



PAPER

OPEN ACCESS

RECEIVED
9 July 2024REVISED
7 September 2024ACCEPTED FOR PUBLICATION
8 October 2024PUBLISHED
22 November 2024

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Efficient mapping of phase diagrams with conditional Boltzmann Generators

Maximilian Schebek^{1,*} , Michele Invernizzi² , Frank Noe^{1,2,3,4} and Jutta Rogal^{1,5} ¹ Fachbereich Physik, Freie Universität Berlin, 14195 Berlin, Germany² Fachbereich Mathematik und Informatik, Freie Universität Berlin, 14195 Berlin, Germany³ Department of Chemistry, Rice University, Houston, TX 77005, United States of America⁴ Microsoft Research AI4Science, 10178 Berlin, Germany⁵ Department of Chemistry, New York University, New York, NY 10003, United States of America

* Author to whom any correspondence should be addressed.

E-mail: m.schebek@fu-berlin.de**Keywords:** machine learning, statistical mechanics, generative modelsSupplementary material for this article is available [online](#)

Abstract

The accurate prediction of phase diagrams is of central importance for both the fundamental understanding of materials as well as for technological applications in material sciences. However, the computational prediction of the relative stability between phases based on their free energy is a daunting task, as traditional free energy estimators require a large amount of simulation data to obtain uncorrelated equilibrium samples over a grid of thermodynamic states. In this work, we develop deep generative machine learning models based on the Boltzmann Generator approach for entire phase diagrams, employing normalizing flows conditioned on the thermodynamic states, e.g. temperature and pressure, that they map to. By training a single normalizing flow to transform the equilibrium distribution sampled at only one reference thermodynamic state to a wide range of target temperatures and pressures, we can efficiently generate equilibrium samples across the entire phase diagram. Using a permutation-equivariant architecture allows us, thereby, to treat solid and liquid phases on the same footing. We demonstrate our approach by predicting the solid–liquid coexistence line for a Lennard-Jones system in excellent agreement with state-of-the-art free energy methods while significantly reducing the number of energy evaluations needed.

1. Introduction

When studying or designing materials, it is crucial to understand their phase diagrams, that is the relative stability of different phases as a function of temperature and pressure. Experimentally, the measurement of phase diagrams is challenging and often it remains uncertain whether the observed phases are truly the most stable ones and whether all possible phases have been discovered. The accurate and efficient prediction of phase diagrams is, therefore, one of the central challenges in computational materials science [1, 2].

The determination of relative stability requires the evaluation of the free energies of all phases over the entire range of thermodynamic states of interest, which is a computationally complex and demanding task. Ideally, free energy calculations should be performed with *ab initio* accuracy of interatomic interactions using large enough simulation cells and an exhaustive sampling of the phase space. Each of these factors, highly accurate interactions, large number of particles and extensive sampling, can render the calculations prohibitively expensive and, in general, at least one of them needs to be compromised in order to make the approach computationally feasible. On the sampling side, an accurate estimate of the equilibrium distributions requires the exploration of the phase space with correct probability, which is generally achieved through the use of trajectory-based molecular simulation techniques such as Markov chain Monte Carlo (MC) and molecular dynamics (MD). These methods sample the equilibrium distribution by gradually changing the system configuration, thus requiring a large number of simulation steps for generating a

sufficient number of statistically independent samples. Furthermore, traditional estimators for free energy differences, such as free energy perturbation (FEP) [3] and its multistate extension, the multistate Bennett acceptance ratio (MBAR) [4], require sufficient overlap in phase space between adjacent states for convergence. Therefore, a grid of simulation points needs to be defined over the relevant range of thermodynamic conditions, leading to a large number of simulations, all of which must be run until convergence. The required number of grid points for such thermodynamic integration is strongly system and state dependent, making the choice of a suitable grid a tedious trial-and-error procedure [5]. Targeted FEP (TFEP) [6] tackles the overlap problem by defining an invertible map on the configuration space. A suitable map increases the phase space overlap between states which reduces the number of required simulations points. However, defining a suitable map is far from trivial and TFEP has mostly been applied to simple problems where physical intuition was key to identifying such a map [7].

In recent years, machine learning (ML) models have found broad use in molecular simulations. ML interatomic potentials have been shown to strongly reduce the cost of *ab initio* MD simulations while retaining an accuracy comparable to density functional theory or other quantum chemistry methods [8–10]. Lately, these potentials have also been applied in the calculation of phase diagrams [11, 12] providing faster energy evaluations while the sampling challenge remains. Complementary, deep generative models are being developed to generate uncorrelated samples of molecular structures in one go, with the aim of overcoming the sampling problem typical of MD and MC. In particular, the Boltzmann Generator (BG) approach [13] leverages a combination of highly expressive invertible coordinate transformations, such as normalizing flows [14, 15] and statistical mechanics methods such as importance sampling or Metropolis–Hastings MC, to generate independent identically distributed samples of the targeted equilibrium distribution. This approach has emerged both in the molecular sciences [13] as well as in lattice field theory [16, 17]. Within the TFEP-based learned FEP (LFEP), using flows enables to learn flexible mappings for systems where physical intuition is insufficient [18]. This approach has successfully been applied to sample small solvation systems [19], monoatomic solids [20–22] and small molecules with complex quantum-mechanical potentials based on cheaper potentials [23] or at different temperatures to overcome slow modes [24]. However, while LFEP offers an elegant way to learn a mapping between two specific thermodynamic states, it cannot be used efficiently for the calculation of phase diagrams since a new flow model needs to be trained for each thermodynamic state.

In this work, we generalize the LFEP and BG approaches by employing normalizing flows conditioned on the target thermodynamic states. This conditional BG is capable of mapping MD samples from only one reference thermodynamic state across the entire (T, P) -range of the phase diagram with a single trained model. By training to match the equilibrium distributions at all thermodynamic states and appropriately reweighting the generated samples, free energy differences can be very efficiently computed for any temperature and pressure. In this way, the coexistence line between phases can be mapped out with only minimal prior knowledge of its location, while avoiding the need to conduct numerous MD simulations across a grid of thermodynamic states. Using a permutation-equivariant architecture [18], our approach further treats ordered and disordered phases on the same footing. We demonstrate our developed framework by determining the coexistence line between solid and liquid phases of a Lennard-Jones (LJ) system consisting of 180 interacting particles, achieving highly accurate free energy estimates for both phases over the entire range of temperatures and pressures.

2. Theoretical background

2.1. Traditional free energy estimators

We consider the isothermal-isobaric (NPT) ensemble at constant number of particles N , temperature T and pressure P . For this ensemble, the reduced potential $u(\mathbf{x}, V)$ is defined as [4]

$$u(\mathbf{x}, V) = \beta(U(\mathbf{x}) + PV), \quad (1)$$

where $U(\mathbf{x})$ is the potential energy, V is the volume of the simulation box at configuration \mathbf{x} and $\beta = 1/k_B T$ with Boltzmann's constant k_B . The reduced potential defines the equilibrium distribution $q(\mathbf{x}, V)$ by

$$q(\mathbf{x}, V) = Z^{-1} \exp(-u(\mathbf{x}, V)), \quad (2)$$

where $Z = \int_{\Gamma} dV d\mathbf{x} q(\mathbf{x}, V)$ is the configurational partition function given by the integral over the configuration space Γ .

From the configurational partition function, the dimensionless configurational Gibbs free energy $f = \beta G$ can be obtained as

$$f = -\log Z. \quad (3)$$

Since evaluating the integral over configuration space is unfeasible for most interaction potentials, evaluating the absolute free energy is usually impossible. A more tractable approach consists in evaluating the difference in free energy, $\Delta f_{AB} = f_B - f_A$, between two states A and B characterized by two different reduced potentials u_A and u_B . In particular, many of the most widely used free energy estimators are based on the Zwanzig identity [3], the central result of FEP:

$$\Delta f_{AB} = -\log \mathbb{E}_{(\mathbf{x}, V) \sim q_A} [\exp(-\Delta u_{AB}(\mathbf{x}, V))]. \quad (4)$$

Here, $\Delta u_{AB}(\mathbf{x}, V) = u_B(\mathbf{x}, V) - u_A(\mathbf{x}, V)$ is the difference in the reduced potential between the two states and $\mathbb{E}_{(\mathbf{x}, V) \sim q_A}$ denotes the expectation value evaluated using samples drawn from q_A . While being formally exact, the convergence of FEP crucially depends on the overlap between the corresponding configurational distributions $q_A(\mathbf{x}, V)$ and $q_B(\mathbf{x}, V)$, making it a highly biased and noisy estimator in practical calculations, even for small molecules [25].

A statistically more optimal estimator can be obtained using samples from both states of interest as done in BAR and its multistate extension MBAR [4]. Nevertheless, the overlap problem remains which is commonly solved by defining a sequence of intermediate Hamiltonians between the two states of interest. However, such multi-staged approaches require samples from all intermediate states and it is not clear how best to define the intermediate stages.

TFEP [6] tackles the problem of having intermediate states by defining an invertible map \mathcal{M} transforming samples (\mathbf{x}, V) from q_A to samples $(\mathbf{x}', V') = \mathcal{M}(\mathbf{x}, V)$ from a new target distribution q'_A which shares a larger overlap with q_B than q_A . For a perfect map, $q'_A = q_B$. Similarly, the inverse map \mathcal{M}^{-1} transforms samples from q_B to q'_B and $q'_B = q_A$ for a perfect map. Using the determinant of the Jacobian of the mapping, $\det J_{\mathcal{M}}(\mathbf{x}, V)$, the generated distribution can be computed analytically using the change of variable theorem

$$q'_A(\mathbf{x}', V') = q_A(\mathbf{x}, V) |\det J_{\mathcal{M}}(\mathbf{x}, V)|^{-1}. \quad (5)$$

Samples from the generated distribution can be reweighted to the target distribution by assigning a statistical weight w , which can be computed using various algorithms [26]. Here, we follow [13] and define the importance weights of the generated samples as

$$w(\mathbf{x}, V) \propto \frac{q_B(\mathcal{M}(\mathbf{x}, V))}{q'_A(\mathcal{M}(\mathbf{x}, V))}. \quad (6)$$

The dimensionless free energy difference between A and B can then be obtained as

$$\Delta f_{AB} = -\log \mathbb{E}_{(\mathbf{x}, V) \sim q_A} [w(\mathbf{x}, V)]. \quad (7)$$

For the probability distributions of interest (equation (2)), the importance weights are given by

$$\log w(\mathbf{x}, V) = u_A(\mathbf{x}, V) - u_B(\mathcal{M}(\mathbf{x}, V)) + \log |\det J_{\mathcal{M}}(\mathbf{x}, V)|. \quad (8)$$

The weights can further be utilized to evaluate how well the learned distribution approximates the target distribution through the Kish effective sample size (ESS) [27]

$$\text{ESS} = \frac{[\sum_i w(\mathbf{x}_i, V_i)]^2}{\sum_i [w(\mathbf{x}_i, V_i)]^2}. \quad (9)$$

The ESS provides a rough measure for how many uncorrelated samples would result in a MC estimator of the same statistical performance (typically reported relative to the total number of generated samples).

The key idea of LFEP [18] consists in parametrizing \mathcal{M} as a normalizing flow which removes the need of crafting a suitable mapping based on physical intuition only.

2.2. Normalizing flows

Normalizing flows are a class of invertible deep generative models which are able to generate samples from a target distribution [14, 28]. In contrast to other generative models, the use of invertible transformations allows to evaluate the generated distribution using the change-of-variables technique, making them an attractive sampling tool for the physical sciences. Normalizing flows aim to learn the optimal set of parameters θ^* of an invertible function $f_{\theta}(\mathbf{x})$ with trainable parameters θ that maps samples \mathbf{x} from a prior distribution $q_A(\mathbf{x})$ into samples from a target distribution $q_B(\mathbf{x})$. In practice, the learned set of parameters θ will be different from θ^* and the flow is trained by minimizing the Kullback–Leibler (KL) divergence, $D_{\text{KL}}(q'_A || q_B)$, between the generated distribution $q'_A(\mathbf{x})$ and $q_B(\mathbf{x})$.

BGs target equilibrium distributions similar to equation (2). In this case, the KL divergence can be expressed as [13, 18]:

$$D_{\text{KL}}(q'_A || q_B) = -\mathbb{E}_{\mathbf{x} \sim q_A} [\log w(\mathbf{x})] - \Delta f_{AB}, \quad (10)$$

where the importance weights are defined similarly to equation (6) as $w(\mathbf{x}) \propto q_B(f_\theta(\mathbf{x})) / q'_A(f_\theta(\mathbf{x}))$ and Δf_{AB} is the free energy difference between A and B . Since only the importance weights depend on the trainable parameters through the function $f_\theta(\mathbf{x})$ while Δf_{AB} is constant, the corresponding loss function simplifies to

$$\mathcal{L}_{AB}(\theta) = -\mathbb{E}_{\mathbf{x} \sim q_A} [\log w(\mathbf{x})] \geq \Delta f_{AB} \quad (11)$$

and provides an upper bound for the free energy difference between A and B .

For efficient training of normalizing flows, the determinant of the Jacobian of the transformation must be easy to calculate (see equation (6)). For this purpose, coupling layers [29, 30] were designed which split the input data into two channels and transform one of the channels conditioned on the other. This yields a triangular Jacobian whose determinant can be evaluated analytically. The entire input vector can be transformed by stacking several coupling layers, whereby the channel to be transformed changes.

Flows can further be trained in a conditional setting which allows the modeling of conditional target distributions [31, 32]. A conditional normalizing flow $f_\theta(\mathbf{x}|\mathbf{c})$ is trained to learn a mapping between a prior distribution $q_A(\mathbf{x})$ and a conditional target distribution $q_B(\mathbf{x}'|\mathbf{c})$ for a vector of conditioning variables \mathbf{c} [31, 32]. In this case, the change of variable reads

$$q'_A(\mathbf{x}'|\mathbf{c}) = q_A(\mathbf{x}) |\det J_f(\mathbf{x}|\mathbf{c})|^{-1}. \quad (12)$$

For energy-based training, the trainable parts of the conditional KL divergence are given by the average of the KL divergences between the generated and the target distributions conditioned on each target state over the entire range of \mathbf{c} , yielding the following loss function:

$$\mathcal{L}(\theta) = -\mathbb{E}_{\mathbf{c} \sim p_c} \mathbb{E}_{\mathbf{x} \sim q_A} [\log w(\mathbf{x}|\mathbf{c})] \quad (13)$$

with the conditional importance weights defined as $w(\mathbf{x}|\mathbf{c}) \propto q_B(f_\theta(\mathbf{x}|\mathbf{c})|\mathbf{c}) / q'_A(f_\theta(\mathbf{x}|\mathbf{c})|\mathbf{c})$. In practice, conditioning a coupling flow architecture can be easily achieved by concatenating the conditioning variable to the unchanged part before transforming the other part [31]. While originally developed in the context of image generation tasks [31, 32], the idea of conditioning flows has recently been applied to the sampling of rare events and to lattice field theories [33, 34].

2.3. Computing coexistence lines

The coexistence line in the (T, P) diagram of two phases of a system is defined by equal total Gibbs free energies $\tilde{G}(T, P)$. For a given state (T_i, P_i) , \tilde{G}_i is the sum of the configurational contribution (equation (3)) and the kinetic contribution:

$$\tilde{G}_i = G_i + \frac{3N}{\beta_i} \log \Lambda(T_i), \quad (14)$$

where $\Lambda(T_i)$ is the thermal De-Broglie wavelength $\Lambda(T_i) = (\frac{h^2}{2\pi m k_B T_i})^{1/2}$ with Planck's constant h and the particle mass m .

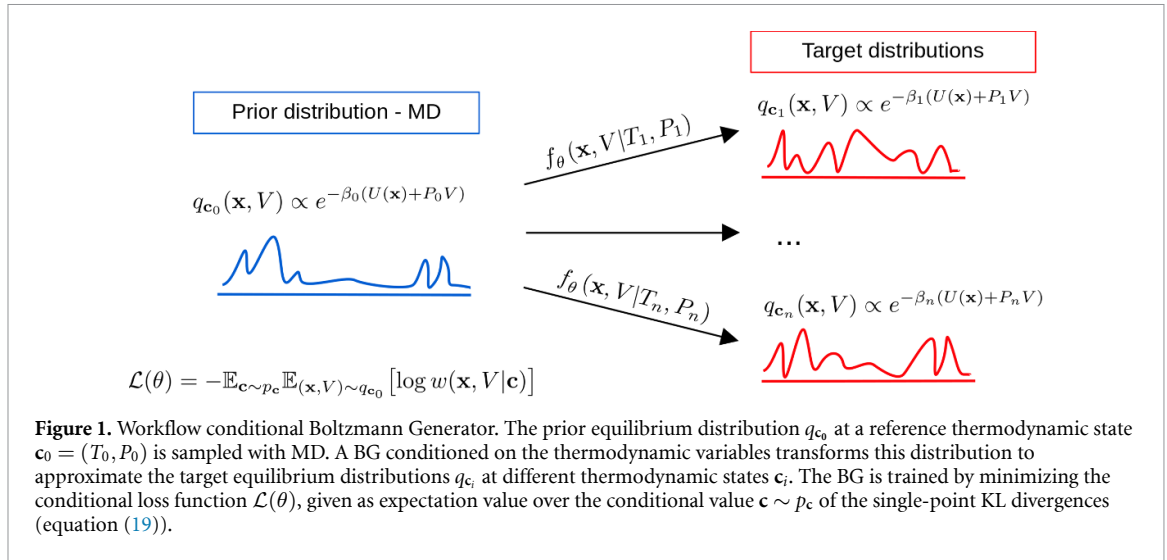
Using free energy estimators such as MBAR or TFEP, the configurational contribution G for each phase can be obtained by combining relative differences within phases with the absolute free energy G_{ref} at a reference state $(T_{\text{ref}}, P_{\text{ref}})$. Writing the difference in the dimensionless free energy between the reference point and a state i , $\Delta f_{\text{ref},i}$ explicitly as

$$\Delta f_{\text{ref},i} = \beta_i G_i - \beta_{\text{ref}} G_{\text{ref}}, \quad (15)$$

the absolute Gibbs free energy can be obtained as

$$G_i = \frac{1}{\beta_i} \Delta f_{\text{ref},i} + \frac{\beta_{\text{ref}}}{\beta_i} G_{\text{ref}}. \quad (16)$$

G_{ref} is not directly available from simulations of the system of interest. Instead, it can be obtained by connecting the physical system to a model system whose Helmholtz free energy at the reference point,



$F_{\text{ref}}^0 = F^0(T_{\text{ref}}, V_{\text{ref}})$ with $V_{\text{ref}} = \langle V \rangle_{T_{\text{ref}}, P_{\text{ref}}}$, can be computed analytically. Then, the reference free energy can be computed as

$$G_{\text{ref}} = F_{\text{ref}}^0 + \Delta F_{\text{ref}}^0 + P_{\text{ref}} V_{\text{ref}} \quad (17)$$

with $\Delta F_{\text{ref}}^0 = F_{\text{ref}} - F_{\text{ref}}^0$ being the difference in Helmholtz free energy between the model and the physical system at the reference temperature. This quantity can be computed defining a set of intermediate Hamiltonians between physical and model system for which NVT MD simulations are performed. The resulting energy histograms serve as input for MBAR. In this work, we use as reference systems for the solid and the liquid phase the Einstein crystal [5] and the Uhlenbeck–Ford (UF) model [35], respectively (see SI for details).

3. Conditional framework

3.1. Conditioning BGs on temperature and pressure

The key idea behind our approach is to compute free energy differences over a wide range of thermodynamic states without the need to run MD simulations at each state point. This is achieved by training a BG conditioned on temperature and pressure, which learns the mapping from a prior distribution at a reference thermodynamic state to a whole family of target distributions across thermodynamic space.

Figure 1 illustrates the workflow of our approach. We first sample the prior distribution at the reference state in the NPT ensemble with MD and save samples consisting of the atomic configuration \mathbf{x} and the box volume V . In this work, we consider rectangular boxes and only allow for isotropic scaling, but the proposed approach is completely general and can easily be extended to fully triclinic simulation boxes. Samples (\mathbf{x}, V) from the prior distribution q_{c_0} of the reference state $\mathbf{c}_0 = (T_0, P_0)$ with temperature T_0 and pressure P_0 are transformed by a flow conditioned on the thermodynamic target state $\mathbf{c} = (T, P)$ according to $(\mathbf{x}', V') = f_\theta(\mathbf{x}, V|\mathbf{c})$ in order to approximate equilibrium distributions at multiple different thermodynamic conditions. Conditioning the transformation on temperature and pressure is achieved by feeding the thermodynamic state as an additional input to the model during training (see section 3.2) [31].

To optimize the flow parameters, we employ conditional energy-based training (equation (13)), avoiding the need to sample the target distributions. Writing the conditional NPT target equilibrium distribution as

$$q(\mathbf{x}, V|\mathbf{c}) \propto e^{-u(\mathbf{x}, V|\mathbf{c})} \quad (18)$$

with the reduced potential defined in equation (1) and adjusting the importance weights in equation (8) to the conditional setting, this yields the following loss function:

$$\mathcal{L}(\theta) = -\mathbb{E}_{(T, P) \sim p_{\mathbf{c}}} \mathbb{E}_{(\mathbf{x}, V) \sim q_{c_0}} [\log w(\mathbf{x}, V|\mathbf{c})] \quad (19)$$

$$= \mathbb{E}_{(T, P) \sim p_{\mathbf{c}}} \mathbb{E}_{(\mathbf{x}, V) \sim q_{c_0}} [-\beta_0(U(\mathbf{x}) + P_0V) + \beta(U(\mathbf{x}') + PV') - \log |\det J_f(\mathbf{x}, V|T, P)|]. \quad (20)$$

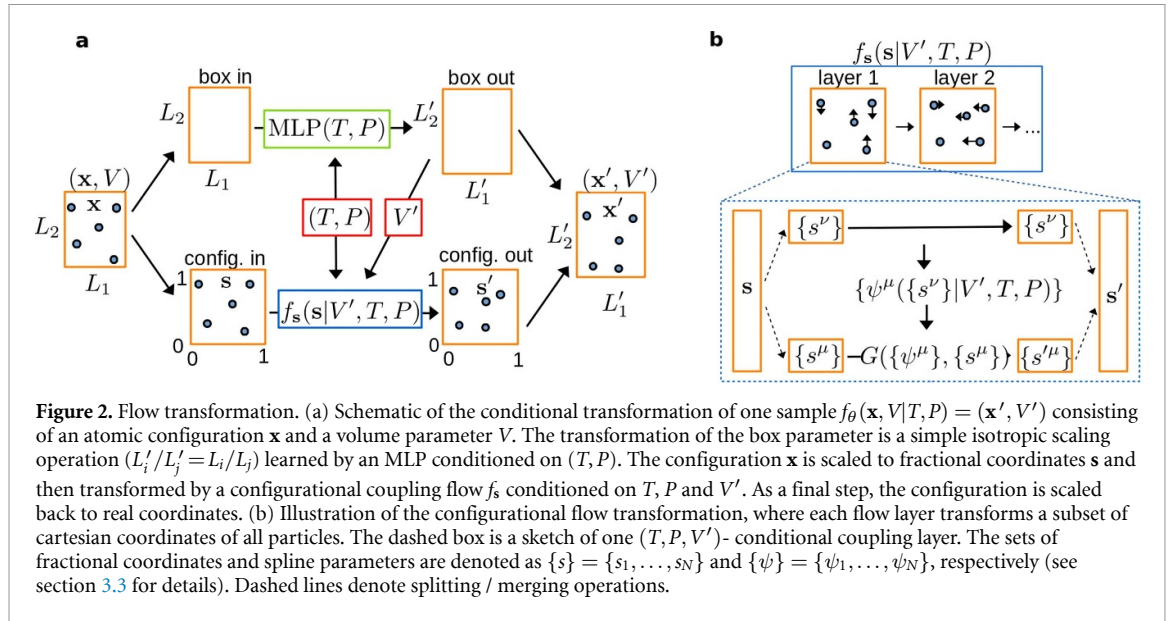


Figure 2. Flow transformation. (a) Schematic of the conditional transformation of one sample $f_{\theta}(\mathbf{x}, V|T, P) = (\mathbf{x}', V')$ consisting of an atomic configuration \mathbf{x} and a volume parameter V . The transformation of the box parameter is a simple isotropic scaling operation ($L'_i/L'_j = L_i/L_j$) learned by an MLP conditioned on (T, P) . The configuration \mathbf{x} is scaled to fractional coordinates \mathbf{s} and then transformed by a configurational coupling flow $f_{\mathbf{s}}$ conditioned on T, P and V' . As a final step, the configuration is scaled back to real coordinates. (b) Illustration of the configurational flow transformation, where each flow layer transforms a subset of cartesian coordinates of all particles. The dashed box is a sketch of one (T, P, V') -conditional coupling layer. The sets of fractional coordinates and spline parameters are denoted as $\{s\} = \{s_1, \dots, s_N\}$ and $\{\psi\} = \{\psi_1, \dots, \psi_N\}$, respectively (see section 3.3 for details). Dashed lines denote splitting / merging operations.

The conditioning states $\mathbf{c} = (T, P)$ are drawn from the distribution $p_{\mathbf{c}}$ and for brevity we used the notation $q(\mathbf{x}, V|\mathbf{c}) = q_{\mathbf{c}}(\mathbf{x}, V)$. After training, equilibrium expectation values of an observable $O(\mathbf{x}, V)$ can be computed as

$$\mathbb{E}_{(\mathbf{x}, V) \sim q_{\mathbf{c}}} [O(\mathbf{x}, V)] \approx \frac{\mathbb{E}_{(\mathbf{x}, V) \sim q_{\mathbf{c}_0}} [w(\mathbf{x}, V|\mathbf{c}) O(\mathbf{x}', V')]}{\mathbb{E}_{(\mathbf{x}, V) \sim q_{\mathbf{c}_0}} [w(\mathbf{x}, V|\mathbf{c})]}. \quad (21)$$

The free energy difference between the prior distribution and the target distribution is obtained from TFEP (equation (7)) as

$$\Delta f_{\mathbf{c}_0 \mathbf{c}} = -\log \mathbb{E}_{(\mathbf{x}, V) \sim q_{\mathbf{c}_0}} [w(\mathbf{x}, V|\mathbf{c})]. \quad (22)$$

Compared to temperature steerable flows [36], the proposed workflow is more general and allows for more flexible models, as no requirements need to be placed on the architecture or the prior.

The evaluation of equation (19) involves the average of the KL divergences over multiple thermodynamic target states. One option to choose these target states is to define a set of discrete points which are used throughout training [31, 33]. However, for n conditioning points this results in n times the number of energy evaluations per sample. In addition, the spacing between the points has to be small enough such that the flow is able to efficiently interpolate between them. We overcome these limitations by randomly drawing a new target state for each sample at each evaluation of the conditional loss function (equation (19)) from a uniform distribution defined over the range of thermodynamic conditions of interest. This approach is computationally more efficient and also allows to condition over a continuous range rather than for a fixed number points.

3.2. Flow architecture

The flow transforms a sample of the prior distribution (\mathbf{x}, V) to (\mathbf{x}', V') , where the volume is connected to the box dimensions \mathbf{L} by $V = \prod_{i=1}^3 L_i$. In this work, we focus on the simulation of solids and liquids and therefore apply certain restrictions regarding the shape of the simulation box. We use orthorhombic simulation boxes with $\mathbf{L} = [L_1, L_2, L_3]$ and further only allow for an isotropic scaling of the box ($L_i/L_j = \text{const.}$), such that the box is fully characterized by its volume $V = \prod_{i=1}^3 L_i$.

Our flow is based on a shape-conditional architecture (see figure 2(a)) [21]. Here, the volume is updated by an affine transformation, while the atomic configurations $\mathbf{x} \in \mathbb{R}^{3N}$ are first scaled to fractional coordinates $\mathbf{s} \in [0, 1]^{3N}$ and then updated conditioned on the volume and the thermodynamic state. The volume transformation and the transformation of the position of particle i can be written as

$$V' = (1 + \alpha(T, P)) \cdot V + \beta(T, P), \quad (23)$$

$$\mathbf{s}_i = \mathbf{x}_i \odot \mathbf{L}, \quad (24)$$

$$\mathbf{s}'_i = f_{\mathbf{s}, i}(\mathbf{s}|V', T, P), \quad (25)$$

$$\mathbf{x}'_i = \mathbf{s}'_i \odot \mathbf{L}'. \quad (26)$$

Here, α and β are the output of a multilayer perceptron (MLP) taking as input pressure and temperature, where α is bounded below by -1 . \odot and \oslash denote element-wise multiplication and division, respectively. The transformed box dimensions are given by $\mathbf{L}' = (V'/V)^{1/3}\mathbf{L}$. $f_{s,i}$ is the output for particle i of a (V, T, P) -conditional coupling flow over configurations (see section 3.3), operating on fractional coordinates. After the transformation, \mathbf{s} are transformed to physical coordinates by multiplying with the new box lengths.

The procedure just described gives rise to three contributions to the Jacobian, $J(V)$, $J_{\text{scale}}(\mathbf{x})$ and $J(\mathbf{s})$. $J(V)$ arises from equation (23) and is given by $J(V) = \partial V'/\partial V = \alpha(T, P)$. J_{scale} accounts for the initial global scaling of the coordinates to fractional coordinates and the final scaling of the flow output to physical coordinates (equations (24) and (26)) and is given by $J_{\text{scale}}(\mathbf{x}) = (V'/V)^N$. Finally, $J(\mathbf{s}) = \partial \mathbf{s}'/\partial \mathbf{s}$ accounts for the transformation in equation (25) and can be obtained from f_s .

3.3. Conditional coupling layer

As condensed phase systems can exist in the liquid as well as in the solid phase, it is crucial to treat both phases on the same footing. Concretely, f_s must respect periodic boundary conditions and invariance under particle permutation. In addition, the center of mass needs to be fixed to prevent uncontrolled errors in the free energy prediction due to rigid translations [20].

To achieve these requirements, we extend the coupling layer originally proposed in [18] and refined in [19] to the case of conditional sampling. Figure 2(b) shows our implementation of a (T, P, V) conditional coupling layer. The N -particle configuration \mathbf{s} is split over cartesian coordinates into two subsets $\{s^\nu\} = \{s_1^\nu, \dots, s_N^\nu\}$ and $\{s^\mu\} = \{s_1^\mu, \dots, s_N^\mu\}$, e.g. $\nu = \{x\}$ and $\mu = \{y, z\}$. All coordinates indexed by ν remain unchanged, $s_i^\nu = s_i^\nu$, while the coordinates of atom i carrying index μ are transformed according to

$$s_i^\mu = G(s_i^\mu, \psi_i^\mu), \quad (27)$$

with

$$\psi_i^\mu = C_i^\mu(s_1^\nu, \dots, s_N^\nu | V', T, P). \quad (28)$$

Here, G is a circular spline [37], which naturally accounts for the periodicity of the problem. The parameters of the spline ψ_i^μ are obtained as the output of the network C for the coordinates indexed by μ of particle i , which takes as input circular encodings of the coordinates ν concatenated with T, P and V' . Following [18, 19], C is built as a specified number of transformer blocks each consisting of a multi-headed attention [38] layer followed by an atom-wise MLP. Both layers are implemented as residual updates. The partitioning of ν and μ is changed between layers such that each coordinate gets transformed (see [18, 19] for further details about the architecture).

The invariance of the target density with respect to particle permutation is ensured due to permutation-equivariance of the transformer update [39]. In order to prevent the flow from learning a global shift, we fix one of the atoms and transform the coordinates of the remaining $N - 1$ particles, similar to the approach chosen in [19].

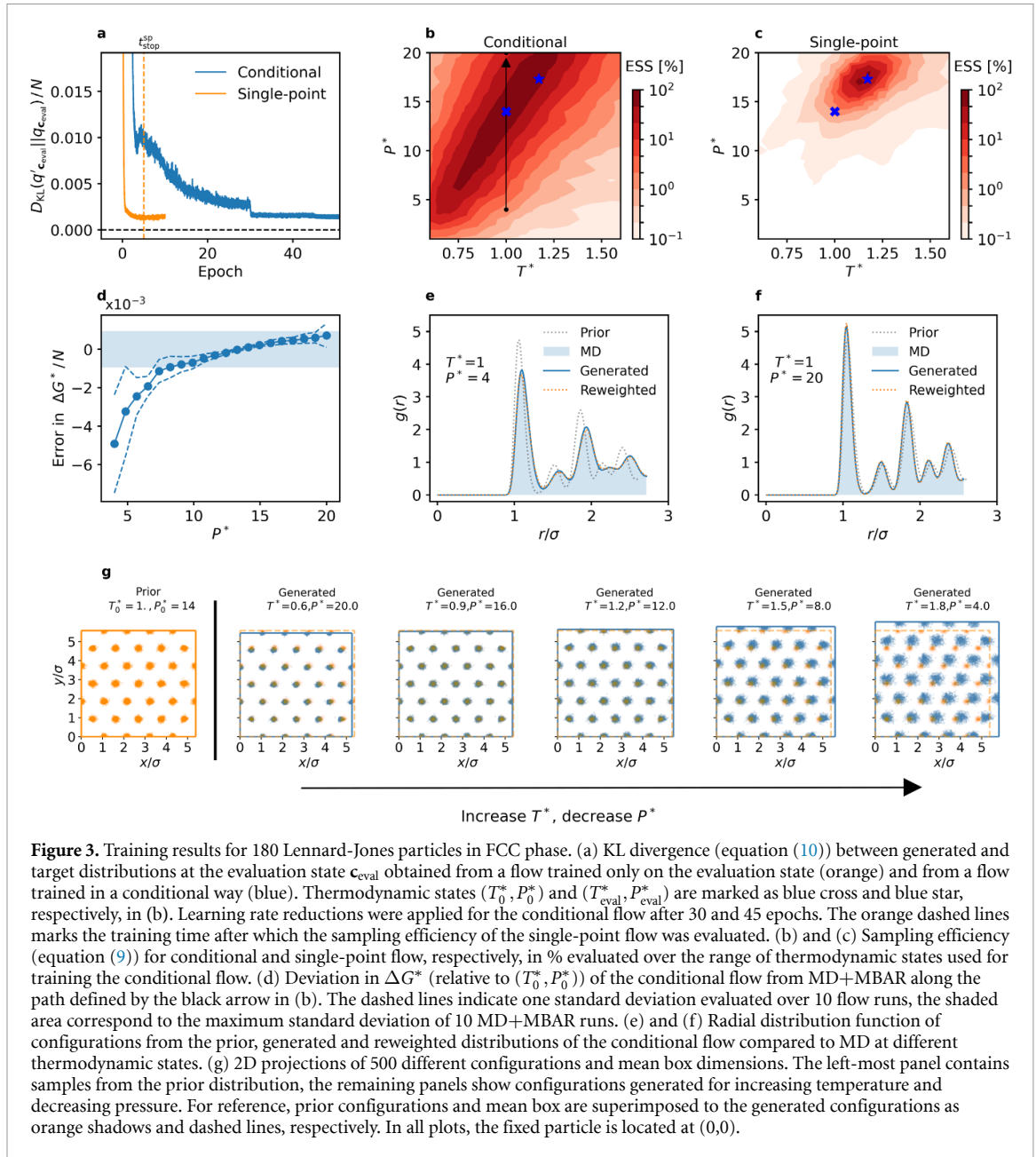
4. Free energy differences and phase diagram

4.1. Training efficiency of conditioned and single-point normalizing flows

The conditioning framework introduced in the previous section has the advantage that samples can be generated over a continuous range of thermodynamic conditions. It requires, however, to train a conditional flow which might converge slower than a single-point flow trained only for one specific temperature and pressure. To compare the convergence of the two flow types, we train both a conditional and single-point flow to transform samples of a 180 particle LJ face-centred cubic (FCC) crystal from a prior thermodynamic state at $\mathbf{c}_0 = (T_0^*, P_0^*)$ with $T_0^* = 1.0, P_0^* = 14.0$ (in reduced LJ units, see Supplementary Information (SI) for computational details). While the single-point flow is trained to generate samples of the equilibrium distribution at an evaluation point $\mathbf{c}_{\text{eval}} = (T_{\text{eval}}^*, P_{\text{eval}}^*)$ with $T_{\text{eval}}^* = 1.2, P_{\text{eval}}^* = 17$, the conditional flow is trained to generate samples for $T^* \in [0.6, 1.6]$ and $P^* \in [1, 20]$. Extensive statistics over the training process is collected by running 10 independent models (see SI for details).

Models were built and trained using JAX [40], Equinox [41], Optax [42], Distrax [42] and jax-md [43]. Molecular dynamics simulations were performed with the OpenMM package [44].

To measure the performances of the two flow types over training time, we evaluate the KL divergence (equation (10), with Δf obtained from standard MD+MBAR simulations) between the generated distribution and the target distribution at the evaluation point (figure 3(a)). The KL divergence of the single-point BG quickly converges within a few epochs, after which an increase in divergence indicates the onset of overfitting. The KL divergence of the conditional BG is much more noisy in the first 20 epochs of



training, with a temporary increase before continuing to decrease. The convergence behavior can be explained by the fact that the conditional BG requires more training steps as it learns a whole family of mappings simultaneously. After applying a learning rate reduction, D_{KL} becomes less noisy and converges to a similar value as the single-point BG. Even though training the conditional BG until convergence requires approximately 10 times more epochs than training the single-point BG, the additional training cost is amortized by the wide range of thermodynamic states covered by the conditional BG.

This becomes evident when comparing the statistical performance of the two flow types over the entire (T, P) -range, which is evaluated using the effective sampling size (ESS) [27] (equation (9)). The ESS over the entire (T, P) -range for the conditional and single-point flows is shown in figures 3(b) and (c), respectively. While the single-point flow reaches high efficiencies close to the evaluation point, the efficiency quickly decays and is close to zero for all other (T, P) -conditions. In contrast, the conditional flow has a very high sampling efficiency of more than 60% over a wide range of thermodynamic states. We found that within LFEP already a few tens of effective samples are sufficient to obtain an accurate estimate of free energy differences. Within the regions of largest efficiencies, highly accurate free energy predictions can already be achieved evaluating only around 100 samples. Similarly, using 2000 test samples, efficiencies as low as 1 % are sufficient to compute free energy differences with high accuracy. Remarkably, similar observations regarding sampling efficiency and free energy differences hold for the liquid phase, although the range of large sampling efficiencies was slightly smaller. For the corresponding efficiency map of the liquid, we refer to the SI.

The sampling efficiency in figure 3(b) is almost constant along a diagonal line in (T, P) crossing the prior state. This may reflect the fact that along this line the flow has to learn only a small transformation, as an increase in temperature leads to an increase in the atomic displacements and volume, whereas an increase in pressure has the opposite effect.

To assess the accuracy of the flow as free energy estimator, using 2000 test samples, we compute the Gibbs free energy difference ΔG^* (equation (22)) with respect to the prior along the path $P^* \in [4, 20]$ for a fixed $T^* = 1.0$ and evaluate the error compared to results obtained with MBAR used across a set of MD simulations (figure 3(d)). Overall, we observe that both methods are in excellent agreement over the entire range. The accuracy only decreases slightly with increasing distance from the prior and decreasing sampling efficiency. Nevertheless, even for the furthest point, $P^* = 4$, the mean deviation per particle is only $5 \cdot 10^{-3}$, while the absolute value in ΔG^* between $P^* = 4$ and $P^* = 20$ is around 14.5 per particle, that is the maximum relative error along the line is less than 0.05%. This demonstrates that the conditional BG can be used for the efficient evaluation of free energy differences also for states far away from the prior.

Furthermore, we illustrate the ability of the model to accurately capture pairwise correlations over varying thermodynamic conditions by computing the radial distribution function (RDF) $g(r)$ [5] for configurations at the start and end point of the path (figures 3(e) and (f)) and compare to MD. For both (T, P) -conditions, the RDFs resulting from samples of the generated distributions are nearly indistinguishable from the MD results, even before reweighting, while samples from the prior distribution clearly yield different RDFs. Any remaining deviations between the mapped and MD results are resolved by reweighting the generated samples (equation (21)). The ability of the conditional BG model to generate samples with varying displacements and volumes is further illustrated in figure 3(g), where 2D-projections of 500 generated configurations with increasing temperature and decreasing pressure are shown.

4.2. Mapping out coexistence lines

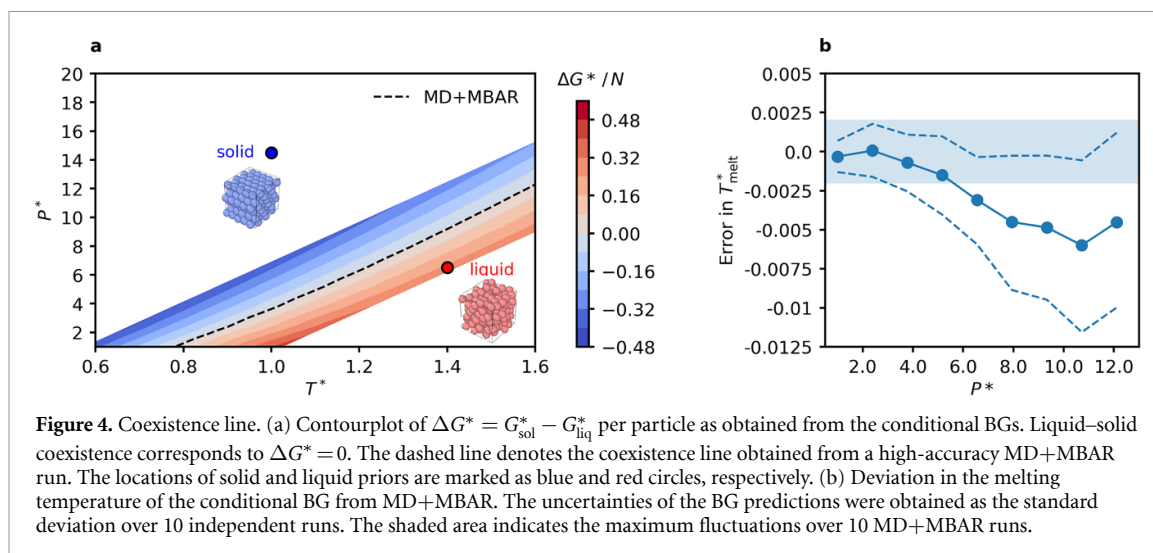
Having demonstrated the efficiency of the conditional BG for accurately computing free energy differences over a wide range of thermodynamic states, we now apply this methodology to the exploration of the phase diagram of the LJ system.

Determining the coexistence line requires to find the thermodynamic states of equal free energy. As discussed in the introduction, this is a tedious and expensive task using traditional free energy estimators as a suitable set of discrete thermodynamic states needs to be identified and full MD/MC simulations, including potentially large equilibration times, need to be executed at each point [45]. In the case of liquid–solid coexistence lines, an additional complication arises from the fact that suitable states, in which both the solid and the liquid phase can be simulated and no phase transition occurs during the simulation time, need to be selected by hand.

Using the LJ system as an example, we demonstrate that our framework based on (T, P) -conditioned flows can alleviate these issues. Specifically, we compute the coexistence line between the solid FCC and liquid phase. The only user-defined input to our workflow consists in choosing appropriate locations of the prior states for the two phases of interest. If a point of coexistence is already known, this point is a good choice as the sampling efficiency is highest close to the prior location (see figure 3) which allows to compute the coexistence line with only a small number of samples. To emphasize that our approach works without explicit knowledge of the coexistence region, we define a broad range of thermodynamic states with $T^* \in [0.6, 1.6]$ and $P^* \in [1, 20]$ in which the coexistence line is expected [46]. Based on physical intuition, we place the prior for the solid at low temperature and high pressure ($T_{\text{sol}}^* = 1, P_{\text{sol}}^* = 14$) and for the liquid at high temperature and low pressure ($T_{\text{liq}}^* = 1.4, P_{\text{liq}}^* = 6$). Conditional flows for both phases are trained over the expected region of coexistence (details are given in the SI). After training, we define a grid of points in the (T^*, P^*) diagram with a spacing of $(\Delta T^*, \Delta P^*) = (0.05, 1.4)$ and use the BGs to evaluate the relative free energies to the respective prior based on 2000 samples.

To obtain the difference in free energy between the two phases, the absolute free energies at the prior locations are required. These are determined by evaluating the free energy difference to analytically or numerically tractable reference systems. For the solid phase, this could also be achieved using normalizing flows, as shown in [21]. Here, we instead chose the more traditional and computationally more efficient approach of MBAR together with MD simulations. For the solid phase, the Einstein crystal [2] serves as a reference, while the liquid phase is connected to the UF model [35] (see section 2.3 for details).

In figure 4(a), the free energy difference between the solid and liquid phase around the coexistence line is shown as a function of temperature and pressure (see SI for free energy differences over the entire (T, P) -range considered during training). The coexistence line predicted by the BG is in excellent agreement with the reference values, obtained from high-accuracy MD+MBAR simulations (dashed line in figure 4(a)), with only small deviations in the high-temperature region. In figure 4(b), the error in the melting temperatures T_{melt}^* as a function of pressure are shown as determined by the BG in comparison to



MD+MBAR. For the low-pressure regions, the melting temperatures predicted by the BG are well within the error bars of the reference calculations. Only for higher pressures, deviations become slightly larger. Overall, the point of equal free energy and therefore the melting temperatures, can be determined with a maximum (mean) deviation accuracy of less than $0.01 T^*$.

4.3. Computational efficiency

A direct comparison of the computational costs for mapping out an entire phase diagram is not entirely straightforward as it depends on a variety of factors. Here, we focus on the number of required energy evaluations which will dominate our flow-based approach as well as traditional free energy estimators for more accurate and expensive interaction models, including ML potentials and *ab initio* approaches.

For MBAR, the number of energy evaluations is determined by the discretization of the (T, P) -range and the number of samples needed for convergence at a given grid density. For the BG, generating a sufficient number of prior samples constitutes the main computational investment. The LJ solid in this study melts when the coexistence line is overstepped by around $0.1 T^*$ or $1 P^*$. Consequently, the grid spacing for MBAR should be at maximum $(\Delta T^*, \Delta P^*) = (0.05 T^*, 0.5 P^*)$ resulting in a total of 800 simulation points per phase. At this grid spacing, at least 100 MD samples per state separated by 500 decorrelation steps need to be collected for MBAR to converge the coexistence line with similar accuracy as the BG, where each MD simulation additionally requires at least 20 000 initial equilibration steps. In total, this amounts to 56 million energy evaluations per phase. For the chosen flow architecture, including more than 20 000 prior samples does not substantially improve the sampling efficiency. Generating these MD samples with the same number of decorrelation steps requires around 10 million energy evaluations, while training for 50 epochs adds less than one million. The final evaluation of 2000 samples on the grid of 800 points adds 1.6 million energy evaluations. Within this rough estimate, the number of required energy evaluation is already reduced by a factor of 5 for the system presented here. In addition, flow-based generation of samples does not require the evaluation of forces which, for expensive potentials, constitutes an additional advantage.

5. Discussion

The introduced BG framework based on conditional normalizing flows provides a novel approach to efficiently calculate differences in free energies over a wide range of thermodynamic states. Contrary to traditional approaches, such as thermodynamic integration or expanded ensembles, it does not require an explicit discretization of the temperature and pressure space. For both solid and liquid phases, flows trained in a conditional fashion achieve large sampling efficiencies even far away from the prior states. As a consequence, only minimal prior knowledge about the system and the melting region is required to map out coexistence lines between two phases with high accuracy. Our workflow, therefore, constitutes a first step towards the efficient and simple calculation of phase diagrams using normalizing flows.

The combination of MD prior data with normalizing flows and the subsequent connection to tractable reference systems allows to compute (T, P) -dependent absolute free energies of the solid phase much more efficiently than purely flow-based approaches [19, 21]. In addition, our approach is equally applicable to evaluate absolute free energies of the liquid phase which is currently out of reach with purely flow-based methods [47]. Nevertheless, the connection to the tractable reference systems remains computationally

expensive. As a next step, a flow-based framework to directly compute free energy differences between phases, ideally utilizing the prior samples of the two phases, would be desirable.

A second direction concerns the system sizes which can be tackled with our approach. In our studies as well as in previous research [48], it was found that the sampling efficiency decreases significantly with increasing system sizes. For the application to realistic systems, thousands of atoms may be required to accurately reproduce experimental results which is unfeasible using current flow-based sampling methodologies. A possible solution could consist in training a transferable model on small systems that can then efficiently be applied to larger ones. With an increase in sampling efficiency, the application of our approach to more accurate potentials as well as the inclusion of molecular crystals [49] becomes feasible and is planned in future work. By advancing the aforementioned fields, we aim to improve the utility and applicability of our approach to enhance the understanding of complex phase behaviour in realistic systems.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/maxschebek/flow_diagrams.

Acknowledgment

M S and J R acknowledge financial support from Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 ‘Scaling Cascades in Complex Systems’, Project Number 235221301, Project B08 ‘Multiscale Boltzmann Generators’. JR acknowledges financial support from DFG through the Heisenberg Programme Project 428315600. MI acknowledges support from the Humboldt Foundation for a Postdoctoral Research Fellowship. We thank Leon Klein for insightful discussions.

ORCID iDs

Maximilian Schebek  <https://orcid.org/0009-0003-4308-3926>

Michele Invernizzi  <https://orcid.org/0000-0002-3328-6557>

Frank Noé  <https://orcid.org/0000-0003-4169-9324>

Jutta Rogal  <https://orcid.org/0000-0002-6268-380X>

References

- [1] Yu Chew P and Reinhardt A 2023 Phase diagrams-Why they matter and how to predict them *J. Chem. Phys.* **158** 030902
- [2] Vega C, Sanz E, Abascal J L F and Noya E G 2008 Determination of phase diagrams via computer simulation: methodology and applications to water, electrolytes and proteins *J. Phys. Condens. Matter* **20** 153101
- [3] Zwanzig R W 1954 High-temperature equation of state by a perturbation method I. Nonpolar gases *J. Chem. Phys.* **22** 1420–6
- [4] Shirts M R and Chodera J D 2008 Statistically optimal analysis of samples from multiple equilibrium states *J. Chem. Phys.* **129** 124105
- [5] Frenkel D and Smit B 2002 Understanding molecular simulation: from algorithms to applications *Computational Science Series* vol 1, 2nd edn (Academic)
- [6] Jarzynski C 2002 Targeted free energy perturbation *Phys. Rev. E* **65** 046122
- [7] Schieber N P and Shirts M R 2019 Configurational mapping significantly increases the efficiency of solid-solid phase coexistence calculations via molecular dynamics: determining the FCC-HCP coexistence line of Lennard-Jones particles *J. Chem. Phys.* **150** 164112
- [8] Schütt K T, Saucedo H E, Kindermans P-J, Tkatchenko A and Müller K-R 2018 SchNet - a deep learning architecture for molecules and materials *J. Chem. Phys.* **148** 241722
- [9] Batzner S, Musaelian A, Sun L, Geiger M, Mailoa J P, Kornbluth M, Molinari N, Smidt T E and Kozinsky B 2022 E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials *Nat. Commun.* **13** 2453
- [10] Zhang L, Han J, Wang H, Car R and Weinan E 2018 Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics *Phys. Rev. Lett.* **120** 143001
- [11] Niu H, Bonati L, Piaggi P M and Parrinello M 2020 Ab initio phase diagram and nucleation of gallium *Nat. Commun.* **11** 2654
- [12] Zhang L, Wang H, Car R and Weinan E 2021 Phase diagram of a deep potential water model *Phys. Rev. Lett.* **126** 236001
- [13] Noé F, Olsson S, Köhler J and Wu H 2019 Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning *Science* **365** eaaw1147
- [14] Tabak E G and Vanden-Eijnden E 2010 Density estimation by dual ascent of the log-likelihood *Commun. Math. Sci.* **8** 217–33
- [15] Jimenez Rezende D and Mohamed S 2015 Variational inference with normalizing flows *Proc. 32nd Int. Conf. on Int. Conf. on Machine Learning - ICML* vol 37 pp 1530–8
- [16] Albergio M S, Kanwar G and Shanahan P E 2019 Flow-based generative models for markov chain monte carlo in lattice field theory *Phys. Rev. D* **100** 034515
- [17] Nicoli K A, Anders C J, Funcke L, Hartung T, Jansen K, Kessel P, Nakajima S and Stornati P 2021 Estimation of thermodynamic observables in lattice field theories with deep generative models *Phys. Rev. Lett.* **126** 032001
- [18] Wirnsberger P, Ballard A J, Papamakarios G, Abercrombie S, Racanière S, Pritzel A, Jimenez Rezende D and Blundell C 2020 Targeted free energy estimation via learned mappings *J. Chem. Phys.* **153** 144112

- [19] Wirnsberger P, Papamakarios G, Ibarz B, Racanière S, Ballard A J, Pritzel A and Blundell C 2022 Normalizing flows for atomic solids *Mach. Learn. Sci. Technol.* **3** 025009
- [20] Ahmad R and Cai W 2022 Free energy calculation of crystalline solids using normalizing flows *Model. Simul. Mat. Sci. Eng.* **30** 065007
- [21] Wirnsberger P, Ibarz B and Papamakarios G 2023 Estimating Gibbs free energies via isobaric-isothermal flows *Mach. Learn. Sci. Technol.* **4** 035039
- [22] van Leeuwen S, de Alba Ortíz A P and Dijkstra M 2023 A Boltzmann generator for the isobaric-isothermal ensemble (arXiv:2305.08483)
- [23] Rizzi A, Carloni P and Parrinello M 2023 Free energies at QM accuracy from force fields via multimap targeted estimation *Proc. Natl Acad. Sci.* **120** e2304308120
- [24] Invernizzi M, Krämer A, Clementi C and Noé F 2022 Skipping the replica exchange ladder with normalizing flows *J. Phys. Chem. Lett.* **13** 11643–9
- [25] Mey A S J S et al 2020 Best practices for alchemical free energy calculations *Living J. Comput. Mol. Sci.* **2** 18378
- [26] Olehnovics E, Michelle Liu Y, Mehio N, Sheikh A Y, Shirts M and Salvalaglio M 2024 Assessing the accuracy and efficiency of free energy differences obtained from reweighted flow-based probabilistic generative models *J. Chem. Theory Comput.* **20** 5913–22
- [27] Kish L 1965 Sampling organizations and groups of unequal sizes *Am. Sociol. Rev.* **30** 564–72
- [28] Papamakarios G, Nalisnick E, Jimenez Rezende D, Mohamed S and Lakshminarayanan B 2021 Normalizing flows for probabilistic modeling and inference *J. Mach. Learn. Res.* **22** 1–64
- [29] Dinh L, Krueger D and Bengio Y 2014 Nice: Non-linear independent components estimation (arXiv:1410.8516)
- [30] Dinh L, Sohl-Dickstein J and Bengio S 2017 Density estimation using Real NVP *Int. Conf. on Learning Representations* (available at: <https://openreview.net/forum?id=HkpbmH9lx>)
- [31] Ardizzone L, Lüth C, Kruse J, Rother C and Köthe U 2019 Guided image generation with conditional invertible neural networks (arXiv:1907.02392)
- [32] Winkler C, Worrall D, Hooeboom E and Welling M 2019 Learning Likelihoods with conditional normalizing flows (arXiv:1912.00042)
- [33] Falkner S, Coretti A, Romano S, Geissler P L and Dellago C 2023 Conditioning Boltzmann generators for rare event sampling *Mach. Learn. Sci. Technol.* **4** 035050
- [34] Gerdes M, de Haan P, Rainone C, Bondesan R and Cheng M C N 2023 Learning lattice quantum field theories with equivariant continuous flows *SciPost Phys.* **15** 238
- [35] Paula Leite R, Freitas R, Azevedo R and de Koning M 2016 The Uhlenbeck-Ford model: Exact virial coefficients and application as a reference system in fluid-phase free-energy calculations *J. Chem. Phys.* **145** 194101
- [36] Dibak M, Klein L, Krämer A and Noé F 2021 Temperature steerable Flows and Boltzmann generators *Phys. Rev. Res.* **4** L042005
- [37] Jimenez Rezende D, Papamakarios G, Racaniere S, Albergo M, Kanwar G, Shanahan P and Cranmer K 2020 Normalizing flows on tori and spheres *Proc. 37th Int. Conf. on Machine Learning* vol 119 pp 8083–92
- [38] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Lukasz and Polosukhin I 2017 Attention is all you need *Proc. 31st Int. Conf. on Neural Information Processing Systems* pp 6000–10
- [39] Köhler J, Klein L and Noé F 2020 Equivariant flows: Exact likelihood generative learning for symmetric densities *Proc. 37th Int. Conf. on Machine Learning* vol 119 pp 5361–70
- [40] Bradbury J et al 2018 JAX: composable transformations of Python+NumPy programs (available at: <http://github.com/jax-ml/jax>)
- [41] Kidger P and Garcia C 2021 Equinox: neural networks in JAX via callable PyTrees and filtered transformations *Differentiable Programming Workshop at Neural Information Processing Systems 2021* (arXiv:2111.00254)
- [42] Babuschkin I et al 2020 The DeepMind JAX Ecosystem (available at: <http://github.com/google-deeppmind>)
- [43] Schoenholz S S and Cubuk E D 2020 JAX M.D. a framework for differentiable physics *Advances in Neural Information Processing Systems* vol 33 (available at: <https://papers.nips.cc/paper/2020/file/83d3d4b6c9579515e1679aca8cbc8033-Paper.pdf>)
- [44] Eastman P et al 2017 Openmm 7: rapid development of high performance algorithms for molecular dynamics *PLOS Comput. Biol.* **13** e1005659
- [45] Chodera J D 2016 A simple method for automated equilibration detection in molecular simulations *J. Chem. Theory Comput.* **12** 1799–805
- [46] Bizjak A, Urbič T and Vlachy V 2009 Phase diagram of the lennard-jones system of particles from the cell model and thermodynamic perturbation theory *Acta Chim. Slov.* **56** 166–71
- [47] Coretti A, Falkner S, Geissler P and Dellago C 2022 Learning mappings between equilibrium states of liquid systems using normalizing flows (arXiv:2208.10420)
- [48] Abbott R et al 2023 Aspects of scaling and scalability for flow-based sampling of lattice QCD *Eur. Phys. J. A* **59** 257
- [49] Köhler J, Invernizzi M, De Haan P and Noé F 2023 Rigid body flows for sampling molecular crystal structures *Proc. 40th Int. Conf. on Machine Learning* vol 202 pp 17301–26