Thesis for The Degree of Licentiate of Philosophy

Learning Molecular Dynamics with Generative Models

From Equilibrium to Nonequilibrium Systems

Johann Flemming Gloy

Department of Computer Science and Engineering
Chalmers University of Technology | University of Gothenburg
Gothenburg, Sweden, 2026

Learning Molecular Dynamics with Generative Models
From Equilibrium to Nonequilibrium Systems

Johann Flemming Gloy

ISSN 1652-876X

Department of Computer Science and Engineering
Division of Data Science and Artificial Intelligence
AI and Machine Learning in the Natural Sciences Lab
Chalmers University of Technology | University of Gothenburg
SE-412 96 Göteborg,
Sweden
Phone: +46(0)31 772 1000

Printed by Chalmers Digitaltryck,
Gothenburg, Sweden 2026.

To my family
Für meine Familie

Learning Molecular Dynamics with Generative Models
From Equilibrium to Nonequilibrium Systems

Johann Flemming Gloy

Department of Computer Science and Engineering
Chalmers University of Technology | University of Gothenburg

Abstract

Statistical mechanics provides a broad theoretical framework for modeling molecular and biological systems, but in practice, the underlying dynamical equations are often not analytically tractable. Consequently, molecular simulations have been the workhorse of statistical mechanics for the past seventy years. However, with the advent of machine learning and generative models, these data-driven methods are starting to complement simulations by providing fast surrogates in scenarios where direct simulation is prohibitively expensive.

In this thesis, we discuss how generative models can enhance traditional simulation methods in both equilibrium and nonequilibrium settings.

In equilibrium sampling with continuous normalizing flow-based Boltzmann generators, likelihood evaluations scale unfavorably with system size. We show how this issue can be alleviated, demonstrating speedups of up to 100× on Lennard-Jones systems.

Nonequilibrium settings encompass a wider range of systems. We briefly discuss some of the generative modeling methods appropriate in this setting and present an extension of implicit transfer operator models to nonautonomous domains. By combining flow map matching with a physically grounded short-time inductive bias, we accurately model both long- and short-time behavior of nonautonomous systems.

This work concludes with a discussion of the broader role of generative machine learning methods in computational statistical mechanics, pointing out new applications and possible future research directions.

Keywords

Generative models, Boltzmann sampling, Boltzmann generators, nonequilibrium statistical mechanics

List of Publications

Appended publications

This thesis is based on the following publications:

[**Paper I**] **Johann Flemming Gloy**, Simon Olsson, *HollowFlow: Efficient Sample Likelihood Evaluation using Hollow Message Passing*
Advances in Neural Information Processing Systems 38 (NeurIPS 2025).

[**Paper II**] **Johann Flemming Gloy**, Simon Olsson, *Generative Transition Density Models for Nonautonomous Dynamics*
Preprint, not published elsewhere.

Acknowledgment

First of all, I would like to thank my supervisor Simon Olsson for his great mentorship, supervision and support throughout the first half of my PhD. I also want to thank my co-supervisor Fredrik Johansson and my examiner Ross King for their valuable feedback and continuous support. I am grateful to Morten Hjorth-Jensen for agreeing to serve as the discussion leader during my licentiate defense.

I would especially like to thank Rocío Mercado for her feedback and fruitful discussions. Furthermore, I am very grateful to all my colleagues and friends in the AIMLeNS and AIME research labs and the DSAI division at Chalmers University of Technology. I would like to especially thank Selma, Beatrice, Pablo, Giannis, Ross, Juan, Christopher, Stefano, Télio, Guy, Kelsey, Richard, Farzaneh, Aleksandar, Hang, Mathias, Emma, Flavio, Chenxiao, Nils, Sofia, Alberto, Tanmay and all my many other wonderful colleagues. I am also grateful to the administrative staff, who continuously work in the background. I especially thank Andrea Brodén for her constant availability and help.

Last, but certainly not least, I thank my family and friends, and especially my partner, Hannah.

This work was generously funded by the Wallenberg AI, Autonomous Systems and Software Program (WASP), funded by the Knut and Alice Wallenberg Foundation, for which I am very grateful.

Contents

Part I

Summary

Chapter 1

Introduction

For centuries, science was primarily driven by experimental observations in combination with the theoretical discovery of fundamental laws via inductive reasoning. Even if all fundamental laws were well understood, they would not directly suffice to explain all phenomena in nature. The interplay of seemingly simple principles can lead to unexpected and complex phenomena.

There is a sentiment, famously attributed to Stephen Hawking, that, while the 20th century was the century of physics, the 21st will be the century of complexity.

One example of such complexity is the dynamics of biomolecules and their interactions. Molecular dynamics (MD) is one of the principal computational tools for studying such complex phenomena by means of numerical simulations. MD emerged in the late 1950s [1], following Markov Chain Monte Carlo simulations [2]. The field developed and improved dramatically over the decades with the development of more capable computers, more precise force fields describing the interactions between atoms and more advanced algorithms [3], [4], [5], [6], [7]. In the 1960s, the first realistic simulations of matter emerged [8], [9] while it took until 1977 to simulate the first protein [10], [11]. Even today, however, large-scale MD simulations are limited to microseconds at best while longer timescales often are desirable for biological applications [12]. Over the past 20 years, machine learning models, while being an example of emergent complexity themselves, have become a powerful surrogate modeling technique. This includes in particular machine-learned force fields that can deliver the precision of ab initio quantum calculations at a fraction of the cost [11], [13], [14], [15], [16]. Machine learning methods reach, however, well beyond force field approximations [17], [18]. In particular, the recent success of generative models [19], [20], [21] launched the new research direction of applying them to predicting equilibrium system configurations [22], [23] or dynamical transitions [24], [25], [26]. These generative modeling techniques are the subject of this thesis.

We first give some background on equilibrium and nonequilibrium statistical mechanics – the theoretical ground on which molecular dynamics rests. Then we briefly summarize machine-learning-based generative modeling techniques

together with a very brief introduction to groups and symmetries that these models can take into account. Bringing these topics together, we discuss how generative modeling can be used to enhance traditional MD techniques. Starting from Boltzmann generators, a technique to learn the distribution of equilibrium configurations, we move on to techniques that aim to approximate the transition density.

We proceed with a summary of the two papers that this thesis is based on. The first paper introduces a method that makes likelihood evaluation in continuous normalizing flows scalable, enabling their use in high-dimensional Boltzmann generators. In the second paper we move beyond equilibrium and generalize an existing framework for generative transition density modeling to nonequilibrium settings with nonautonomous dynamics.

We conclude with a brief summary and a discussion of possible future research directions.

Chapter 2

Background

2.1 Stochastic Dynamics of Molecular Systems

The dynamics of systems with many degrees of freedom, like molecules or materials, are inherently difficult to study. Even though the underlying dynamical equations are often known, or can at least be modeled precisely, solving them exactly is often impossible and we have to rely on numerical methods. But numerical methods quickly hit their limits [27]. This applies especially to quantum mechanics where the size of the state space grows exponentially with the size of the system [28]. Even if we rely on classical approximations, resolving the dynamics of all degrees of freedom of a system is often not feasible. Thus, one only models the dynamics of an interesting subsystem, while treating the interaction with the environment, i.e., the rest of the system, as thermal noise. In this situation, the environment is often referred to as the heat bath. The temperature of the heat bath controls the strength of the noise. This results in a type of equations called Langevin dynamics which we detail below.

2.1.1 Langevin Dynamics

Consider a system with a configurational space Ω, typically $\Omega \subseteq \mathbb{R}^{3N}$, where N is the number of particles of the system we aim to describe. The time evolution of a configuration $\mathbf{x} \in \Omega$ of such a system can then be modeled by the following equation [29], [30]:

$$m_i \ddot{x}_{i\alpha} = -\gamma \dot{x}_{i\alpha} - \frac{\partial V(\mathbf{x})}{\partial x_{i\alpha}} + \xi_{i\alpha}(t), \tag{2.1}$$

where the index $i \in \{1, ..., N\}$ runs over all particles and $\alpha \in \{1, 2, 3\}$ over the physical space dimensions. V is the potential energy of the system and ξ is Gaussian white noise defined by

$$\langle \xi_{i\alpha}(t) \xi_{j\beta}(t') \rangle = 2\gamma T \delta_{ij} \delta_{\alpha\beta} \delta(t - t'). \tag{2.2}$$

T is the temperature. γ and T are related to the diffusion coefficient – or more generally diffusion tensor – by the Einstein relation $D = T/\gamma$[1]. $\delta(t)$ is the Dirac delta distribution. More formally, the Langevin equation can be understood as a stochastic differential equation and the noise term can mathematically be defined as the formal derivative of a Wiener process [31].

In a microscopic setting, the inertia term $m_i\ddot{x}$ can often be neglected: Let R_i denote the radius of particle i. As $m_i\ddot{x} \sim R^3$ and $\gamma\dot{x}_i \sim R$ (Stokes' law), the inertia term becomes negligible for $R \to 0$. The resulting first-order equation is called overdamped Langevin equation:

$$\gamma\dot{x}_{i\alpha} = -\frac{\partial V(\mathbf{x})}{\partial x_{i\alpha}} + \xi_{i\alpha}(t). \tag{2.3}$$

More generally, the drift force does not have to be the negative gradient of a potential, it can also be a more general force $\mathbf{f}$:

$$\gamma\dot{x}_{i\alpha} = f_{i\alpha}(\mathbf{x}) + \xi_{i\alpha}(t). \tag{2.4}$$

To simplify notation, the following discussion will focus on the overdamped Langevin equation.

2.1.2 From Langevin Dynamics to the Fokker-Planck Equation

The Langevin equation describes the stochastic dynamics of a given system as a single realization of a stochastic process. However, it is often of interest to describe the same dynamics as the time evolution of a probability distribution $P(\mathbf{x}, t)$. We will give a brief overview of how the two descriptions are linked [30]. We start from the Chapman-Kolmogorov equation, describing how the density $P(\mathbf{x}, t)$ changes in a time step τ:

$$P(\mathbf{x}, t+\tau) = \int P(\mathbf{x}, t+\tau|\mathbf{x}', t)P(\mathbf{x}', t)d\mathbf{x}'. \tag{2.5}$$

The goal is to derive an expression for the differential $\partial P(\mathbf{x}, t)/\partial t$. For that, one can expand the integrand into a Taylor series in $\Delta = \mathbf{x} - \mathbf{x}'$, called the Kramers-Moyal expansion. By Pawula's theorem one can show that for a Langevin equation with white Gaussian noise only the first two terms contribute [30]. This results in the following partial differential equation describing the time evolution of the probability density associated with the Langevin equation above, where we summarize both indices from above into a single index running from 1 to $3N$ to ease notation:

$$\frac{\partial}{\partial t}P(\mathbf{x}, t) = -\frac{1}{\gamma}\sum_i \frac{\partial}{\partial x_i} f_i(\mathbf{x})P(\mathbf{x}, t) + \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j}(D_{ij}P(\mathbf{x}, t)). \tag{2.6}$$

Its counterpart for underdamped dynamics is known as the Klein-Kramers equation.

[1] We work in units where the Boltzmann constant k_B is equal to one throughout this thesis.

An important special case, known as the Smoluchowski equation, is when the force $\mathbf{f}$ can be written as the negative gradient of a potential energy function $V(\mathbf{x})$:

$$\frac{\partial}{\partial t}P(\mathbf{x},t) = \frac{1}{\gamma}\sum_i \frac{\partial}{\partial x_i}\left(\left(\frac{\partial}{\partial x_i}V(\mathbf{x})\right)P(\mathbf{x},t)\right) + \sum_{ij}\frac{\partial^2}{\partial x_i \partial x_j}(D_{ij}P(\mathbf{x},t)). \tag{2.7}$$

2.1.3 The Boltzmann Equilibrium

When the potential V is independent of time and the diffusion coefficient is constant, the Fokker-Planck equation admits a stationary solution with zero probability current known as the Boltzmann distribution [30], [32]

$$P_{\mathrm{eq}}(\mathbf{x}) = \frac{1}{Z}\exp(-\beta V(\mathbf{x})), \tag{2.8}$$

where $Z = \int \exp(-\beta V(\mathbf{x}))d\mathbf{x}$ is the partition function and $\beta = 1/T$ is the inverse temperature. It is often of practical interest to compute averages of observables $O : \Omega \to \mathbb{R}$ over the Boltzmann distribution, i.e.,

$$\langle O\rangle = \int_\Omega O(\mathbf{x})P_{\mathrm{eq}}(\mathbf{x})d\mathbf{x}. \tag{2.9}$$

O can have a variety of shapes and interpretations, for example, in polymer physics one might be interested in the radius of gyration [33] describing the overall size of the polymer. In the case of protein folding, one is often interested in calculating the free energy profile [32] $F(q) = -T\log[\langle\delta(S(\mathbf{x}) - q)\rangle]$ with respect to a collective variable $S(\mathbf{x})$.

Computing these averages, i.e., integrals, in practice is challenging. Analytically, they are only tractable in special cases and one usually has to rely on numerical methods, such as Monte Carlo integration, to compute approximate solutions. However, even that is usually extremely difficult. Monte Carlo methods approximate these integrals as [32]

$$\langle O\rangle = \int_\Omega O(\mathbf{x})P_{\mathrm{eq}}(\mathbf{x})d\mathbf{x} = \lim_{N\to\infty}\frac{1}{N}\sum_i O(\mathbf{x}_i), \quad \mathbf{x}_i \sim P_{\mathrm{eq}}. \tag{2.10}$$

This requires access to a large number of uncorrelated samples $\mathbf{x}_i \sim P_{\mathrm{eq}}$ that faithfully cover all relevant regions of the potentially high-dimensional space Ω. Generating these samples remains expensive despite decades of methodological development [2], [32]. Traditional methods [6] generate samples by directly integrating eq. (2.3) or by Markov chain Monte Carlo (MCMC) methods [2], where a move is proposed and then accepted according to a criterion that ensures convergence to the Boltzmann distribution [32]. However, in general, these techniques have to rely on local moves, while relevant regions of the configuration space are often separated by large energy barriers resulting in correlated samples and very slow convergence rates. Recently, machine learning

techniques have emerged [22] that can alleviate some of these problems. We will give a brief summary of these in section 2.3.

However, even if the computation of Boltzmann averages were solved, we would still be unable to compute important properties of many systems. Most importantly, one is often interested in dynamical observables, i.e., observables that depend on time like correlation functions. Even more challenging are systems that are not in or do not even admit a Boltzmann equilibrium of the form eq. (2.8).

2.1.4 Nonequilibrium Statistical Mechanics

Nonequilibrium systems that do not admit a Boltzmann distribution mainly fall into one of two largely distinct categories. Firstly, a system may be driven. This could occur if the force is explicitly time-dependent or if it is non-conservative, i.e., not the gradient of a potential energy function. These systems either reach a nonequilibrium steady state or never approach a stationary distribution at all [30], [32].

Secondly, there are systems that relax to equilibrium extremely slowly so that the equilibrium distribution cannot be reached on experimentally and computationally feasible timescales. In particular, glassy dynamics [34], [35], [36] falls into this category.

Both types of nonequilibrium systems require a different theoretical and computational toolbox to study them compared to equilibrium systems, such as stochastic thermodynamics [37].

In the computational and modeling context, it is often of interest to look at possibly time-dependent transition probabilities

$$P(\mathbf{x}, t \mid \mathbf{y}, s), \tag{2.11}$$

and entire dynamical trajectories. These transition densities can then be used to compute dynamic two-time observables such as correlation functions [29]:

$$C(t, s) = \int\int O_1(\mathbf{x}) O_2(\mathbf{y}) P(\mathbf{x}, t \mid \mathbf{y}, s) P(\mathbf{y}, s) d\mathbf{y} d\mathbf{x} \tag{2.12}$$

In practice, especially in high-dimensional systems, these computations are in general analytically intractable and computationally expensive.

However, some analytical relations known as fluctuation theorems can be derived exactly and even hold far from equilibrium. Detailed fluctuation theorems [38], [39] relate the probability of the forward trajectory to that of the same trajectory traced backwards in time, showing that the forward trajectory is exponentially more likely than its time-reversed counterpart. Starting from detailed fluctuation theorems, one can derive integral fluctuation theorems, such as the Jarzynski equality [40] in terms of work and free energy, or a related equation by Seifert [39] in terms of the total entropy production ΔS_{tot} along a trajectory:

$$\langle e^{-\Delta S_{\mathrm{tot}}} \rangle = 1, \tag{2.13}$$

which implies the second law of thermodynamics $\langle \Delta S_{\text{tot}} \rangle \geq 0$ via Jensen's inequality. As entropy production is a path-dependent quantity, this provides further motivation for modeling full transition densities for arbitrary times s and t.

Just as in the equilibrium case, generative models can be used to approximate these transition densities from data [24], [41], which we discuss next after a brief overview of generative modeling techniques.

2.2 Learning Probability Distributions with Generative Models

Generative models aim to approximate an unknown or intractable distribution $\tilde{\rho}_1$ from samples by parameterizing a model of the distribution ρ_1 that is easy to sample from.

Generative modeling is a broad field with a vast number of recent developments. This includes variational autoencoders, energy-based models, autoregressive models, generative adversarial networks, flow and diffusion based models and more. See, e.g., [20], [42], [43] for a review. We will focus on flow- and diffusion-based models here.

Flow- and diffusion-based models Flow- and diffusion based models use different techniques to learn a transformation ϕ such that the pushforward of an easy-to-sample prior distribution ρ_0 (typically a Gaussian) under ϕ is equal to ρ_1, closely matching the target distribution $\tilde{\rho}_1$. They can mainly be classified into the following three categories

- **Normalizing Flows** The transformation ϕ is parametrized directly by an invertible neural network [44].
- **Continuous normalizing flows** The transformation ϕ is parametrized as the solution to an initial value problem $\dot{x}_s = b_s(x_s)$, such that $x_0 \sim \rho_0,\ x_1 \sim \rho_1$. The vector field b is learned [45], [46].
- **Diffusion models** The transformation ϕ is parametrized as the solution to a stochastic differential equation $\mathrm{d}x_s = b_s(x_s)\,\mathrm{d}s + \sigma_s dW_s$, such that $x_0 \sim \rho_0,\ x_1 \sim \rho_1$[2] Here, dW_s is a Wiener process. The vector field b is learned [19], [21].

The last two categories have been unified by Albergo et al. [21] in the stochastic interpolant framework.

We will give more details on continuous normalizing flows below, as this is the main modeling paradigm applied in this thesis. The exposition closely follows parts from the background chapter of the second appended paper.

[2]The time interval and the direction of time depend on the precise formulation. Here we follow [21].

2.2.1 Continuous Normalizing Flows

As mentioned, a continuous normalizing flow (CNF) parametrizes the diffeomorphism[3] ϕ that pushes an easy-to-sample prior density ρ_0 forward to a density ρ_1 as the solution of an initial value problem, specified by a learnable vector field $b : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$:

$$\dot{x}_s = b_s(x_s),\ x_0 \sim \rho_0,\ x_1 \sim \rho_1. \tag{2.14}$$

To efficiently learn the vector field from data, several related approaches have recently been proposed [45], [48]. Albergo et al. [21] gave a general framework that unifies these approaches by defining a stochastic interpolant as a stochastic process $I : [0,1] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$

$$I_s(x_0, x_1) = \alpha_s x_0 + \beta_s x_1 \tag{2.15}$$

where $\alpha, \beta : [0,1] \to [0,1]$ are continuously differentiable functions satisfying the boundary conditions $\alpha_0 = \beta_1 = 1, \alpha_1 = \beta_0 = 0$.

The pair (x_0, x_1) is drawn from a coupling $\rho(x_0, x_1)$ whose marginals are ρ_0 and ρ_1: $\int_{\mathbb{R}^d} \rho(x_0, x_1) dx_0 = \rho_1(x_1)$ and $\int_{\mathbb{R}^d} \rho(x_0, x_1) dx_1 = \rho_0(x_0)$. The coupling $\rho(x_0, x_1)$ can be chosen in different ways. It is often of practical and theoretical interest to reduce the transport cost between the two distributions ρ_0 and ρ_1, a field known as optimal transport (OT). In practice, this is often done with a technique known as mini-batch OT, in which the samples within a batch are permuted to minimize the mean square distance between base and target samples [49]. Symmetries of ρ_0 and ρ_1 can also be considered [50], [51] which we will discuss below in more detail.

The vector field b can then be written as a conditional expectation

$$b_s(x) = \mathbb{E}[\dot{I}_s | I_s = x]. \tag{2.16}$$

This expectation is usually intractable. Instead, b_s can be learned by minimizing the following loss or closely related objectives [21], [45], [48], [49], [52]:

$$\mathcal{L}_b(\hat{b}) = \int_0^1 \mathbb{E}_{x_0, x_1}[\|\hat{b}_s(I_s) - \dot{I}_s\|^2] ds. \tag{2.17}$$

An important special case is flow matching [45], [48] where $\alpha_s = 1 - s$ and $\beta_s = s$. As sampling from the model requires numerical integration of eq. (2.14), several approaches have been proposed to learn its solution instead [52], [53], [54]. Boffi et al. recently proposed a generalized framework called flow map matching that encapsulates these approaches which we describe in more detail in the second paper appended to this thesis.

While flow-map-based models lower the computational cost of sampling, calculating likelihoods under the model remains a bottleneck. Mathematically, computing likelihoods is straightforward. The ODE eq. (2.14) together with the prior ρ_0 gives rise to a time-dependent distribution ρ_s given by the continuity equation [21]

$$\partial_s \rho_s = -\nabla \cdot (\rho_s b_s). \tag{2.18}$$

[3] under standard regularity conditions [47]

Solving eq. (2.18), we get the change in log-probability under the CNF,

$$\Delta \log \rho^{\mathrm{CNF}} := \log \rho_1(x_1) - \log \rho_0(x_0) = -\int_0^1 \nabla \cdot b_s(x_s)\, ds. \tag{2.19}$$

In practice, the integral in eq. (2.19) has to be discretized in time for numerical evaluation. At every time step, the divergence of the vector field needs to be calculated via d backwards passes through the model making it $O(d)$ more expensive than a normal sampling step [55]. For high-dimensional applications where likelihood evaluation is required and approximate evaluations via, e.g., the Hutchinson estimator [56] are not sufficient, this is one of the main bottlenecks of continuous normalizing flows. Boltzmann generators [22] (see section 2.3), which sample from physical equilibrium distributions, are a prominent example of such an application. Precisely the issue of slow exact likelihoods in Boltzmann generators is addressed in the first paper appended to this thesis.

2.2.2 Symmetries in Machine Learning Models

Many systems in physics are subject to some symmetry constraints. For example, translating or rotating a molecule in space does not change its properties. Incorporating these symmetries can yield significant advantages in scientific modeling applications such as Boltzmann generators [50], [51]. In machine learning, the study of how to incorporate these symmetries is the field of geometric deep learning. Below, we give a brief introduction to some of the core concepts. A more extensive review can be found in [57].

Groups and Equivariance Symmetries are typically described by groups and their actions. With these concepts, a function can then be defined to be invariant (not change) or equivariant (change in some predictable way) under some group transformation. We give a few core mathematical concepts below to make this notion precise.

A *group* is a set G together with an associative binary operation, an identity element and inverses. Important examples are the permutation group and groups of transformations in three dimensions, in particular the rotation group $SO(3)$, the orthogonal group $O(3)$, which additionally includes reflections, and the Euclidean group $E(3)$, which includes rotations, reflections and translations.

The symmetries of a system are described by actions of a group on the configuration space of the system. A *group action* is a map $G \times X \to X$ compatible with the group operation.

Functions often change under a group action in a controlled way. This is called equivariance. More precisely, let $f : X \to Y$ and let $\triangleright_X$ and $\triangleright_Y$ be two actions of the group G on the sets X and Y, respectively. The function f is called *equivariant* under the group actions $\triangleright_X$ and $\triangleright_Y$ if $f(g \triangleright_X x) = g \triangleright_Y f(x)$ for all $x \in X$ and $g \in G$. If $\triangleright_Y$ is trivial, we have $f(g \triangleright_X x) = f(x)$ and the function is called *invariant*.

Equivariant graph neural networks Equivariant graph neural networks are graph neural networks [16] that are equivariant under some group action.

There is a vast amount of literature on that subject; see e.g., [58] for a review. Graph neural networks provide a natural framework for modeling point clouds in three-dimensional Euclidean space.

Briefly, a graph neural network operates on a graph $\mathcal{G} = (V, E)$. The vertices V exchange information about their features along the edges E via message functions. The messages are then pooled and subsequently used to update the vertex features via update functions. This process is iterated and finally, a readout function is applied to all features to obtain the final output of the network.

The properties of these point clouds are often equivariant or invariant under translations and rotations and invariant under permutations of identical particles. In this case, the group action consists of translating or rotating the point cloud in three-dimensional Euclidean space or of permuting the individual particles.

Translation invariance is often handled by letting the network operate on the pairwise distances, which naturally are translation invariant. Equivariance under rotation is usually more complicated and generally obtained by ensuring that the message and update functions are equivariant. Many specialized architectures have been developed to handle these symmetries [57], [58]. In particular, we are using the PaiNN architecture [59] in the first appended paper. Very briefly, PaiNN uses two types of node features, scalar (invariant) and vector (equivariant), with message and update functions that respect their respective transformation properties under $O(3)$.

Equivariant continuous normalizing flows When modeling a probability density ρ_1 that is invariant under some group action $\triangleright$ of some group G, i.e., $\rho_1(g \triangleright x) = \rho_1(x)$, it can be beneficial to constrain the model to obey this symmetry. Köhler et al. [51] show, under some additional technical conditions, that an equivariant vector field as in eq. (2.14), together with an invariant prior ρ_0 leads to an invariant target ρ_1. Additionally, Klein et al. [50] introduce equivariant OT. In this extension of mini-batch OT, the samples of the prior and target distributions are not only aligned by permutations in the mini-batch but also by finding the optimal element $g \in G$ to further reduce the transport cost, where G is the symmetry group under which the two distributions are invariant. Empirically, they demonstrate a significant increase in model performance compared to a model that does not take symmetries into account. We apply these techniques in the first appended paper.

2.3 Learning Equilibrium Distributions – Boltzmann Generators

Boltzmann generators [22] tackle the computational problem of equilibrium sampling by leveraging generative models to directly draw uncorrelated samples from the Boltzmann distribution eq. (2.8). The main idea is to train a (continuous) normalizing flow or a diffusion model such that the target distribution ρ_1 is as close as possible to the Boltzmann distribution P_{eq}. An important

requirement is the tractability of the model likelihood, since observables are typically computed as Monte Carlo averages:

$$\langle O \rangle = \int P_{\text{eq}}(\mathbf{x}) O(\mathbf{x}) d\mathbf{x} \approx \frac{1}{N} \sum_{i=1}^{N} O(\mathbf{x}_i), \quad \mathbf{x}_i \sim P_{\text{eq}}. \tag{2.20}$$

However, the samples from the model distribution ρ_1 are only approximately samples from the Boltzmann distribution. If these samples were used directly in eq. (2.20) to approximate the integral, this would lead to biased expectation values. Thus, one typically employs reweighting

$$\langle O \rangle = \int P_{\text{eq}}(\mathbf{x}) O(\mathbf{x}) d\mathbf{x} = \int \rho_1(\mathbf{x}) \frac{P_{\text{eq}}(\mathbf{x})}{\rho_1(\mathbf{x})} O(\mathbf{x}) d\mathbf{x} \tag{2.21}$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} \frac{P_{\text{eq}}(\mathbf{x}_i)}{\rho_1(\mathbf{x}_i)} O(\mathbf{x}_i), \quad \mathbf{x}_i \sim \rho_1. \tag{2.22}$$

If we define weights $w_i = \exp(-\beta V(\mathbf{x}_i))/\rho_1(\mathbf{x}_i)$, we can write $\langle O \rangle$ as

$$\langle O \rangle \approx \frac{\sum_{i=1}^{N} w_i O(\mathbf{x}_i)}{\sum_{i=1}^{N} w_i}, \tag{2.23}$$

making it a tractable computation not relying on the partition function. This still requires access to the energy function, which is typically available, and the likelihood under the model, the evaluation of which is one of the practical challenges discussed below.

Training Data Training the Boltzmann generator requires samples from the distribution that we try to model. This can, to some extent, be overcome by energy-based training [22], [60]. However, energy-based training often suffers from mode collapse, even though there has been some recent progress in mitigating this [61].

Small effective sample size The reweighting procedure described above often leads to a small effective sample size (ESS), i.e., a proxy for the number of statistically independent samples [62], [63],

$$\text{ESS} = \frac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} w_i^2}, \tag{2.24}$$

even if the learned distribution ρ_1 is only slightly off from the ground truth P_{eq}. This problem gets worse the higher the dimension of the system [50], [63], [64].

Slow likelihood evaluation Evaluating the likelihood under the model is only fast for a subclass of normalizing flows, which, however, have limited expressiveness. For the more expressive continuous normalizing flows, likelihood evaluation is generally slow and scales as $\mathcal{O}(d)$ per integration step, where d is

the system dimension. Calculating the likelihood becomes impractical in high dimensions [55], which are, however, often the cases of practical interest. This problem has been addressed in several recent works [55], [65], including the first paper appended to this thesis, which is related in spirit to [55].

2.4 Learning Transition Densities – Implicit Transfer Operator Models

As detailed above, Boltzmann generators can only approximate equilibrium distributions and are thus fundamentally limited to equilibrium systems. However, many real-world systems are inherently out of equilibrium or, even for equilibrium systems, one might be interested in dynamical observables.

One possibility to extend Boltzmann generators to be able to model dynamics instead of only the Boltzmann equilibrium is to model the transition density $P_\tau(y \mid x)$. There are various modeling approaches in that spirit, including, but not limited to, Markov state models [66], [67], [68], VAMPnets [69] and Koopman operator estimation techniques [70].

These techniques are built on writing the probabilistic evolution in terms of the linear operator $\mathcal{P}_\tau$, defined as

$$\mathcal{P}_\tau P(x,t) = \int P_\tau(x \mid y) P(y,t) dy. \tag{2.25}$$

This operator, called the Perron-Frobenius operator [71], is the solution operator of the Fokker-Planck equation (2.6). The Koopman operator $\mathcal{K}_\tau$, which propagates an observable f instead of the density P, is the L^2 adjoint of $\mathcal{P}_\tau$:

$$\mathcal{K}_\tau f(x,t) = \int P_\tau(y \mid x) f(y,t) dy. \tag{2.26}$$

The eigenvalues and eigenfunctions of these transfer operators encode important dynamical information, such as relaxation times and metastable sets [71]. The data-driven methods described above directly approximate a transfer operator by discretizing the state space or by projecting onto a finite set of basis functions.

A different and particularly promising method employs conditional generative models to directly learn the transition density [24] without explicitly modeling the underlying transfer operator. Under the name implicit transfer operator (ITO) models, this approach has shown great success in modeling molecular dynamics at time steps orders of magnitude larger than what is stable in numerical integration [24], [72]. Crucially, these models have also demonstrated the ability to generalize across chemical space to unseen molecules [41] and the ability to incorporate new types of data into the training procedure [73]. However, all these models have so far been limited to autonomous systems [24], [41], [72], [73] or to fixed time step size τ[74]. The second paper appended to this thesis addresses these limitations by providing an extension of the ITO framework to nonautonomous systems.

Chapter 3

Summary of Included Papers

3.1 HollowFlow: Efficient Sample Likelihood Evaluation using Hollow Message Passing

The first paper addresses the problem of slow likelihood calculation explained in more depth in section 2.3.

Problem Drawing uncorrelated samples from the Boltzmann distribution is a long-standing challenge in computational statistical mechanics. Traditional approaches rely on iterative methods such as Markov chain Monte Carlo or Langevin dynamics. While these approaches provide asymptotically unbiased samples, they exhibit slow convergence because transitions between metastable states are rare, leaving adjacent samples highly correlated. Boltzmann generators provide an alternative method by using generative modeling to produce uncorrelated samples directly. However, the samples from the model need to be reweighted into the true equilibrium distribution to provide accurate Monte Carlo averages. This reweighting requires evaluation of the likelihood of the model samples. For continuous normalizing flows, this is often impractical in high-dimensional settings as it requires $\mathcal{O}(d)$ backward passes per integration step through the model, rendering it d times slower than the sample generation itself, where d is the system size.

Contribution We propose HollowFlow, a continuous normalizing flow architecture whose Jacobian structure allows likelihoods to be computed in $\mathcal{O}(1)$ backward passes rather than $\mathcal{O}(d)$, enabling, in principle, Boltzmann-generator likelihoods at previously infeasible scales. To achieve this, we propose a new type of graph neural network based on the line graph. This architecture is non-backtracking for an arbitrary number of message passing steps, i.e., information from any node of the graph can never return to itself. As a consequence, the Jacobian of the vector field that the graph neural network parametrizes

can be decomposed into a block-diagonal and a block-hollow part. We then exploit this structure to compute the likelihood using only $\mathcal{O}(1)$ instead of the usually required $\mathcal{O}(d)$ backward passes per step.

We further demonstrate that our framework generalizes and can be applied to any equivariant and nonequivariant graph neural network architecture.

We theoretically prove that our model computes correct likelihoods and is up to $\mathcal{O}(d^2)$ faster than previous methods. We finally validate this model by training a Boltzmann generator on two Lennard-Jones systems of different size, demonstrating that our method is up to $100\times$ faster per drawn sample on the larger system. This provides a significant step towards scaling Boltzmann generators to larger molecular systems.

3.2 Generative Transition Density Models for Nonautonomous Dynamics

This paper generalizes implicit transfer operator models, i.e., generative models of the transition density, to nonautonomous settings. A broader background explanation is provided in section 2.4.

Problem While Boltzmann generators provide a powerful, machine-learning-based method to sample equilibrium configurations, kinetic properties cannot be captured at all. Modeling the transition density instead of the equilibrium distribution closes this gap. While there is a rich line of work on modeling the transition density by explicitly modeling the underlying transfer operator, direct generative modeling of the transition density has shown particular promise in the form of implicit transfer operator (ITO) models [24], [41], [72], [73]. However, these models have been limited to autonomous settings, rendering them unsuitable for a number of modeling tasks such as molecular motors [75].

Contribution To close this gap, we develop a framework based on flow map matching [52] and stochastic interpolants [21] that allows for modeling nonautonomous dynamics. We find that accurately modeling the dynamics at short timescales is particularly difficult in some settings, even though accuracy in this regime is crucial when computing path-dependent quantities such as entropy production. To address this, we introduce a short-time inductive bias based on the Euler-Maruyama transition density. Our framework allows for smooth interpolation between the physics-based short-time dynamics and the fully data-driven long-time regime. The resulting model is both physically grounded and data-driven, applicable in nonautonomous nonequilibrium settings. We demonstrate the capability of our framework on one equilibrium (Müller-Brown potential) and two nonequilibrium (moving harmonic trap and stochastic resonance) systems, thereby showing its ability to estimate path-dependent quantities such as entropy from data while still being accurate at single-step sampling over long time lags. Using stochastic resonance, a paradigmatic example of a periodically driven nonequilibrium system, as a test case, we further demonstrate that our framework captures nontrivial nonequilibrium effects and generalizes across parameter space. This extends generative transition density models to nonequilibrium settings previously inaccessible to ITO-style approaches.

Chapter 4

Discussion and Future Work

This thesis investigates generative machine learning methods and their applications in computational statistical mechanics. In particular, we reviewed how generative models can help to solve long-standing computational challenges in statistical physics.

It is important to mention that computational statistical mechanics and the applications of machine learning therein in general and of generative modeling in particular are large and rapidly developing fields. Its methodological toolbox grows rapidly and a complete discussion of the potentials and limitations of these methods is impractical and beyond the scope of this thesis. Instead, we will focus on the two modeling paradigms that directly relate to the two works appended to this thesis: Boltzmann generators and generative models of transition densities.

The idea of applying generative models to Boltzmann sampling has had and continues to have transformative impact. As one of the first methods to allow one-shot generation of uncorrelated samples, it has the potential to outperform traditional approaches such as MCMC. While we addressed one of their limitations, i.e., the slow evaluation of likelihoods, several challenges remain. The reliance on extensive training datasets is one of the more important problems as their generation still requires traditional numerical simulations. While there has been work on purely energy-based training [22], [60], [61], these approaches often have other problems like mode collapse. Other interesting directions include generalizability across chemical space, i.e., being able to sample the configurations of molecules that the model has not seen during training [25]. A valuable end goal would be a foundation model that can accurately sample the equilibrium distribution when conditioned on an arbitrary molecular graph. While there is work in this direction [76], full generalizability has not been achieved yet but does not seem fully out of reach.

Models of transition densities present an even more general approach. While, for equilibrium systems, repeated sampling of transitions will generate samples of the stationary distribution, these models also give access to kinetic properties.

Scaling these models to larger systems and making them fully generalizable across biochemical space is also an important research direction in this setting. Again, there is substantial work in this direction [41], yet broader generalizability remains to be achieved. Instead of focusing on scaling these models to larger and more general settings, we discussed a more fundamental extension in this thesis, focusing on nonautonomous systems. Additionally, extensions and applications of these modeling approaches to other nonequilibrium systems remain to be explored. In one ongoing project we focus on modeling glassy dynamics, i.e., systems that do not reach equilibrium on timescales accessible by established numerical techniques. However, nonequilibrium physics does not end there. It would also be interesting to see if we can model, for example, active matter or other systems where non-Markovian effects come into play – another direction we plan to pursue in the future.

More fundamentally, it would be of great interest to develop a mathematically more rigorous framework for generative models of transition densities. Currently, some ad-hoc modeling and engineering choices need to be made during training and dataset design, including, but not limited to, (possibly time-dependent) prior distributions, long- and short-time inductive biases and the inclusion of further information about the system at hand. While the second paper provides a stepping stone in that direction, it does not present a general mathematical framework. To achieve this, the stochastic interpolant framework [21] might offer a good starting point.

To summarize, machine learning in general and generative modeling in particular have shown great success and even more potential in the field of computational statistical mechanics. Yet, there is room for improvement and further exploration. In particular, broadening the applicability of these models to a larger class of systems is an interesting methodological direction whereas transferability and scalability are two of the main practical challenges.

Bibliography

[1] B. J. Alder and T. E. Wainwright, "Studies in molecular dynamics. i. general method," *The Journal of Chemical Physics*, vol. 31, no. 2, pp. 459–466, Aug. 1959, ISSN: 0021-9606. DOI: 10.1063/1.1730376. eprint: https://pubs.aip.org/aip/jcp/article-pdf/31/2/459/18817177/459_1_online.pdf. [Online]. Available: https://doi.org/10.1063/1.1730376 (cit. on p. 3).

[2] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller, "Equations of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, Aug. 1953, ISSN: 0021-9606. DOI: 10.1063/1.1699114 (cit. on pp. 3, 7).

[3] L. Verlet, "Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules," *Phys. Rev.*, vol. 159, pp. 98–103, 1 1967. DOI: 10.1103/PhysRev.159.98. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.159.98 (cit. on p. 3).

[4] J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes," *Journal of Computational Physics*, vol. 23, no. 3, pp. 327–341, 1977, ISSN: 0021-9991. DOI: https://doi.org/10.1016/0021-9991(77)90098-5. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0021999177900985 (cit. on p. 3).

[5] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," *Phys. Rev. Lett.*, vol. 55, pp. 2471–2474, 22 1985. DOI: 10.1103/PhysRevLett.55.2471. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.55.2471 (cit. on p. 3).

[6] S. Nose, "A molecular dynamics method for simulations in the canonical ensemble," *Molecular Physics*, vol. 52, no. 2, pp. 255–268, 1984. DOI: 10.1080/00268978400101201. eprint: https://doi.org/10.1080/00268978400101201. [Online]. Available: https://doi.org/10.1080/00268978400101201 (cit. on pp. 3, 7).

[7] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Phys. Rev. A*, vol. 31, pp. 1695–1697, 3 1985. DOI: 10.1103/PhysRevA.31.1695. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.31.1695 (cit. on p. 3).

[8] J. B. Gibson, A. N. Goland, M. Milgram and G. H. Vineyard, "Dynamics of radiation damage," *Phys. Rev.*, vol. 120, pp. 1229–1253, 4 1960. DOI: 10.1103/PhysRev.120.1229. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.120.1229 (cit. on p. 3).

[9] A. Rahman, "Correlations in the motion of atoms in liquid argon," *Phys. Rev.*, vol. 136, A405–A411, 2A 1964. DOI: 10.1103/PhysRev.136.A405. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.136.A405 (cit. on p. 3).

[10] J. A. McCammon, B. R. Gelin and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, no. 5612, pp. 585–590, Jun. 1977. DOI: 10.1038/267585a0 (cit. on p. 3).

[11] G. Ciccotti, C. Dellago, M. Ferrario, E. R. Hernández and M. E. Tuckerman, "Molecular simulations: Past, present, and future (a topical issue in epjb)," *The European Physical Journal B*, vol. 95, no. 1, p. 3, 2022, ISSN: 1434-6036. DOI: 10.1140/epjb/s10051-021-00249-x. [Online]. Available: https://doi.org/10.1140/epjb/s10051-021-00249-x (cit. on p. 3).

[12] M. C. Zwier and L. T. Chong, "Reaching biological timescales with all-atom molecular dynamics simulations," *Current Opinion in Pharmacology*, vol. 10, no. 6, pp. 745–752, 2010, Endocrine and metabolic diseases/New technologies - the importance of protein dynamics, ISSN: 1471-4892. DOI: https://doi.org/10.1016/j.coph.2010.09.008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1471489210001463 (cit. on p. 3).

[13] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Phys. Rev. Lett.*, vol. 98, p. 146 401, 14 2007. DOI: 10.1103/PhysRevLett.98.146401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.98.146401 (cit. on p. 3).

[14] A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," *Phys. Rev. Lett.*, vol. 104, p. 136 403, 13 2010. DOI: 10.1103/PhysRevLett.104.136403. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.104.136403 (cit. on p. 3).

[15] Z. Li, J. R. Kermode and A. De Vita, "Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces," *Phys. Rev. Lett.*, vol. 114, p. 096 405, 9 2015. DOI: 10.1103/PhysRevLett.114.096405. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.114.096405 (cit. on p. 3).

[16] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals and G. E. Dahl, *Neural message passing for quantum chemistry*, 2017. arXiv: 1704.01212 [cs.LG]. [Online]. Available: https://arxiv.org/abs/1704.01212 (cit. on pp. 3, 11).

[17] W. Chen and A. L. Ferguson, "Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration," *Journal of Computational Chemistry*, vol. 39, no. 25, 2079–2102, 2018, ISSN: 1096-987X. DOI: 10.1002/jcc.25520. [Online]. Available: http://dx.doi.org/10.1002/jcc.25520 (cit. on p. 3).

[18] J. Wang et al., "Machine learning of coarse-grained molecular dynamics force fields," *ACS Central Science*, vol. 5, no. 5, pp. 755–767, May 2019. DOI: 10.1021/acscentsci.8b00913 (cit. on p. 3).

[19] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan and S. Ganguli, *Deep unsupervised learning using nonequilibrium thermodynamics*, 2015. arXiv: 1503.03585 [cs.LG]. [Online]. Available: https://arxiv.org/abs/1503.03585 (cit. on pp. 3, 9).

[20] M. Nareklishvili, N. Polson and V. Sokolov, *Generative modeling: A review*, 2025. arXiv: 2501.05458 [stat.CO]. [Online]. Available: https://arxiv.org/abs/2501.05458 (cit. on pp. 3, 9).

[21] M. S. Albergo, N. M. Boffi and E. Vanden-Eijnden, *Stochastic interpolants: A unifying framework for flows and diffusions*, 2025. arXiv: 2303.08797 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2303.08797 (cit. on pp. 3, 9, 10, 17, 20).

[22] F. Noé, S. Olsson, J. Köhler and H. Wu, "Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning," *Science*, vol. 365, no. 6457, eaaw1147, 2019. DOI: doi/10.1126/science.aaw1147 (cit. on pp. 3, 8, 11–13, 19).

[23] J. Viguera Diez, S. Romeo Atance, O. Engkvist and S. Olsson, "Generation of conformational ensembles of small molecules via surrogate model-assisted molecular dynamics," *Machine Learning: Science and Technology*, vol. 5, no. 2, p. 025 010, Apr. 2024, ISSN: 2632-2153. DOI: 10.1088/2632-2153/ad3b64. [Online]. Available: http://dx.doi.org/10.1088/2632-2153/ad3b64 (cit. on p. 3).

[24] M. Schreiner, O. Winther and S. Olsson, *Implicit transfer operator learning: Multiple time-resolution surrogates for molecular dynamics*, 2023. arXiv: 2305.18046 [physics.chem-ph]. [Online]. Available: https://arxiv.org/abs/2305.18046 (cit. on pp. 3, 9, 14, 17).

[25] S. Lewis et al., "Scalable emulation of protein equilibrium ensembles with generative deep learning," *Science*, vol. 389, no. 6761, eadv9817, 2025. DOI: 10.1126/science.adv9817. eprint: https://www.science.org/doi/pdf/10.1126/science.adv9817. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.adv9817 (cit. on pp. 3, 19).

[26] L. Klein et al., *Timewarp: Transferable acceleration of molecular dynamics by learning time-coarsened dynamics*, 2023. arXiv: 2302.01170 [stat.ML]. [Online]. Available: https://arxiv.org/abs/2302.01170 (cit. on p. 3).

[27] R. P. Feynman, "Simulating physics with computers," *International Journal of Theoretical Physics*, vol. 21, no. 6, pp. 467–488, 1982, ISSN: 1572-9575. DOI: 10.1007/BF02650179. [Online]. Available: https://doi.org/10.1007/BF02650179 (cit. on p. 5).

[28] B. Fauseweh, "Quantum many-body simulations on digital quantum computers: State-of-the-art and future challenges," *Nature Communications*, vol. 15, no. 1, p. 2123, 2024, ISSN: 2041-1723. DOI: 10.1038/s41467-024-46402-9. [Online]. Available: https://doi.org/10.1038/s41467-024-46402-9 (cit. on p. 5).

[29] C. W. Gardiner, *Stochastic Methods A Handbook for the Natural and Social Sciences*, 4th ed. Berlin: Springer-Verlag, 2009, ISBN: 978-3-540-70712-7 (cit. on pp. 5, 8).

[30] H. Risken, *The Fokker–Planck Equation: Methods of Solution and Applications*, 2nd ed. Berlin, Heidelberg: Springer, 1996, ISBN: 978-3-642-61544-3. DOI: 10.1007/978-3-642-61544-3 (cit. on pp. 5–8).

[31] E. Parzen, *Stochastic Processes*. Dover Publications, 2015 (cit. on p. 6).

[32] M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press, Aug. 2023, ISBN: 9780198825562. DOI: 10.1093/oso/9780198825562.001.0001. [Online]. Available: https://doi.org/10.1093/oso/9780198825562.001.0001 (cit. on pp. 7, 8).

[33] M. Rubinstein and R. H. Colby, *Polymer Physics*. Oxford University Press, 2003, ISBN: 9780198520597 (cit. on p. 7).

[34] J.-L. Barrat and L. Berthier, "Computer simulations of the glass transition and glassy materials," *Comptes Rendus. Physique*, vol. 24, no. S1, 57–72, Apr. 2024, ISSN: 1878-1535. DOI: 10.5802/crphys.129. [Online]. Available: http://dx.doi.org/10.5802/crphys.129 (cit. on p. 8).

[35] L. Berthier and G. Biroli, "Theoretical perspective on the glass transition and amorphous materials," *Rev. Mod. Phys.*, vol. 83, pp. 587–645, 2 2011. DOI: 10.1103/RevModPhys.83.587. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.83.587 (cit. on p. 8).

[36] Z. G. Nicolaou, H. Cho, Y. Zhang, J. N. Kutz and S. L. Brunton, *Signature of glassy dynamics in dynamic modes decompositions*, 2025. arXiv: 2502.10918 [cond-mat.dis-nn]. [Online]. Available: https://arxiv.org/abs/2502.10918 (cit. on p. 8).

[37] U. Seifert, "Stochastic thermodynamics, fluctuation theorems and molecular machines," *Reports on Progress in Physics*, vol. 75, no. 12, p. 126 001, Dec. 2012. DOI: 10.1088/0034-4885/75/12/126001 (cit. on p. 8).

[38] G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," *Physical Review E*, vol. 60, no. 3, 2721–2726, 1999, ISSN: 1095-3787. DOI: 10.1103/physreve.60.2721. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.60.2721 (cit. on p. 8).

[39] U. Seifert, "Entropy production along a stochastic trajectory and an integral fluctuation theorem," *Phys. Rev. Lett.*, vol. 95, p. 040 602, 4 2005. DOI: 10.1103/PhysRevLett.95.040602. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.95.040602 (cit. on p. 8).

[40] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Physical Review Letters*, vol. 78, no. 14, 2690–2693, Apr. 1997, ISSN: 1079-7114. DOI: 10.1103/physrevlett.78.2690. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.78.2690 (cit. on p. 8).

[41] J. V. Diez, M. Schreiner and S. Olsson, *Transferable generative models bridge femtosecond to nanosecond time-step molecular dynamics*, 2025. arXiv: 2510.07589 [physics.chem-ph]. [Online]. Available: https://arxiv.org/abs/2510.07589 (cit. on pp. 9, 14, 17, 20).

[42] Y. Cheng, Y. Gong, Y. Liu, B. Song and Q. Zou, "Molecular design in drug discovery: A comprehensive review of deep generative models," *Briefings in Bioinformatics*, vol. 22, no. 6, bbab344, Nov. 2021. DOI: 10.1093/bib/bbab344 (cit. on p. 9).

[43] S. Bond-Taylor, A. Leach, Y. Long and C. G. Willcocks, "Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, 7327–7347, Nov. 2022, ISSN: 1939-3539. DOI: 10.1109/tpami.2021.3116668. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2021.3116668 (cit. on p. 9).

[44] I. Kobyzev, S. J. Prince and M. A. Brubaker, "Normalizing flows: An introduction and review of current methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, 3964–3979, Nov. 2021, ISSN: 1939-3539. DOI: 10.1109/tpami.2020.2992934. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2020.2992934 (cit. on p. 9).

[45] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel and M. Le, *Flow matching for generative modeling*, 2023. arXiv: 2210.02747 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2210.02747 (cit. on pp. 9, 10).

[46] X. Liu, C. Gong and Q. Liu, *Flow straight and fast: Learning to generate and transfer data with rectified flow*, 2022. arXiv: 2209.03003 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2209.03003 (cit. on p. 9).

[47] E. Coddington and N. Levinson, *Theory Of Ordinary Differential Equations*. McGraw-Hill Education, 1984, ISBN: 978-0070992566 (cit. on p. 10).

[48] Q. Liu, "Rectified flow: A marginal preserving approach to optimal transport," *arXiv preprint arXiv:2209.14577*, 2022 (cit. on p. 10).

[49] A. Tong et al., "Improving and generalizing flow-based generative models with minibatch optimal transport," *Transactions on Machine Learning Research*, 2024, Expert Certification, ISSN: 2835-8856. [Online]. Available: https://openreview.net/forum?id=CD9Snc73AW (cit. on p. 10).

[50] L. Klein, A. Krämer and F. Noé, *Equivariant flow matching*, 2023. arXiv: 2306.15030 [stat.ML]. [Online]. Available: https://arxiv.org/abs/2306.15030 (cit. on pp. 10–13).

[51] J. Köhler, L. Klein and F. Noe, "Equivariant flows: Exact likelihood generative learning for symmetric densities," in *Proceedings of the 37th International Conference on Machine Learning*, H. D. III and A. Singh, Eds., ser. Proceedings of Machine Learning Research, vol. 119, PMLR, Jul. 2020, pp. 5361–5370. [Online]. Available: https://proceedings.mlr.press/v119/kohler20a.html (cit. on pp. 10–12).

[52] N. M. Boffi, M. S. Albergo and E. Vanden-Eijnden, *How to build a consistency model: Learning flow maps via self-distillation*, 2025. arXiv: 2505.18825 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2505.18825 (cit. on pp. 10, 17).

[53] Y. Song, P. Dhariwal, M. Chen and I. Sutskever, *Consistency models*, 2023. arXiv: 2303.01469 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2303.01469 (cit. on p. 10).

[54] C. Chi et al., *Diffusion policy: Visuomotor policy learning via action diffusion*, 2024. arXiv: 2303.04137 [cs.RO]. [Online]. Available: https://arxiv.org/abs/2303.04137 (cit. on p. 10).

[55] R. T. Q. Chen and D. Duvenaud, "Neural networks with cheap differential operators," *CoRR*, vol. abs/1912.03579, 2019. arXiv: 1912.03579. [Online]. Available: http://arxiv.org/abs/1912.03579 (cit. on pp. 11, 14).

[56] M. Hutchinson, "A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines," *Communications in Statistics - Simulation and Computation*, vol. 19, no. 2, 433–450, Jan. 1990, ISSN: 1532-4141. DOI: 10.1080/03610919008812866. [Online]. Available: http://dx.doi.org/10.1080/03610919008812866 (cit. on p. 11).

[57] M. Weiler, P. Forré, E. Verlinde and M. Welling, *Equivariant and Coordinate Independent Convolutional Networks, A Gauge Field Theory of Neural Networks*. 2023. [Online]. Available: https://maurice-weiler.gitlab.io/cnn_book/EquivariantAndCoordinateIndependentCNNs.pdf (cit. on pp. 11, 12).

[58] R. Kondor, "The principles behind equivariant neural networks for physics and chemistry," *Proceedings of the National Academy of Sciences*, vol. 122, no. 41, e2415656122, 2025. DOI: 10.1073/pnas.2415656122. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2415656122. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.2415656122 (cit. on p. 12).

[59] K. T. Schütt, O. T. Unke and M. Gastegger, "Equivariant message passing for the prediction of tensorial properties and molecular spectra," *CoRR*, vol. abs/2102.03150, 2021. arXiv: 2102.03150. [Online]. Available: https://arxiv.org/abs/2102.03150 (cit. on p. 12).

[60] H. Schopmans, C. von Klitzing and P. Friederich, *Efficient training of boltzmann generators using off-policy log-dispersion regularization*, 2026. arXiv: 2602.03729 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2602.03729 (cit. on pp. 13, 19).

[61] C. von Klitzing, D. Blessing, H. Schopmans, P. Friederich and G. Neumann, *Learning boltzmann generators via constrained mass transport*, 2026. arXiv: 2510.18460 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2510.18460 (cit. on pp. 13, 19).

[62] L. Kish, *Survey Sampling*. John Wiley & Sons, 1965 (cit. on p. 13).

[63] S. Moqvist, W. Chen, M. Schreiner, F. Nüske and S. Olsson, "Thermodynamic interpolation: A generative approach to molecular thermodynamics and kinetics," *Journal of Chemical Theory and Computation*, vol. 21, no. 5, 2535–2545, Feb. 2025, ISSN: 1549-9626. DOI: 10.1021/acs.jctc.4c01557. [Online]. Available: http://dx.doi.org/10.1021/acs.jctc.4c01557 (cit. on p. 13).

[64] J. F. Gloy and S. Olsson, *Hollowflow: Efficient sample likelihood evaluation using hollow message passing*, 2025. arXiv: 2510.21542 [stat.ML]. [Online]. Available: https://arxiv.org/abs/2510.21542 (cit. on p. 13).

[65] D. Rehman, T. Akhound-Sadegh, A. Gazizov, Y. Bengio and A. Tong, *Falcon: Few-step accurate likelihoods for continuous flows*, 2025. arXiv: 2512.09914 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2512.09914 (cit. on p. 14).

[66] G. R. Bowman, C. Clementi and X. Huang, "Introduction to markov state modeling of conformational dynamics," *Journal of Chemical Theory and Computation*, vol. 22, no. 6, pp. 2665–2667, 2026, ISSN: 1549-9618. DOI: 10.1021/acs.jctc.6c00351. [Online]. Available: https://doi.org/10.1021/acs.jctc.6c00351 (cit. on p. 14).

[67] V. S. Pande, K. Beauchamp and G. R. Bowman, "Everything you wanted to know about markov state models but were afraid to ask," *Methods*, vol. 52, no. 1, pp. 99–105, Sep. 2010. DOI: 10.1016/j.ymeth.2010.06.002 (cit. on p. 14).

[68] J. D. Chodera and F. Noé, "Markov state models of biomolecular conformational dynamics," *Current Opinion in Structural Biology*, vol. 25, pp. 135–144, Apr. 2014. DOI: 10.1016/j.sbi.2014.04.002 (cit. on p. 14).

[69] A. Mardt, L. Pasquali, H. Wu and F. Noé, "Vampnets for deep learning of molecular kinetics," *Nature Communications*, vol. 9, no. 1, p. 5, Jan. 2018, ISSN: 2041-1723. DOI: 10.1038/s41467-017-02388-1. [Online]. Available: https://doi.org/10.1038/s41467-017-02388-1 (cit. on p. 14).

[70] F. Nüske, P. Gelß, S. Klus and C. Clementi, "Tensor-based computation of metastable and coherent sets," *Physica D: Nonlinear Phenomena*, vol. 427, p. 133 018, 2021, ISSN: 0167-2789. DOI: https://doi.org/10.1016/j.physd.2021.133018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167278921001755 (cit. on p. 14).

[71] S. Klus et al., "Data-driven model reduction and transfer operator approximation," *Journal of Nonlinear Science*, vol. 28, no. 3, pp. 985–1010, Jun. 2018, ISSN: 1432-1467. DOI: 10.1007/s00332-017-9437-7. [Online]. Available: https://doi.org/10.1007/s00332-017-9437-7 (cit. on p. 14).

[72] J. V. Diez, M. J. Schreiner, O. Engkvist and S. Olsson, "Boltzmann priors for implicit transfer operators," in *The Thirteenth International Conference on Learning Representations*, 2025. [Online]. Available: https://openreview.net/forum?id=pRCOZllZdT (cit. on pp. 14, 17).

[73] P. Antoniadis, B. Pavesi, S. Olsson and O. Winther, *Protein language model embeddings improve generalization of implicit transfer operators*, 2026. arXiv: 2602.11216 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2602.11216 (cit. on pp. 14, 17).

[74] Y. Chen, M. Goldstein, M. Hua, M. S. Albergo, N. M. Boffi and E. Vanden-Eijnden, *Probabilistic forecasting with stochastic interpolants and föllmer processes*, 2024. arXiv: 2403.13724 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2403.13724 (cit. on p. 14).

[75] C. Bustamante, D. Keller and G. Oster, "The physics of molecular motors," *Accounts of Chemical Research*, vol. 34, no. 6, pp. 412–420, Jun. 2001, ISSN: 0001-4842. DOI: 10.1021/ar0001719. [Online]. Available: https://doi.org/10.1021/ar0001719 (cit. on p. 17).

[76] L. Klein and F. Noé, *Transferable boltzmann generators*, 2025. arXiv: 2406.14426 [stat.ML]. [Online]. Available: https://arxiv.org/abs/2406.14426 (cit. on p. 19).