ODE solver. Figure 1 below shows an example of a CNF transformation of an initial density ρ_0 into an endpoint density ρ_1.

Inspired by non-equilibrium statistical mechanics, the *denoising diffusion probabilistic models* [14] (DDPM) learns to invert a deterministic noise process. The transformation of data into a simple noise distribution can be achieved using a continuous-time SDE. This SDE can be reversed when the score of the distribution at each intermediate time step is known, denoted by $\nabla_x \log \rho_t(x)$, which is learned through neural networks.

Most importantly, the diffusion model, coupled with score matching [15] objectives, presents an important idea: the mapping of a simple distribution ρ_0 to another distribution ρ_1 via a flexible pathway. This stands in contrast to maximum likelihood training for CNFs, which requires continuous simulation of ODE during training. Diffusion specifies the pathway as a diffusion process, thereby

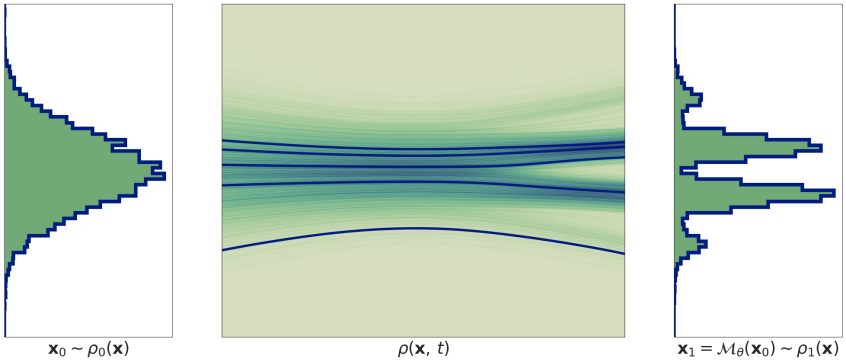

Figure 1: Illustration of how an initial density ρ_0 continuously transforms into an endpoint density ρ_1 such that $\mathbf{x}_1 = \mathcal{M}(\mathbf{x}_0) \sim \rho_1(\mathbf{x})$.

offering a simulation-free training paradigm. This notion is highly inspiring. With this simulation-free training approach, several related works, such as *flow matching* [16] and *rectified flow* [17], have emerged. A unified framework bridging diffusion and flow models is proposed and named *stochastic interpolants* [18].

Enhanced sampling methods Various methods have been proposed to optimize sampling processes. Generally, these methods fall into three main branches: biasing methods, path sampling methods, and generalized ensemble methods.

Biasing methods seek to identify and learn lower-dimensional manifolds that capture the slow or dominant patterns of system dynamics. Typical approaches include VAE [19] and the Variational Approach to Conformational Dynamics [20], which are integrated into a general framework for dimension reduction in dynamical systems.

Path sampling methods strategically start multiple short parallel simulations in regions that will lead to a more accurate description of the ensemble. They generally take advantage of analysis methods such as Markov State Models (MSM) [21] to combine the statistics and kinetics of these simulations. There are different ways to choose adaptive seeding strategies, and one possible way is to use reinforcement learning by setting up a reward function.

Generalized ensemble methods sample the system across multiple thermodynamic ensembles simultaneously or sequentially. Each ensemble has its own set of thermodynamic parameters. One intuitive idea is to transition to a higher temperature ensemble, which can accelerate barrier crossing, improving the exploration of energy landscapes and the sampling of rare events.

Many methods combine multiple of these classes at the same time [22]. Continuous-time generative models cannot solve the sampling problem independently for now; they need to be coupled with existing methods to enhance the sampling [3].

3 Research Questions

Extensions of Flow Matching While the flow matching framework is relatively recent [16], it has seen significant development through various proposed extensions. A noteworthy contribution to this field is the Stochastic Interpolant framework [18], which offers a comprehensive foundation for future investigations, both theoretically and in practical applications.

One central topic is optimal transport [23] and Schrödinger bridges [24]. By solving corresponding problems for the flow field or score field, sampling from joint distribution (ρ_0, ρ_1) can be optimized, leading to more stable models for training and faster inference. Iterative algorithms [25] have been proposed to obtain the exact Schrödinger bridge, while with the drawback of possibly accumulating errors. Direct approximation of the continuous-time Schrödinger bridge without simulation [26] can be achieved by computing the entropic optimal transport plan via efficient algorithms.

Another important topic is generalizing to other spaces and data types. In many scientific domains, the distributions of interest are supported on *non-Euclidean* spaces. Riemannian Flow Matching [27], a relative framework, was proposed for training flow matching on manifolds, where further applications can be explored. Also, in practice, we often deal with *discrete data* like molecules and proteins. Discrete Flow Models [28] were proposed, which use Continuous Time Markov Chains to achieve the discrete equivalent of continuous space flow matching. Dirichlet flow matching [29] uses mixtures of Dirichlet distributions as probability paths and applies them for generative modeling of discrete categorical data and sequences, e.g., DNA sequences. Further research can be conducted based on the above work on biological sequence generation and discrete data applications.

Problem of Amortized Sampling To train continuous-time generative models for sampling unnormalized density, typically, we require a sufficient number of initial samples that cover most of the modes, which can be quite restrictive. Moreover, while flow matching supports mapping between arbitrary densities [18], both the initial and target distributions necessitate high-quality data, with initial samples often relying on Markov Chain Monte Carlo and MD methods. Therefore, there is a need, and practical benefit, in learning in the absence of extensive data.

The Boltzmann generator [4] needs sufficient data for training, while learning-based neural samplers [30] like the Path Integral Sampler [31] and Denoising Diffusion Sampler [32] attempt to learn in the absence or scarcity of data. However, these samplers require simulation during learning, which makes scaling challenging. Flow AIS Bootstrap (FAB) [33] is another approach that combines Annealed Importance Sampling samples with (equivariant) normalizing flow training. FAB involves computing the model likelihood in its loss function, which also needs simulation during training. Recently proposed, the Iterated Denoising Energy Matching (iDEM) [34] uses a stochastic score matching objective leveraging solely the energy function and its gradient—requiring no data samples—to train a diffusion-based sampler, representing the current state-of-the-art and showing promise for potential applications. Further exploration of iDEM represents a promising direction for future research.

Applications and side interests Flow matching has been effectively applied in image generation applications such as Stable Diffusion 3 [35]. Other applications focusing on latent spaces, such as Scalable Interpolant Transformers [36] and Neural Network Diffusion [37], exhibit promising potential for further study.

Many studies have used flow [38] or diffusion [39] models for molecular applications. It is natural to extend and explore potential applications using recent stochastic interpolant frameworks. Recent related works, such as equivariant flow matching [40], FoldFlow-based approaches [41], and FrameFlow [42], serve as successful examples. During my master's thesis, we successfully extended flow matching to the Implicit Transfer Operator (ITO) [39] framework to approximate $\rho_\tau(\mathbf{x}_{t+\Delta t}|\mathbf{x}_t)$, focusing on molecular dynamics rather than molecular design. Simultaneously, we attempt to generalize ITO across thermodynamic variables. Many potential research topics can be explored by combining these with enhanced sampling methods [43].

Regarding side interests, the development of structured, uncertainty-aware generative models as inference machines is particularly appealing. One related work is GFlowNets [44], which also shares connections with iDEM [34]. I would be excited to witness new generative models surpass autoregressive in LLMs and Multi-modal models, potentially paving the way towards AGI.

4 Research Plans

In contrast to traditional engineering fields, the AI research community operates within a rapidly evolving, challenging, and competitive environment. Often, significant shifts can occur within a one-year timeframe. Therefore, it's crucial to stay open to new ideas and strategically choose research directions. Unlike purely application-driven research, which typically follows a clear path of trial and optimization goals utilizing existing frameworks, my proposed area will inevitably involve theoretical exploration and iterative experimentation. Hence, it is wise to establish a general goal and a one-year plan, recognizing that questions arise organically during daily research and ideas emerge through accumulation and discussions.

The *overall goal* of this PhD is to improve and apply continuous-time generative models to address the challenge of sampling unnormalized densities. Although this topic has gained increased attention,

particularly over the past two years, it originally has a rich research history within the biophysics community. Consequently, domain-specific knowledge should be addressed during the research process.

Over the upcoming year, spanning from the *pre-admission* phase to the first few months of my PhD journey, I have outlined several key objectives in Table 1. Initially, my focus will be on concluding my thesis work, with a specific emphasis on applying kinetic states transformation method to molecular systems. With optimism for successful outcomes, our intention is to formalize and release our code as open-source, with the goal of publishing a paper in either JCTC or JCP. Furthermore, I aspire to deepen my understanding of biophysics, motivated by the domain knowledge requirement for this program. To this end, I will learn more about SDE, statistical physics, and the JAX tool. Moreover, I will make experiments with iDEM and FAB (codes are available). While iDEM exhibits promise, its original success might be due to the smooth nature of the Lennard-Jones potential. I would be curious to see its applicability to small molecular systems. I aim to conduct a thorough investigation into its limitations to gain a comprehensive understanding. Concurrently, I will actively seek out collaboration opportunities with prospective supervisors to ensure alignment with our research objectives.

Table 1: Key Objectives

Topics	Description	Period
Master Thesis	Generalize ITO to kinetic states, Publish	$-$Aug., 2024
Domain Knowledge	SDE, Statistical Physics, JAX, Biophysics	$\sim$
iDEM and FAB	Applicability for small molecular systems	$-$Nov., 2024
Collaboration	Align research objectives with Supervisors	$\sim$ Admission

5 Discussions

Research on both continuous-time generative models and sampling unnormalized densities is crucial. Integrating structural and causal information into generative models is important for AI safety, ensuring the development of an explainable and trustworthy framework toward AGI. Furthermore, generative models are powerful tools for scientific discovery [45] and productivity enhancement.

Sampling unnormalized densities has been a longstanding challenge [46]. Utilizing AI methods holds promise to either replace or enhance current traditional frameworks. One direct outcome is the potential reduction in computational power required for molecular simulations, aligning with sustainability goals. Additionally, this advancement could accelerate the process of designing new drugs and materials, ultimately benefiting humanity.

Acknowledgments

The author wants to thank my supervisor Simon Olsson and teammate Selma Moqvist for their helpful discussions during my master's thesis. Special thanks to Jakub M. Tomczak for his informative blog posts on generative models, which have been invaluable to the community. I'm also grateful to Nikolay Malkin for his support and insightful discussions.

References

[1] Mark E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simultion*. Oxford University Press, Oxford, 2 edition, 2023.

[2] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Elsevier, 2023.

[3] Shams Mehdi, Zachary Smith, Lukas Herron, Ziyue Zou, and Pratyush Tiwary. Enhanced sampling with machine learning: A review. *arXiv preprint arXiv:2306.09111*, 2023.

[4] Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

[5] Grant M Rotskoff. Sampling thermodynamic ensembles of molecular systems with generative neural networks: Will integrating physics-based models close the generalization gap? *Current Opinion in Solid State and Materials Science*, 30:101158, 2024.

[6] Bernt Øksendal and Bernt Øksendal. *Stochastic differential equations*. Springer, 2003.

[7] Mathias Schreiner, Ole Winther, and Simon Olsson. Implicit transfer operator learning: Multiple time-resolution models for molecular dynamics. In A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 36449–36462. Curran Associates, Inc., 2023.

[8] E Weinan, Tiejun Li, and Eric Vanden-Eijnden. *Applied stochastic analysis*, volume 199. American Mathematical Soc., 2021.

[9] Afonso S Bandeira, Antoine Maillard, Richard Nickl, and Sven Wang. On free energy barriers in gaussian priors and failure of cold start mcmc for high-dimensional unimodal distributions. *Philosophical Transactions of the Royal Society A*, 381(2247):20220150, 2023.

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[12] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[13] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[14] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[15] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

[16] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

[17] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

[18] Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*, 2023.

[19] Lixin Sun, Jonathan Vandermause, Simon Batzner, Yu Xie, David Clark, Wei Chen, and Boris Kozinsky. Multitask machine learning of collective variables for enhanced sampling of rare events. *Journal of Chemical Theory and Computation*, 18(4):2341–2353, 2022.

[20] Luigi Bonati, GiovanniMaria Piccini, and Michele Parrinello. Deep learning the slow modes for rare events sampling. *Proceedings of the National Academy of Sciences*, 118(44):e2113533118, 2021.

[21] Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Senne, Martin Held, John D Chodera, Christof Schütte, and Frank Noé. Markov models of molecular kinetics: Generation and validation. *The Journal of chemical physics*, 134(17), 2011.

[22] Jérôme Hénin, Tony Lelièvre, Michael R Shirts, Omar Valsson, and Lucie Delemotte. Enhanced sampling methods for molecular dynamics simulations. *arXiv preprint arXiv:2202.04164*, 2022.

[23] 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. *arXiv preprint arXiv:2302.00482*, 2023.

[24] Christian Léonard. A survey of the schr\" odinger problem and some of its connections with optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.

[25] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709, 2021.

[26] Alexander Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free schr\" odinger bridges via score and flow matching. *arXiv preprint arXiv:2307.03672*, 2023.

[27] Ricky TQ Chen and Yaron Lipman. Riemannian flow matching on general geometries. *arXiv preprint arXiv:2302.03660*, 2023.

[28] Andrew Campbell, Jason Yim, Regina Barzilay, Tom Rainforth, and Tommi Jaakkola. Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. *arXiv preprint arXiv:2402.04997*, 2024.

[29] Hannes Stark, Bowen Jing, Chenyu Wang, Gabriele Corso, Bonnie Berger, Regina Barzilay, and Tommi Jaakkola. Dirichlet flow matching with applications to dna sequence design. *arXiv preprint arXiv:2402.05841*, 2024.

[30] Marcin Sendera, Minsu Kim, Sarthak Mittal, Pablo Lemos, Luca Scimeca, Jarrid Rector-Brooks, Alexandre Adam, Yoshua Bengio, and Nikolay Malkin. On diffusion models for amortized inference: Benchmarking and improving stochastic control and sampling. *arXiv preprint arXiv:2402.05098*, 2024.

[31] Qinsheng Zhang and Yongxin Chen. Path integral sampler: a stochastic control approach for sampling. *arXiv preprint arXiv:2111.15141*, 2021.

[32] Francisco Vargas, Will Grathwohl, and Arnaud Doucet. Denoising diffusion samplers. *arXiv preprint arXiv:2302.13834*, 2023.

[33] Laurence Illing Midgley, Vincent Stimper, Gregor NC Simm, Bernhard Schölkopf, and José Miguel Hernández-Lobato. Flow annealed importance sampling bootstrap. *arXiv preprint arXiv:2208.01893*, 2022.

[34] Tara Akhound-Sadegh, Jarrid Rector-Brooks, Avishek Joey Bose, Sarthak Mittal, Pablo Lemos, Cheng-Hao Liu, Marcin Sendera, Siamak Ravanbakhsh, Gauthier Gidel, Yoshua Bengio, et al. Iterated denoising energy matching for sampling from boltzmann densities. *arXiv preprint arXiv:2402.06121*, 2024.

[35] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. *arXiv preprint arXiv:2403.03206*, 2024.

[36] Nanye Ma, Mark Goldstein, Michael S Albergo, Nicholas M Boffi, Eric Vanden-Eijnden, and Saining Xie. Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers. *arXiv preprint arXiv:2401.08740*, 2024.

[37] Kai Wang, Zhaopan Xu, Yukun Zhou, Zelin Zang, Trevor Darrell, Zhuang Liu, and Yang You. Neural network diffusion. *arXiv preprint arXiv:2402.13144*, 2024.

[38] Hao Wu, Jonas Köhler, and Frank Noé. Stochastic normalizing flows. *Advances in Neural Information Processing Systems*, 33:5933–5944, 2020.

[39] Mathias Schreiner, Ole Winther, and Simon Olsson. Implicit transfer operator learning: Multiple time-resolution surrogates for molecular dynamics. *arXiv preprint arXiv:2305.18046*, 2023.

[40] Leon Klein, Andreas Krämer, and Frank Noé. Equivariant flow matching. *Advances in Neural Information Processing Systems*, 36, 2024.

[41] Avishek Joey Bose, Tara Akhound-Sadegh, Kilian Fatras, Guillaume Huguet, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael Bronstein, and Alexander Tong. Se (3)-stochastic flow matching for protein backbone generation. *arXiv preprint arXiv:2310.02391*, 2023.

[42] Jason Yim, Andrew Campbell, Andrew YK Foong, Michael Gastegger, José Jiménez-Luna, Sarah Lewis, Victor Garcia Satorras, Bastiaan S Veeling, Regina Barzilay, Tommi Jaakkola, et al. Fast protein backbone generation with se (3) flow matching. *arXiv preprint arXiv:2310.05297*, 2023.

[43] Juan Viguera Diez, Sara Romeo Atance, Ola Engkvist, and Simon Olsson. Generation of conformational ensembles of small molecules via surrogate model-assisted molecular dynamics. 2023.

[44] Nikolay Malkin, Salem Lahlou, Tristan Deleu, Xu Ji, Edward Hu, Katie Everett, Dinghuai Zhang, and Yoshua Bengio. Gflownets and variational inference. *arXiv preprint arXiv:2210.00580*, 2022.

[45] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

[46] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.