

Accurate Lattice Free Energies of Packing Polymorphs from Probabilistic Generative Models

Published as part of *Journal of Chemical Theory and Computation* special issue "Machine Learning and Statistical Mechanics: Shared Synergies for Next Generation of Chemical Theory and Computation".

Edgar Olehnovics, Yifei Michelle Liu, Nada Mehio, Ahmad Y. Sheikh, Michael R. Shirts, and Matteo Salvalaglio*



Cite This: *J. Chem. Theory Comput.* 2025, 21, 2244–2255



Read Online

ACCESS |



Metrics & More

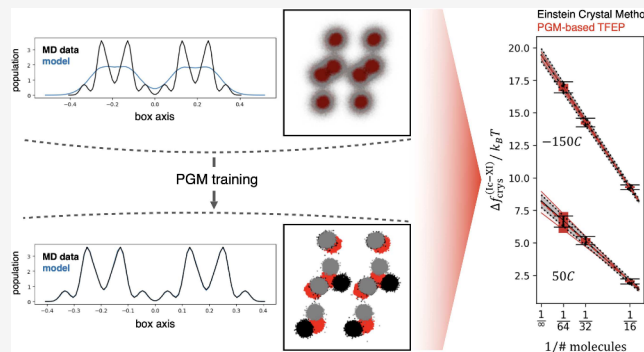


Article Recommendations



Supporting Information

ABSTRACT: Finite-temperature lattice free energy differences between polymorphs of molecular crystals are fundamental to understanding and predicting the relative stability relationships underpinning polymorphism, yet are computationally expensive to obtain. Here, we implement and critically assess machine-learning-enabled targeted free energy calculations derived from flow-based generative models to compute the free energy difference between two ice crystal polymorphs (Ice XI and Ic), modeled with a fully flexible empirical classical force field. We demonstrate that even when remapping from an analytical reference distribution, such methods enable a cost-effective and accurate calculation of free energy differences between disconnected metastable ensembles when trained on locally ergodic data sampled exclusively from the ensembles of interest. Unlike classical free energy perturbation methods, such as the Einstein crystal method, the targeted approach analyzed in this work requires no additional sampling of intermediate perturbed Hamiltonians, offering significant computational savings. To systematically assess the accuracy of the method, we monitored the convergence of free energy estimates during training by implementing an overfitting-aware weighted averaging strategy. By comparing our results with ground-truth free energy differences computed with the Einstein crystal method, we assess the accuracy and efficiency of two different model architectures, employing two different representations of the supercell degrees of freedom (Cartesian vs quaternion-based). We conduct our assessment by comparing free energy differences between crystal supercells of different sizes and temperatures and assessing the accuracy in extrapolating lattice free energies to the thermodynamic limit. While at low temperatures and in small system sizes, the models perform with similar accuracy. We note that for larger systems and high temperatures, the choice of representation is key to obtaining generalizable results of quality comparable to that obtained from the Einstein crystal method. We believe this work to be a stepping stone toward efficient free energy calculations in larger, more complex molecular crystals.



1. INTRODUCTION

Free energy estimates are key to quantitatively understanding physical-chemical processes ranging from solubility to binding affinity.^{1–4} In the case of polymorphism of molecular crystals in particular, accurate and efficient lattice free energies are needed to determine the relative stability of crystal packings, which is important in many industries, including chemicals, pharmaceuticals, semiconductors, and food products.^{5–9} An accurate prediction of the relative thermodynamic stability of polymorphs at different temperatures requires calculating the entropic contribution to free energy. Including such thermodynamic detail is often prohibitively expensive due to the anharmonic nature and the presence of low-frequency modes in the vibrational free energy.^{8,10–15} For free energy calculations to be of industrial relevance, free energy methods are needed

that are suitably cheap and scalable for deployment on larger data sets of putative finite temperature polymorphs while retaining accuracy levels on par with states of the art classical anharmonic free energy methods.^{8,12,13} Given an atomistic model of the system of interest and a putative crystal structure, its finite-temperature lattice free energy can be rigorously computed using statistical mechanics computational techniques, often based on the principles of free energy perturbation and

Received: November 26, 2024

Revised: January 31, 2025

Accepted: February 7, 2025

Published: February 21, 2025



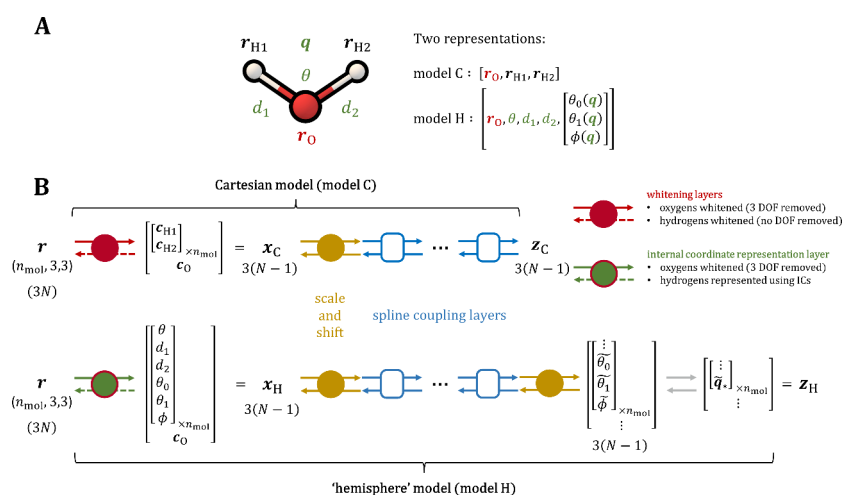


Figure 1. (A) Illustration of a water molecule with two representations (model C and model H) of the 9 DOFs. (B) Illustration of the two model architectures compared in this work (model C; top and model H; bottom). Representations: Both models start by transforming the Cartesian coordinates of the system r , containing N atoms, as discussed in the main text, to create a vector x containing $3(N - 1)$ marginal variables. These variables are then rescaled to a fixed model range (e.g., $[-1, 1]$ or $[-1, 1]$, depending on the type of variable). Mapping: Both models used one-dimensional monotonically increasing rational quadratic splines⁴⁰ (with five trainable knots) to transform x in an elementwise manner according to coupled flow architecture.⁴¹ This was implemented as in ref.³⁹ In each system, 4 coupling layers were used in the case of the Hemisphere model, and a maximum number of coupling layers was used in the Cartesian model. The partitioning of variables in each layer was based on the heuristic discussed in SI Figure S1.⁴² The coupling layers invertibly transform x to z . The maps were trained to approach a uniform remapped distribution in z . For this reason, in model H, the variables associated with quaternions are transformed back to 4D unit vectors on the specified hemisphere. The details are discussed in the main text. The amount of data used for training and validating the PGMs was 200,000 samples (extracted from a 20 ns trajectory) in each system containing 16 or 32 molecules and 400,000 samples (extracted from a 40 ns trajectory) in each system with 64 molecules. A training:validation split of 3:1 was used during each training run.

thermodynamic integration.^{16,17} Several approaches have been proposed in the literature, including various implementations of the Einstein crystal method,^{17–19} diabatic approaches drawing inspiration from lattice-switch Monte Carlo and Marcus theory^{20–22} and enhanced sampling methods based on introducing biasing potentials,^{23–25} to name a few. While these approaches are theoretically well established and formally yield fully anharmonic lattice-free energies, their computational cost is frequently too demanding for large-scale deployment. For this reason, the current state of the field relies on the ad hoc application of exact methods such as the Einstein crystal and the more widespread use of methods based on the harmonic approximation, which are nowadays routinely deployed to carry out a final refinement of zero-Kelvin energy landscapes.^{6,12}

One of the main limitations of formally exact methods based on statistical mechanics is that they typically require a significant amount of computing effort to sample intermediate, unphysical states, which increases the computational cost without providing inherent physical insights.^{16,17} Taking the Einstein crystal method as the “gold standard” example of a physics-based, formally exact approach for calculating lattice-free energies, we note that its convergence requires devising a chain of states with overlapping configurational distributions that connect the Hamiltonian describing the physical crystal polymorph to the Einstein crystal Hamiltonian. To obtain convergence, neighboring Boltzmann distributions along such a chain of states must sample overlapping configurations, and in each state, configurations must be sampled ergodically. Thus, sampling sets of tens to hundreds of perturbed Hamiltonians are typically required to achieve well-converged free energy estimates for a single putative crystal form.

The emergence of machine learning methods capable of training deterministic bijective coordinate transformations that directly map between the probability distributions of two

arbitrarily different metastable states of interest, in this case, the polymorphic cell of interest and the Einstein crystal reference state offers a promising alternative to the costly direct sampling of intermediate, unphysical states.^{26–31} We have recently shown that such approaches, coupled with BAR³² and MBAR³³ reweighting techniques for free energy estimation can yield very accurate free energy differences between conformational states of isolated molecules.^{34–39}

Here, we take stock of the lessons learned in ref.³⁹ to critically compare two different methods of representing the configurations of flexible molecules in a crystal for training invertible flow-based probabilistic generative models (PGMs) that can effectively map between two metastable water ice crystal structures (Ice XI and Ic), modeled with an empirical classical force field. In particular, we assess the efficiency and accuracy of the lattice-free energy estimates obtained by ML-enabled targeted free energy estimation against ground truth estimates obtained with an Einstein Crystal method. We examine the effect of different temperatures and system sizes on the convergence of free energy estimates and the quality of extrapolating these results to the thermodynamic limit. We note that the two temperatures examined (-150 and 50 °C) are meant to stress test the methods. For instance, at higher temperatures, we can expect the configurational distributions to be significantly less localized and more anharmonic, making them more challenging to model using neural networks.

II. PROBABILISTIC GENERATIVE MODELS FOR LATTICE FREE ENERGY CALCULATION

To quantitatively estimate lattice free energies, we build and expand on our earlier work,³⁹ where we identified several useful heuristics for applying reweighted PGMs to obtain accurate free energy estimates. The only MD data that is required for training

and reweighting the PGMs is that which is sampled in the physical macrostates of interest. This approach, therefore, entirely avoids sampling multiple intermediate Hamiltonians (details provided in the *Simulations Setup* section), typical of free energy perturbation methods such as the Einstein Crystal.^{39,43} In ref.³⁹ we used isolated small molecules in a vacuum as representative case studies, with $3N - 9$ out of $3N - 6$ of the relevant intramolecular degrees of freedom (DOFs) represented using internal coordinates.⁴⁴ Those PGMs relied on a *mixed coordinate* representation of the given molecule.³⁴ Specifically, we kept three of the atoms represented in Cartesian coordinates to serve as a starting point for a bijective reconstruction of all remaining atoms of the molecule from the internal coordinates. We referred to these three Cartesian atoms as the *Cartesian block*. When dealing with isolated molecules in a vacuum, the Cartesian block only has three relevant DOFs. However, in systems containing more than one molecule, like crystal supercells, all nine DOFs of the Cartesian block of each molecule must be jointly modeled by the PGM.

Here, we use flow-based PGMs to compute the Helmholtz FE differences between crystal supercells using two ice polymorphs (Ic and XI) as simple but representative examples of systems requiring the explicit representation of all nine DOFs for all Cartesian blocks. For such systems, it is important to consider the most cost-effective way of representing a Cartesian block inside the PGM. In this regard, we highlight ref.³⁰ which has contributed a powerful method of representing the Cartesian block based on a unit-quaternion to describe the rotational state of a given molecule. In particular, they have demonstrated that building a normalizing flow model with this representation allows learning an accurate approximation of the Boltzmann distribution over DOFs belonging to atoms that are not covalently bonded and, thus, obtaining FE estimates of ice XI crystal supercells modeled with rigid water molecules.³⁰ In the following, we compare two PGM architectures, respectively implementing a quaternions-based representation inspired by ref.³⁰ and a direct Cartesian representation. These two models only differ in how the hydrogen atoms in each water molecule are represented. The first representation is based directly on their Cartesian coordinates (model C, illustrated in Figure 1A), while the second relies on unit-quaternions and internal coordinates (model H, illustrated in Figure 1B). Details of the models are discussed in sections IIA and IIB, respectively. To assess their accuracy, efficiency, and ability to deal with an increasing number of DOFs, we extended our comparison to three system sizes and two temperatures.

To maintain consistency in comparing free energy differences computed from the two models and the ground truth, we established the following ground rules:

- Both types of models were trained and evaluated on the same data, following the same protocol. For instance, the training validation split was fixed to 3:1 in all cases. This MD data was taken from $\lambda = 1$ simulations defined in section IIH to establish a fair comparison with the Einstein crystal method. Each model was trained and evaluated on batches of 1,000 and 10,000 configurations, respectively.
- The three DOFs defining the position of the center of mass were removed from both types of flows identically by *whitening* the Cartesian coordinates of all oxygen atoms, as discussed in section IIC.

- Each spline coupling layer (represented in blue in Figure 1) used a specific nonrandom conditioning protocol (displayed in Figure S1). The number of trainable parameters was adjusted by varying the number of coupling layers to minimize the variance of FE estimates obtained using the BAR_V estimator (defined in section IID) evaluated on the models during training. Models were trained for as long as necessary to observe a convergent behavior in the running average of these estimates, as defined in section IIE.

Moreover, we note that it was possible to construct the trainable part of both models using standard spline coupling layers, analogously to our previous work in isolated molecules.³⁹ As such, both models (C and H) used standard coupling layers³⁹ that are not inherently equivariant to permutations of any DOFs. It was possible to use this approach because none of the water molecules in the supercells we compared swapped lattice sites within the MD-accessible time scales. We note that PGM architectures can be equipped with equivariant functionality by replacing the standard multilayered perceptrons with attention-based coupling mechanisms.^{26–28,30,31} The following sections describe in detail the approach adopted in models C and H, respectively.

A. Cartesian Representation of Hydrogen Atoms. In the Cartesian model (model C in Figure 1B), the positions of the hydrogen atoms in the whole system $[r_{H1}, r_{H2}]$ were represented directly with their Cartesian coordinates. Meanwhile, the oxygen atoms in the system (r_O) were instead represented using whitened Cartesian coordinates, as described in section IIC. Taken together, the variable x_C shown in Figure 1, describes all the relevant $N(N - 1)$ marginal DOFs of the ice crystal. Finally, the shift and scale layer (yellow in Figure 1) was initialized to fit the one-dimensional marginal variables of x_C into the working interval of the 1D splines. This step takes advantage of the minimum and maximum values explored by each 1D marginal variable in the MD training and validation data to define the one-dimensional linear transformations to carry out the scale and shift, exactly like in ref.³⁹ Likewise, the spline coupling layers in this work were also implemented according to methods detailed in ref.³⁹ with each marginal variable of model C assigned to a nonperiodic $[-1, 1]$ interval. The uniform log base distribution ($\ln p_0$) in model C was therefore set to³⁹

$$\ln p_0 = -(9n_{\text{mol}} - 3)\ln(2) = -3(N - 1)\ln(2) \quad (1)$$

B. Hemisphere Representation of Hydrogen Atoms. In the hemisphere model (model H in Figure 1B), the positions of the hydrogen atoms were described with the help of unit quaternions. The following paragraphs describe how this representation was implemented.

Reference³⁰ shows that Cartesian coordinates of the two hydrogen atoms (r_{H1} and r_{H2}) of a water molecule $r_{\text{mol}} = [r_O, r_{H1}, r_{H2}]$ can be efficiently represented by a set of independent internal coordinates (x_{mol}), consisting of two bond lengths (d_1, d_2), one bond angle (θ), and one (flip invariant) unit-quaternion ($\pm q \in S^3 \subset \mathbb{R}^4$), as well as r_O .^{30,45} For brevity, only the inverse of this transformation is discussed here. The configuration of a water molecule can be reconstructed from these coordinates using the following equation, where we adopt the approach detailed in ref.⁴⁶ to transform a unit-quaternion into a rotation matrix $R(q)$:^{30,45,46}

$$\mathbf{r}_{\text{mol}} = \mathbf{r}_O + \mathbf{R}(\mathbf{q}) \begin{bmatrix} 0 & 0 & d_2 \sin \theta \\ 0 & 0 & 0 \\ 0 & d_1 & d_2 \cos \theta \end{bmatrix} \quad (2)$$

In relation to eq 2, refs^{30,45} provide the following closed-form expression of the log-volume change:

$$\ln \gamma_{\mathbf{x}_{\text{mol}} \rightarrow \mathbf{r}_{\text{mol}}} = \ln(8d_1^2 d_2^2 \sin \theta) + \frac{1}{2} \ln(4(d_1^2 + d_2^2) + 1) \quad (3)$$

We observe that eq 3 gives the same quantities as a $\frac{1}{2} \sum_{i=1}^9 \ln \lambda_i$, with eigenvalues $\lambda_1, \dots, \lambda_9$ obtained from diagonalizing $\mathbf{J}^T \mathbf{J}$, where $\mathbf{J} \in \mathbb{R}^{9 \times 10}$ is the Jacobian of eq 2 given by automatic differentiation. eq 3 was therefore evaluated when reconstructing a water molecule (i.e., mapping $\mathbf{x}_{\text{mol}} \rightarrow \mathbf{r}_{\text{mol}}$). When mapping in the opposite direction (i.e., $\mathbf{r}_{\text{mol}} \rightarrow \mathbf{x}_{\text{mol}}$), as defined in ref³⁰ the log volume change of $-\ln \gamma_{\mathbf{x}_{\text{mol}} \rightarrow \mathbf{r}_{\text{mol}}}$ was used.^{30,45}

In model H, all of the intramolecular DOFs (i.e., the bond lengths and angles $[d_1, d_2, \theta]_{\mathbf{x}_{\text{mol}}}$, see Figure 1) were directly concatenated to the whitened Cartesian coordinates of the oxygen atoms ($\mathbf{c}_O$), with the latter being treated in the same way as in model C.

The remaining $3n_{\text{mol}}$ DOFs of the system, corresponding to the n_{mol} unit-quaternions (specifying the rotations of the molecule in a supercell relative to a fixed reference frame ensured by the nonrotating supercell vectors), were treated as discussed below.

Each unit-quaternion, representing the rotation of a given molecule in an arbitrary crystal data set, may exist anywhere on the S^3 hyper-sphere. Thus, its four marginal coordinates $[q_0, q_1, q_2, q_3]$ are coupled by the unit-length constraint. Moreover, antipodal quaternions represent equivalent rotation matrices (i.e., $\mathbf{R}(\mathbf{q}) = \mathbf{R}(-\mathbf{q})$). This means that any hemisphere of S^3 is sufficient to describe all possible rotations, or equivalently, only half of S^3 is sufficient to specify a set of all unique quaternion-based representations of all 3D rotations.^{47,48} When considering the curvature of S^3 , or its hemisphere, it is clear that an explicit global parametrization of this space is impossible without encountering singularities. Instead, the recommended singularity-free approach is to work directly with unit vectors embedded in $\mathbb{R}^4$ (implicit representation).⁴⁶ Indeed, this is exactly how ref³⁰ approached the problem of mapping between distributions of rotations. Specifically, two novel flip-equivariant bijectors (*symmetrized Moebius transformations* and *symmetrized projective convex gradient maps*) were developed in ref³⁰ that enable modeling smooth flip-symmetric flows on the surface of any flip-symmetric sphere. In the case of S^3 , these bijectors serve as robust flow architectures for modeling distributions on the rotation manifold (SO3).³⁰ Alternative ways of learning normalized distributions of rotations that do not rely on flows can also be found in the literature and retain the unit-vector representation.^{49,50}

All quaternions are assigned to a specific hemisphere and then globally parametrized using standard hyper-spherical coordinates despite the presence of singularities. This choice was implemented to provide model H with sufficient representational power while minimizing computational costs.

The hemisphere was specified by $\mathbf{q}_* = \text{sign}(q_0)\mathbf{q}$, where $\text{sign}(x) = \frac{x}{|x|}$, and the associated maps are

$$\mathbf{s} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \phi \end{bmatrix} = \begin{bmatrix} \text{atan } 2(\sqrt{q_1^2 + q_2^2 + q_3^2}, q_0) \\ \text{atan } 2(\sqrt{q_2^2 + q_3^2}, q_1) \\ \text{atan } 2(q_3, q_2) \end{bmatrix}$$

$$\mathbf{q}_* = \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} \cos(\theta_0) \\ \sin(\theta_0)\cos(\theta_1) \\ \sin(\theta_0)\sin(\theta_1)\cos(\phi) \\ \sin(\theta_0)\sin(\theta_1)\sin(\phi) \end{bmatrix}$$

; where $\theta_0 \in [0, \frac{\pi}{2}]$, $\theta_1 \in [0, \pi]$, and $\phi \in [-\pi, \pi]$ (4)

The log-volume change of eq 4, in the $\mathbf{s} \rightarrow \mathbf{q}_*$ direction is $\ln(\sin^2(\theta_0) \sin(\theta_1))$. When transforming in the opposite direction, the same expression is multiplied by -1 . The singularities in this quantity are considered at points where $\sin(x) = 0$ for any x , where the volume, and thus density, cannot be defined. For numerical stability of dealing with samples flowing near these points, the arguments of the log were *clipped* to be greater than a small but finite threshold ($1\text{e-}8$). Generally, clipping of this type can introduce errors during inference, which in the flow-based PGMs can be quantified by a mismatch in log volume changes associated with the forward and inverse transformations of the data. For this reason, we systematically checked the inversion accuracy of model H in both directions, before and after training, to ensure that the approximation introduced in the chosen representation is within reasonable tolerance levels. An example of this analysis is discussed in SI Figure S7. One can assume that the small patches of configurational space surrounding the singularities in each molecular configuration are negligible compared to the total volume of configurational space. Thus, the errors associated with samples flowing near these regions may, on average, yield a negligible contribution during reweighting.

Furthermore, since $\mathbf{s}$ does not preserve the antipodal flip-symmetry at the edge of the hemisphere, we note that in the systems we examined so far it was possible to rotate the global reference frame of the supercell data such that, for each molecule in the data set, the marginal distribution in θ_0 is shifted away from both 0 (one of the singularities) and $\frac{\pi}{2}$ (edge of the hemisphere). More generally, one can also note that in a stable crystal system where none of the molecules samples significantly larger rotational angles, and thus their quaternion distribution on S^3 is quite narrow, one can define an arbitrary per-molecule rotation (i.e., a translation of the MD data set on S^3) such that each molecule maps to a formally well behaved joint distribution in the hyper-spherical coordinates ($\mathbf{s}$). In this regime, the three marginal variables of $\mathbf{s}$ can be transformed and treated independently.

That being said, we intentionally did not limit the model H from being capable of generating all possible rotations, even though this constraint can be trivially imposed on the marginals of $\mathbf{s}$, as part of scaling and shifting layer.

After each unit-quaternion is transformed to $\mathbf{s}$, the combined set of $3n_{\text{mol}}$ rotational DOFs of the crystal were concatenated to the rest of the marginal variables mentioned earlier to give $\mathbf{x}_H$ shown in Figure 1B.

Akin to model C, the coupling layers in a model H were trained to flatten the configurational distribution of the MD data, to map it as close as possible to a uniform distribution over z_H (Figure 1). In terms of achieving a uniform distribution of rotations, this corresponds to uniformly sampling the surface of the hemisphere of S^3 . For this reason model H transforms s back to q^* on both sides of the flow, including the *noisy* side (Figure 1). The log base distribution $\ln p_0$ in model H was therefore set to

$$\ln p_0 = -2n_{\text{mol}} \ln \pi - (6n_{\text{mol}} - 3) \ln 2 \quad (5)$$

The first term in eq 5 originates from the fact that a hemisphere of S^3 has a surface area of π^2 . The second term takes into account the 6 remaining nonrotational DOFs in each molecule, with 3 global center-of-mass degrees of freedom of the supercell missing from the flow.

C. Virtual Atoms and Center of Mass Treatment. The classical force field used to model water molecules in the polymorphic forms of ice investigated was TIP4P/Ice: a four-body potential that, alongside the oxygen and two hydrogen atoms, employs an additional virtual atom (r_M). The dynamics of the virtual atom do not contribute any additional DOF to the system because its position can be reconstructed without uncertainty as

$$r_M = r_O + 0.13458335 \times (r_{H1} + r_{H2} - 2r_O) \quad (6)$$

where $r_{i \neq M}$ represents the position of atom i . As such, the virtual atoms were removed before feeding molecular coordinates to any PGM model. When sampling the supercell configurations of ice polymorphs using the model, virtual atoms were reconstructed using eq 6.

The center of mass (COM) of the oxygen atoms was removed from all the MD-generated configurations used in the training or validation of the PGM models as

$$r := r - \langle r_O \rangle_{\text{supercell}} \quad (7)$$

After applying eq 7, in which the periodic boundary conditions (PBCs) of the supercell box are not taken into account, the Cartesian coordinates of each supercell (r) are guaranteed to exist on a $3(N - 1)$ dimensional hyper-plane specified by the fixed center of mass. In other words, eq 7 always creates three singular DOFs in r . These three redundant dimensions were reliably removed from the flow using a PCA whitening layer, initialized on Cartesian coordinates of the oxygen atoms from the entire MD data set (after applying eq 7). The whitening transformation was adopted following the approach of ref.³⁴ The log volume changes corresponding to the whitening and unwhitening transformations are $-\frac{1}{2} \sum_{j=1}^{3(n_{\text{mol}}-1)} \ln \lambda_j$ and $\frac{1}{2} \sum_{j=1}^{3(n_{\text{mol}}-1)} \ln \lambda_j$, respectively, where n_{mol} is the number of molecules, and $\lambda_j > 0$ are the eigenvalues of the covariance matrix initialized on the entire MD data set, as detailed in the SI.³⁴

D. Reweighting the Learned Distribution and Computing Lattice Free Energies. Using the relevant log-base distributions (eqs 1 and 5, in models C and H respectively), the normalized log-probability $\ln(q(r))$ of each PGM model (q) was defined in the following way:³⁹

$$\ln(q(r)) = \ln \gamma_{r \rightarrow z}(r) + \ln(p_0) \quad (8)$$

In eq 8, $\ln \gamma_{r \rightarrow z}(r)$ represents the log volume change associated with transforming r to z , as illustrated in Figure 1.³⁹

Defining the generalized work function $\phi = \beta U + \ln(q)$, one can use the two-state BAR equation (extensively discussed in ref.³⁹) to estimate *absolute* FE of each macrostate corresponding to a polymorph. Here, one of the two data sets fed into the BAR algorithm is sampled from the PGM mode. At the same time, the other originates from the MD data that samples the equilibrium distribution of a given macrostate. Ergodic samples from the model distribution q were obtained by sampling the (uniform) base distribution p_0 and then transforming them via the learned map. Ensemble averages over MD data ($\langle \bullet \rangle_p$) were instead evaluated on locally ergodic configurations, exclusively extracted from the validation data set (i.e., data not included in the PGM training set).³⁹ As such, we refer to this estimator as BAR_V and indicate the absolute free energy estimates in reduced units obtained from this estimator as $f_{\text{crys}}^{\text{BAR}_V}$, where the subscript *crys* indicates the polymorph.

In each system, $f_{\text{crys}}^{\text{BAR}_V}$ was evaluated on random batches of 10^4 validation configurations during model training, with a stride of 25 training batches between evaluations. This reweighing approach is both computationally cheap and accurate, an observation that aligns well with results in prior literature.^{35,36,39} Furthermore, the use of the *pymbar* library to evaluate the BAR free energy estimator provides analytical error bars which quantify the uncertainty of $f_{\text{crys}}^{\text{BAR}_V}$ estimates.⁵¹

E. Averaging Free Energies from Multiple Maps during Training. When training each PGM by maximum likelihood, BAR_V was evaluated multiple times during training, resulting in a set of M different *raw* estimates of crystal free energy ($f_{\text{crys}}^{\text{BAR}_V}(i); i = 1, \dots, M$), that are stochastically distributed around an average, most representative, estimate.

Here, we note that the most accurate of these estimates, i.e., the one associated with lower variance during training by maximum likelihood, coincides with the minimum validation error.³⁶ Employing the same notation of ref.³⁹ we can label such error when evaluated on the same validation batches as the free energy estimate $-f_{\text{crys}}^{\text{AVMD}_V}(i)$.³⁴

We also note that $f_{\text{crys}}^{\text{AVMD}_V}(i)$ increases at the start of training toward a maximum and then decreases gradually with a rate that correlates with overfitting, which our previous work has shown to coincide with increasing variance in the $f_{\text{crys}}^{\text{BAR}_V}(i)$ estimates.³⁹

As such, inspired by ref.³⁸ we adopt the following unsupervised way of averaging multiple $f_{\text{crys}}^{\text{BAR}_V}(i)$ estimates obtained from multiple maps during training:

$$\bar{f}_{\text{crys}} = \left[\sum_{j=1}^M w_j \right]^{-1} \sum_{i=1}^M w_j f_{\text{crys}}^{\text{BAR}_V}(i) \quad (9)$$

where weights are proportional to $w_i = \exp(f_{\text{crys}}^{\text{AVMD}_V}(i))$. This approach enables to obtain running weighted averages $\bar{f}_{\text{crys}}$ that are minimally affected by the large variances seen at the start of training or, in later stages, due to overfitting. Evaluating eq 9 during training yields a converging curve of final FE values (see Figures 3, 4) that in practice takes into account estimates from all the previous versions of the trained maps, and all of the random validation batches encountered. This approach practically extends the *multimap* method described in ref.³⁸ by

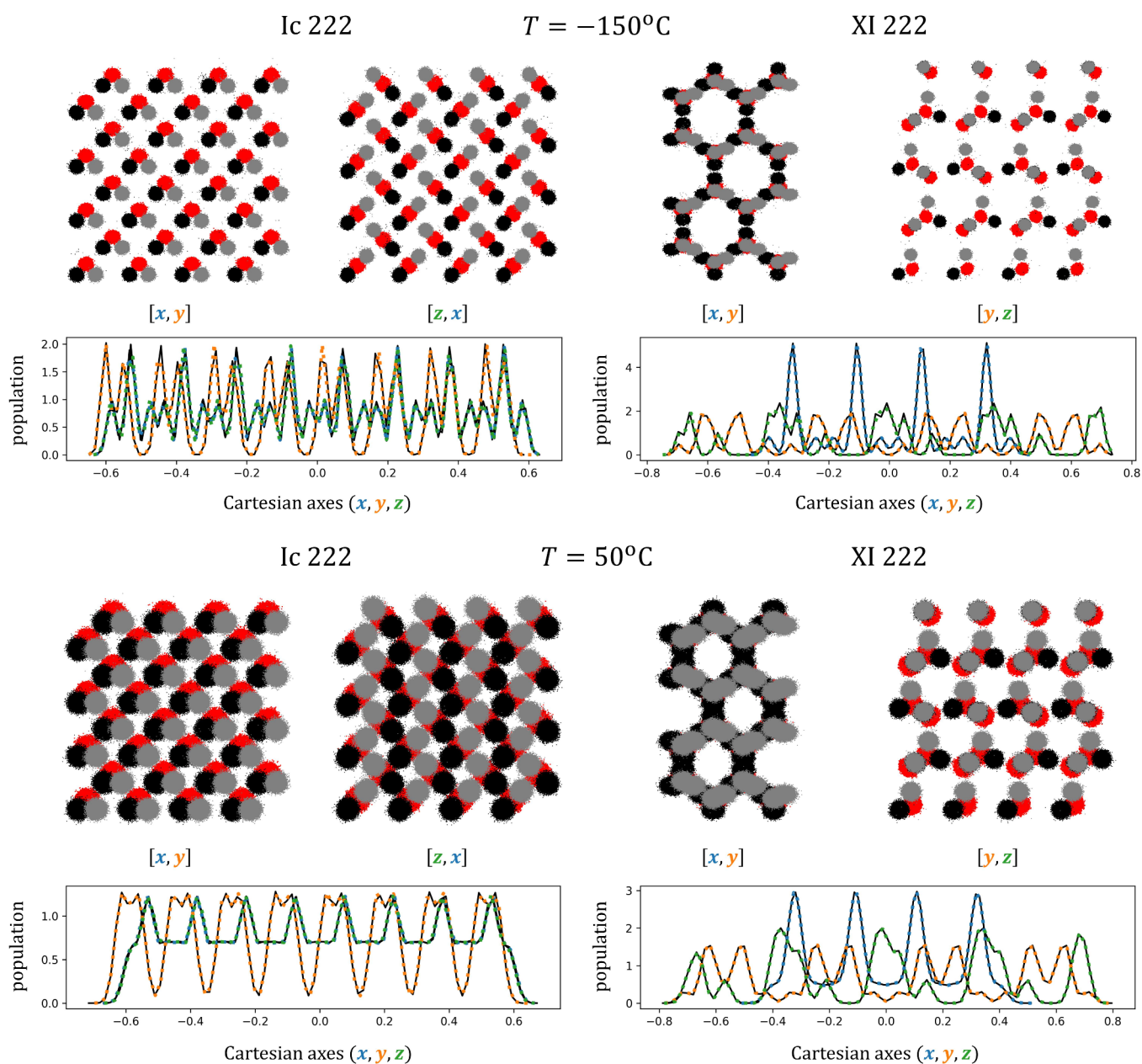


Figure 2. Samples of supercell configurations generated by transforming 20,000 random samples obtained from a uniform p_0 , and using four separate instances of model H, trained to map configurations of polymorphs XI and Ic at low and high temperature, respectively. The figures show that the four models produce physically consistent samples in the Cartesian space of supercell coordinates. The chosen Ice polymorphs do not exhibit proton disorder. By coloring the hydrogen positions in black and gray, we show that the model can consistently map their coordinates, thus enabling a consistent comparison with the ground-truth method. The probability density plots underneath the system's configurations report the marginal density obtained from MD data in black solid, and in color, dashed, the marginal probability computed from the mapped configurations, generated transforming p_0 . The histograms are virtually indistinguishable, with lines almost perfectly overlapped. An analogous figure showing samples from model C is reported in the SI.

introducing a weighting that accounts for model overfitting. In Figures 3 and 4, the standard error bars ($f^{\text{BAR-V}}(i) \pm \text{SE}(i)$) were also averaged using eq 9.

F. PGM Training. Both types of models (C and H) (see Figure 2) were trained and evaluated in the same way as in our previous work (ref³⁹). Specifically, to train a model, the following loss function (eq 11 defined in ref³⁹) was minimized via stochastic gradient descent (using Adam optimizer), with respect to the set of model parameters (Θ ; located inside the coupling layers; blue in Figure 1):

$$\min_{\Theta} L_{\text{ML}}(\Theta); L_{\text{ML}}(\Theta) = f - \langle \phi(\mathbf{r}; \Theta) \rangle_p \quad (10)$$

The averaging in eq 10 is performed over random batches of the locally ergodic MD training data that are distributed according to the true Boltzmann distribution $p(\mathbf{r}) = Z^{-1} \exp(-u(\mathbf{r}))$. The constant of interest $f = -\ln Z$, appearing eq 10, is unknown prior to performing a FE calculation but does not affect the gradient that is used for training the model $\partial_{\Theta} L_{\text{ML}}$. The quantity averaged in eq 10 corresponds to the generalized work function, discussed in section IID. Taking into account eq 8, makes $\partial_{\Theta} L_{\text{ML}} = -\langle \partial_{\Theta} \ln \gamma_{r \rightarrow z}(\mathbf{r}; \Theta) \rangle_p$.

G. MD Simulations. Unit cells of most forms of ice can be found in databases such as *American Mineralogist Crystal Structure Database* and *crystallography open database*, in the form of .cif files. The specific structures used in this paper are Ic⁵² and XI.⁵³ The .cif files were converted to .pdb files using *Mercury*.⁵⁴ Both unit cells contain 8 molecules. In this paper, we are focused on three system sizes (supercells containing 16, 32, and 64 molecules). The supercells containing these numbers of molecules were obtained by replicating the unit cells by translational stacking. Specifically, if the unit cell is 111, then the two respective sets of supercells of Ic and XI were 211,221,222 and 211,212,222. In creating supercells, we generally prefer supercells that are as close to cubic as possible.

All MD simulations were run using OpenMM⁵⁵ using its TIP4P/Ice force field implementation. The long-range electrostatics were treated using the Particle-Mesh Ewald (PME) algorithm, with an error tolerance parameter of 1e-5, a real space cutoff distance of 0.31 nm, and a switching distance of 0.279 nm. The cutoff (0.31 nm) was chosen considering the shortest box length in supercells 211 and 221 of ice Ic, which was 0.6358 nm.

To sample the NVT ensembles, the *LangevinMiddleIntegrator* integrator, with a friction coefficient of 20 ps⁻¹, was used. We note that, at 50 °C, both ice polymorphs are metastable with respect to liquid water; however, within the sampling performed to gather data for training PGM models, no melting events were observed. All simulations were performed with a time step of 2 fs, saving a configuration every 50 timesteps.

Before running the NVT simulations, each of the six supercells mentioned earlier was equilibrated for 1 ns using *MonteCarloAnisotropicBarostat* set to 1 atm at the relevant temperature. Although Open MM does not provide estimates of instantaneous pressure, the average box shapes equilibrated qualitatively well, according to the histograms of the box lengths during the simulation. Out of the 10,000 structures sampled during these simulations, a single structure with a box most similar to the average box was selected as the initial structure for all subsequent NVT simulations.

H. Ground Truth Free Energies with the Einstein Crystal Method. To assess the accuracy of FE estimates obtained from reweighted PGMs, we employed the Einstein Crystal method (ECM) based on ref.⁴³ which is representative of the state-of-the-art in estimating fully anharmonic FEs from MD data. In our Open MM-based implementation of the ECM, we adopted a scalar spring constant (k), leading to the following potential energy expression:⁴³

$$u_i(\mathbf{r}) = \lambda \beta U(\mathbf{r}) + (1 - \lambda) \frac{\beta k}{2} \|\mathbf{r} - \mathbf{r}_0\|^2 \quad (11)$$

In eq 11, λ is a coupling parameter controlling the linear interpolation between the physical ensemble, characterized by potential energy function (U), and the ensemble of the Einstein crystal, described by harmonic potential energy function ($\frac{k}{2} \|\mathbf{r} - \mathbf{r}_0\|^2$), where $\mathbf{r}_0$ can be any static configuration of the system with center of mass shifted to zero, and k is the scalar spring constant. We set $\mathbf{r}_0$ to the initial configuration, with the center of mass removed. The spring constant was fixed to 6×10^3 kJ(mol)⁻¹(nm)⁻² so that the Gaussian distributions associated with the harmonic potential have negligible periodic images along the smallest box length (0.6358 nm).⁴³ The ground truth FE of a given supercell ($f_{\text{cryst}}^{\text{ECM}}$) is then estimated as⁴³

$$f_{\text{cryst}}^{\text{ECM}} = f_0 + \sum_{i=1}^{n_\lambda-1} \Delta f_i; \text{ where } \Delta f_i = f_{\lambda_{i+1}} - f_{\lambda_i} \quad (12)$$

where $\mu_i = \left(\sum_{j=1}^n m_j\right)^{-1} m_i$ are normalized masses m_i of the atoms in the system. In this work, $n_\lambda = 20$ separate simulations were ran with λ parameters $\lambda_1, \dots, \lambda_{20}$ taking on values 0, ..., 1, respectively. The intermediate values, arranged in ascending order between zero and one, are reported in SI Figure 6(A). The same set of lambda values were used in all systems. Since the $\lambda_1 = 1$ simulation samples the translationally invariant ensemble of the unperturbed system of interest, all other simulations were initialized from $\mathbf{r}_0$ with *CMMotionRemover* functionality of *openMM* being active. This removes the average momentum of the center of mass during the simulations. Additionally, every 50 timesteps during these simulations, the center of mass was actively removed from the positions ($\mathbf{r}$) to make sure that all systems remained on the same $3(N-1)$ dimensional subspace. In eq 12, Δf_i was calculated $n_\lambda - 1 = 19$ times using BAR functionality of *pymbar*.⁵¹ This was done several times during the simulations to show how the FE estimates converge as a function of more data becoming available for the BAR calculations (SI Figure 6(B)). The set of *ground truth* FEs reported in this paper utilized all of the available data. The standard error of $f_{\text{cryst}}^{\text{ECM}}$ was estimated by summing the $n_\lambda - 1$ intermediate standard errors provided by *pymbar*. This approach provides a lower-bound estimate of the uncertainty on $f_{\text{cryst}}^{\text{ECM}}$, because this error estimate assumes complete decorrelation between the samples used to compute free energy differences between neighboring replicas. While underestimating the uncertainty of the ground-truth method, this approach is fast (i.e., does not require bootstrapping) and provides a relatively more ambitious target with which to compare the PGM-based methods. Finally, f_0 appearing in eq 12 is defined as

$$f_0 = \frac{3N}{2} \ln\left(\frac{\beta k}{2\pi}\right) + \frac{3}{2} \ln\left(\frac{2\pi}{\beta k} \sum_{i=1}^N \mu_i^2\right) + \ln\left(\frac{n_{\text{mol}}}{V}\right) \quad (13)$$

and takes three additive terms, including the FE of the Einstein crystal, the FE associated with the removal of the center of mass from the Einstein crystal, and the FE associated with removing the center of mass from the unperturbed physical system (where V was set to the volume of the box).⁴³ Please refer to ref⁴³ for further details about these terms.

Using the ground truth method described above (ECM), a total of 1.84 μs of data was sampled, with 0.32 μs of this data (17.4%) sampling the physical ensembles (i.e., $\lambda = 1$). This percentage is >5% because each physical simulation was four times longer than any other simulation of the same ECM run to collect sufficient MD data for later use in the PGM models. Collecting this data in the context of ECM ensured that the average potential energy in each system was the same across all FE methods being compared in this work. Notably, one major methodological discrepancy between the ground truth method and the PGMs can be linked to how the center of mass was removed. In the EC calculations, the center of mass of the supercells was removed based on the position of all atoms (including the hydrogens). Meanwhile, in the PGMs, the center of mass being removed was only based on the oxygen atoms, a more practical choice given the fact that the oxygens are used to anchor Cartesian blocks (eq 7). This methodological difference does introduce small systematic discrepancies between the *absolute* free energy of supercells that cancel out when

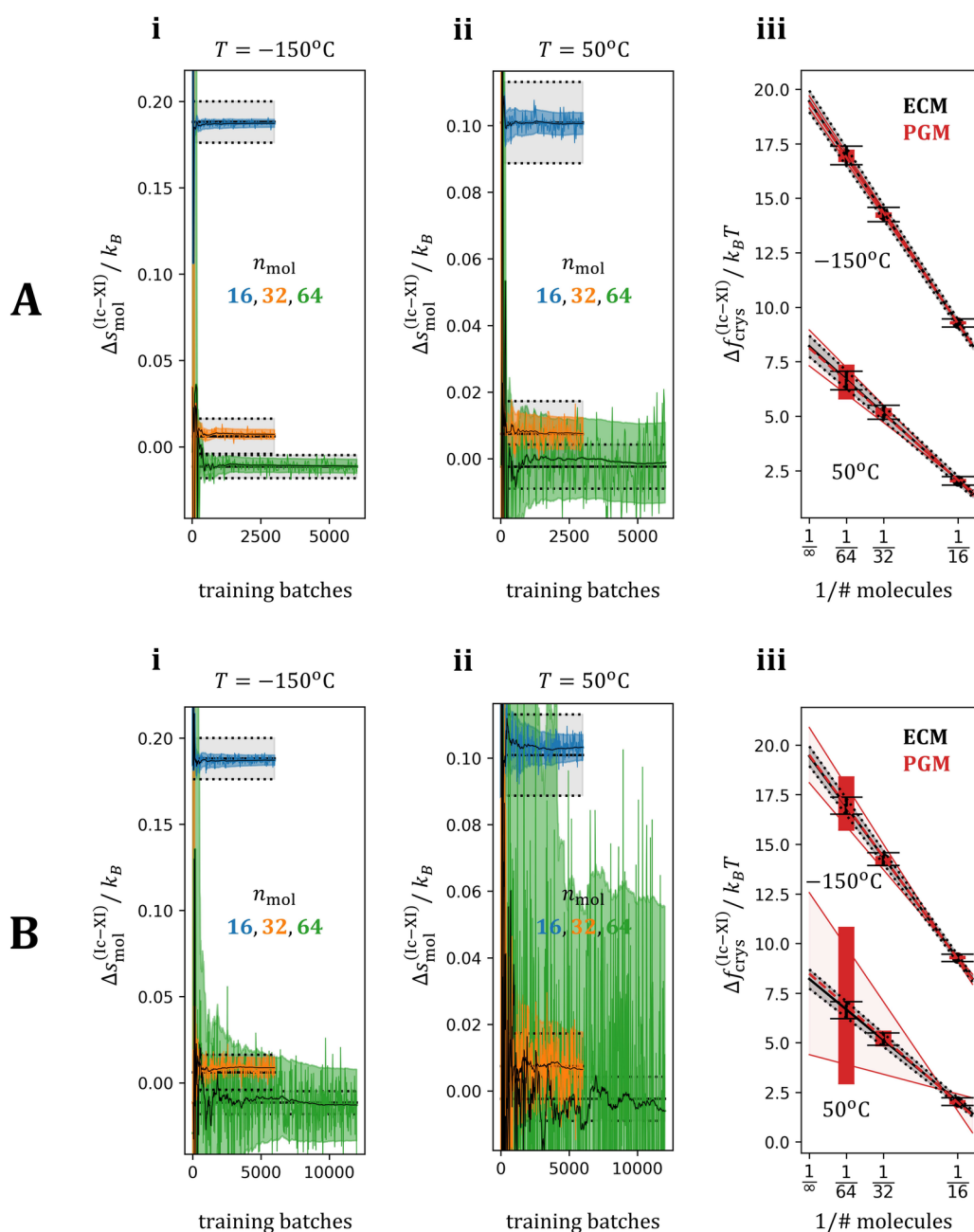


Figure 3. Free energy and entropy differences between ice polymorphs. A: model H, B: model C. The straight, black, solid colored lines represent reference ground truth computed with the ECM. The corresponding error bars are shown using straight black dotted lines. For cases where the running estimate of the PGM-based method renders it difficult to assess the ground truth accurately, we refer the reader to Table 1. Plots labeled with *i* and *ii* report the entropy differences per molecule between forms Ic and XI as a function of the number of training batches (i.e., during training). The estimates from models H and C are shown using colors corresponding to the three system sizes investigated in this study. The fluctuating colored lines indicate the *raw* estimates from the models, and the fluctuating, converging black curves indicate the cumulative weighted averages of these *raw* estimates, obtained as discussed in section IIE). Plots labeled *iii* display the differences in crystal-free energy between the two forms, with a line of best-fit extrapolating to approximate the same quantities in the thermodynamic limit (y-intercept). All transparent areas represent standard error bars (colored: PGM models, black: ECM ground truth).

considering free energy differences between polymorphs, as shown in the Results section.

III. RESULTS

In this section, we quantitatively assess the results from the two types of flow-based PGMs, i.e., models C and H, introduced in sections IIA and IIB, respectively. The ground truth estimates of the lattice free energy were obtained using the Einstein crystal method (ECM), introduced in section IIIH.

Two types of plots are used to visualize the comparisons between methods. Plots showing temperature-reduced crystal free energy differences $\Delta f_{\text{crys}}^{(A-B)} = f_{\text{crys}}^{(A)} - f_{\text{crys}}^{(B)}$ and plots showing temperature-reduced entropy differences per molecule $\Delta s_{\text{mol}}^{(A-B)} = s_{\text{mol}}^{(A)} - s_{\text{mol}}^{(B)}$, where (A, B) are either the two different forms of ice (Ic, XI) (Figure 3) or the two temperatures (−150 °C, 50 °C) (Figure 4).

The average potential energy in each state $\langle U \rangle_i$ is constant across all methods because it is obtained from the same

Table 1. Final Helmholtz Free Energy Differences (in units of $k_B T$) with Their Associated Standard Errors, between Ice Ic and XI Obtained Using Two PGM Models, H and C, and the ECM^a

$\Delta_{\text{Ic-XI}} @ T = -150\text{ }^{\circ}\text{C}$					
n_{mol}	$\Delta\langle f \rangle_{\text{ECM}}$	$\Delta\langle f \rangle_{\text{H}}$	$ \Delta\langle f \rangle_{\text{ECM}} - \Delta\langle f \rangle_{\text{H}} $	$\Delta\langle f \rangle_{\text{C}}$	$ \Delta\langle f \rangle_{\text{ECM}} - \Delta\langle f \rangle_{\text{C}} $
16	9.27 ± 0.19	9.29 ± 0.04	0.014 ± 0.23	9.29 ± 0.05	0.016 ± 0.24
32	14.25 ± 0.33	14.22 ± 0.09	0.033 ± 0.41	14.17 ± 0.18	0.084 ± 0.51
64	16.95 ± 0.43	16.94 ± 0.26	0.009 ± 0.68	17.05 ± 1.32	0.099 ± 1.75
∞	19.44 ± 0.49	19.41 ± 0.28	0.033 ± 0.78	19.49 ± 1.39	0.048 ± 1.89
$\Delta_{\text{Ic-XI}} @ T = 50\text{ }^{\circ}\text{C}$					
n_{mol}	$\Delta\langle f \rangle_{\text{ECM}}$	$\Delta\langle f \rangle_{\text{H}}$	$ \Delta\langle f \rangle_{\text{ECM}} - \Delta\langle f \rangle_{\text{H}} $	$\Delta\langle f \rangle_{\text{C}}$	$ \Delta\langle f \rangle_{\text{ECM}} - \Delta\langle f \rangle_{\text{C}} $
16	2.03 ± 0.20	2.04 ± 0.05	0.005 ± 0.25	2.00 ± 0.06	0.037 ± 0.26
32	5.18 ± 0.32	5.18 ± 0.15	0.002 ± 0.47	5.21 ± 0.36	0.033 ± 0.67
64	6.64 ± 0.42	6.56 ± 0.77	0.075 ± 1.20	6.87 ± 3.93	0.234 ± 4.36
∞	8.21 ± 0.48	8.13 ± 0.82	0.078 ± 1.31	8.48 ± 4.08	0.269 ± 4.57

^aThe accuracy of the PGM estimates is assessed by computing their absolute deviation with respect to the ECM ground truth as $|\Delta\langle f \rangle_{\text{ECM}} - \Delta\langle f \rangle_i|$, where $i = \text{H, C}$. This analysis indicates that while agreement is excellent across all models, model H estimates are closer to ECM despite being trained on the same data. The underlying values of *absolute* FE estimates for ice Ic and XI are reported in SI Figure 6.

underlying set of MD-generated configurations. Hence, by comparing the entropic contribution to the lattice free energy s_i , we introduce a more stringent criterion to assess the accuracy of free energy estimates from PGMs trained by likelihood, akin to our previous work.³⁹

The per-molecule temperature-reduced entropy of polymorph i is defined as

$$s_i = -n_{\text{mol},i}^{-1}(\langle f_i \rangle - \beta\langle U \rangle_i) \quad (14)$$

where state i is specified by the identity of the metastable state (Form) and the physical conditions of the given Canonical ensemble (N, V, T).

Table 1, based on raw data reported in SI Figure S5, reports the lattice Helmholtz free energy differences (in units of $k_B T$) between the two forms of ice, computed using all of the methods (i.e., from models H and C, and from the ECM). These comparisons show a significant level of agreement between the different methods, indicating that the results from the PGMs are accurate. We note, however, that the error bars are higher in larger systems and at higher temperatures because such ensembles occupy a larger volume of configurational space. That being said, we consistently observe smaller error bars from results of model H, compared to model C. This indicates that model H (based on the decoupled, quaternion-based representation of the Cartesian blocks) is better suited for efficiently learning configurational distributions of ice crystals. Moreover, since both models we compare used the same architecture for the trainable part of the model (spline coupling layers; see Figure 1), we expect that the computational gains from using the decoupled representation of the Cartesian block can be generalized to other types of crystals.

Figure 3 reports per-molecule, temperature-reduced entropy difference (in units of k_B) between the two forms of ice (Ic and XI), estimated using the different methods. In these comparisons, the two temperatures are plotted separately. Figures 3A (i, ii) and Figure 3B (i, ii) report estimates obtained from the PGM models H and C respectively, plotted as a function of the training progress. The differences between panels A and B in Figure 3 indicate that model H was able to converge more rapidly and with lower variance compared to model C. These estimates are systematically comparable to the ECM ground truth (straight black lines).

The reason for the lower variance in the BAR_V estimates from model H, compared to model C, can be linked to the differences in the amount of overlap observed between the potential energies of the data and the potential energies of the samples drawn from the models during training (shown in SI Figures 2 and 3). When taken together, the current results show that in the large systems containing 64 molecules and at the higher temperature (50 °C) (green error bars in Figure 3A and B, (ii), both models tend to approach a bottleneck in accuracy.

Both models H and C enable us to consistently extrapolate the lattice free energy to the thermodynamic limit, achieving results consistent with the ECM ground truth (see Figure 3A and B, (iii)). Also, in this case, the lower error associated with model H is evident, and the level of accuracy in the high-temperature simulations is of the same order as the ECM ground truth.

Figure 4 reports per-molecule, temperature-reduced entropy differences between low and high-temperature ensembles of the same systems, based on the estimates from the three methods (PGM models: H (panel A) and C (panel B), and the ground truth method: ECM (all panels)). These estimates are plotted as a function of training progress. Also, in this case, the agreements between the PGM-based and the ground truth (ECM-based) estimates of the temperature-dependent entropy difference in the three system sizes are significant. As for the comparison between polymorphs, estimates obtained from model H are associated with significantly smaller error bars than the estimates from model C, once again underscoring the importance of adopting a physically motivated representation in constructing flow-based maps for targeted FE calculations.

For completeness, samples drawn from the larger models (with 64 molecules) are visualized in Figure 2 and SI Figure 4. There are no significant qualitative differences between the two types of models (model C vs H respectively).

Our results clearly show that both the computational costs associated with training and the numerical precision of the resulting free energy estimates obtained from flow-based PGMs, even when accurately reweighted,³⁹ are strongly affected by the representation of the relevant DOFs.

IV. DISCUSSION AND CONCLUSIONS

In this work, we have demonstrated that given an accurate reweighting scheme,³⁹ the lattice free energy differences between packing polymorphs of flexible ice crystals, modeled

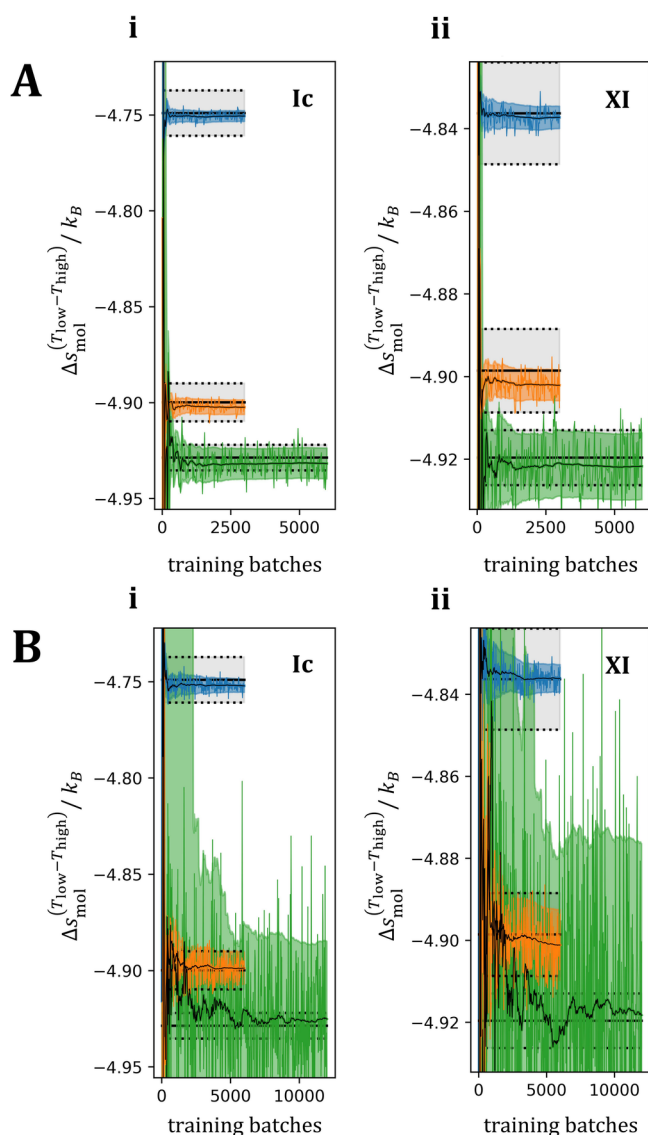


Figure 4. Free energy differences across temperatures, rather than across polymorphs as reported in Figure 3. A: model H, B: model C. The solid, straight black lines represent the ground truth estimates (from ECM). Plots labeled with *i* and *ii* report the entropy differences per molecule between two versions of the same systems at the two temperatures ($T_{\text{low}} = -150^\circ\text{C}$ and $T_{\text{high}} = 50^\circ\text{C}$), as a function training progress. The estimates from models H and C are represented using colors corresponding to the three system sizes investigated in this study. The fluctuating colored curves are the *raw* BAR_V estimates on random subsets of 10,000 points from the validation set, and the black curves are the cumulative weighted averages of these *raw* estimates, obtained as discussed in section IIE). All transparent regions represent standard error bars (colored: PGM models, black: ECM ground truth).

using a classical force field (TIP4P-ice), can be efficiently computed using *targeted* free energy calculations based on normalizing flow models, remapping from a common, flat, analytical base distribution.

Demonstrating the method with a uniform base distribution is significant because it can be generalized to model permutationally invariant distributions, which has been a highlight in prior literature.^{26,27,30,31} Furthermore, training with a flat base distribution generates flows that can be easily composed back-to-back to connect any pair of ensembles of the same dimensionality, thus opening up the possibility of connecting

different packings to a *common* base distribution. This, in turn, enables reweighting with Multistate BAR, as discussed in ref.³⁸ and demonstrated for isolated molecules in ref.³⁹ Furthermore, mapping across packing polymorphs via a flat base distribution enables training a number of maps equal to the number of metastable states, thus scaling favorably with the number of polymorphs investigated and setting the scene for large-scale applications of free energy methods based on PGMs. The results illustrated here show that mapping via a flat base distribution is not only possible but also accurate and computationally cheaper than ground truth with the whole range of temperature conditions of practical relevance for finite-temperature crystal structure prediction.

Our results show that this approach can handle molecular crystals where all the internal degrees of freedom of a given Cartesian block are explicitly modeled. As such, it is a representative example of how Cartesian blocks can be treated in larger molecules. We, therefore, believe that the current work is a useful stepping stone toward assessing the applicability of PGM-based free energy methods to complex and realistic crystal systems, including molecular crystals with multiple conformational degrees of freedom.⁶

Across all tests performed in this work, we note how the extrapolation to the thermodynamic limit of an infinitely large cell shows a convincingly convergent behavior enabling a quantitative removal of finite-size effects from both lattice-free energy differences and entropies. The order of magnitude of the uncertainty is particularly important in these estimates. Our current setup showed that a PGM-based approach such as model H enables matching the ground-truth error levels.

We note that the PGM approach is more computationally efficient than the Einstein crystal method (ECM), as it requires performing only two simulations of the physical end-states of interest vs multiple simulations in perturbed Hamiltonians necessary for converging the ECM. The difference in cost can be quantified by comparing the amount of MD-generated samples used by the two methods. In this regard, PGM requires only $\approx 17\%$ of the configurations needed by the ECM, leading to an increase in MD efficiency of factors between 5 and 6. Given that the FEs from PGMs in the current work have converged quite quickly during training, we anticipate that further efficiency gains can be reached in more complex molecular systems.

We further note that the computational overheads associated with training PGMs by maximum likelihood are *agnostic* to the quality and computational cost of the potential energy functions used to sample the training data. This opens up future possibilities of computing fully anharmonic free energy differences, remapped to force fields at a higher level of theory.^{37,38}

The choice of a common base distribution, the application to FE differences in a wide and practically relevant range of temperatures, the flexible representation of individual molecules, and the application to packing polymorphs with nonoverlapping high-temperature distributions provide a step forward with respect to the literature on PGM-based FE calculations on molecular crystals.³⁰ As such, these results represent a stepping stone toward adopting normalizing flow methods into the anharmonic FE toolbox of practitioners dealing with finite temperature ranking of large sets of molecular crystalline solids. More broadly, when taken together with several key recent developments in this field,^{28,31} we can further expect normalizing flows to begin solving increasingly diverse real-world problems in the foreseeable future.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.4c01612>.

Figure reporting a representative coupling pattern between marginals, Analysis of the numerical error associated with inversion, Description of the whitening approach, potential energy overlaps obtained during training, qualitative comparison of configurations generated by PGM and MD simulation, convergence of the ECM calculations and spacing of the λ replicas, table summary of the absolute FE of every supercell (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Matteo Salvalaglio – Thomas Young Centre and Department of Chemical Engineering, University College London, London WC1E 7JE, United Kingdom; orcid.org/0000-0003-3371-2090; Email: m.salvalaglio@ucl.ac.uk

Authors

Edgar Olehnovics – Thomas Young Centre and Department of Chemical Engineering, University College London, London WC1E 7JE, United Kingdom

Yifei Michelle Liu – Molecular Profiling and Drug Delivery, Research & Development, AbbVie Bioresearch Center, Worcester, Massachusetts 01605, United States

Nada Mehio – Molecular Profiling and Drug Delivery, Research & Development, AbbVie Inc., North Chicago, Illinois 60064, United States

Ahmad Y. Sheikh – Molecular Profiling and Drug Delivery, Research & Development, AbbVie Inc., North Chicago, Illinois 60064, United States; orcid.org/0000-0002-5972-3938

Michael R. Shirts – University of Colorado Boulder, Boulder, Colorado 80309, United States; orcid.org/0000-0003-3249-1097

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.4c01612>

Notes

The authors declare the following competing financial interest(s): A.Y.S., N.M., and Y.M.L. are employees of AbbVie and may own AbbVie stock.

■ ACKNOWLEDGMENTS

E.O. acknowledges funding from Abbvie. M.S. acknowledges funding from the Crystallization in the Real World EPSRC Programme Grant (EP/R018820/1) and from the ht-MATTER UKRI Frontier Research Guarantee Grant (EP/X033139/1). M.R.S. acknowledges support from the National Science Foundation under grant no. CHE-2331939. AbbVie sponsored and partially funded the study, contributed to the design, and participated in the collection, analysis, and interpretation of data and in writing, reviewing, and approval of the final publication.

■ REFERENCES

- (1) Wang, Lingle; Wu, Yujie; Deng, Yuqing; Kim, Byungchan; Pierce, Levi; Krilov, Goran; Lupyan, Dmitry; Robinson, Shaughnessy; Dahlgren, Markus K; Greenwood, Jeremy; et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015**, *137* (7), 2695–2703.
- (2) Hong, Richard S; Mattei, Alessandra; Sheikh, Ahmad Y; Tuckerman, Mark E A data-driven and topological mapping approach for the a priori prediction of stable molecular crystalline hydrates. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (43), No. e2204414119.
- (3) Hong, Richard S; Rojas, Ana V; Bhardwaj, Rajni Miglani; Wang, Lingle; Mattei, Alessandra; Abraham, Nathan S; Cusack, Kevin P; Pierce, M Olivia; Mondal, Sayan; Mehio, Nada; et al. Free energy perturbation approach for accurate crystalline aqueous solubility predictions. *J. Med. Chem.* **2023**, *66* (23), 15883–15893.
- (4) Duarte Ramos Matos, Guilherme; Kyu, Daisy Y; Loeffler, Hannes H; Chodera, John D; Shirts, Michael R; Mobley, David L Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the freesolv database. *Journal of Chemical & Engineering Data* **2017**, *62* (5), 1559–1569.
- (5) Yang, Mingjun; Dybeck, Eric; Sun, Guangxu; Peng, Chunwang; Samas, Brian; Burger, Virginia M; Zeng, Qun; Jin, Yingdi; Bellucci, Michael A; Liu, Yang; et al. Prediction of the relative free energies of drug polymorphs above zero kelvin. *Cryst. Growth Des.* **2020**, *20* (8), 5211–5224.
- (6) Firaha, Dzmitry; Liu, Yifei Michelle; Streek, Jacco van de; Sasikumar, Kiran; Dietrich, Hanno; Helfferich, Julian; Aerts, Luc; Braun, Doris E; Broo, Anders; DiPasquale, Antonio G; et al. Predicting crystal form stability under real-world conditions. *Nature* **2023**, *623* (7986), 324–328.
- (7) Mattei, A.; Hong, R. S.; Dietrich, H.; Firaha, D.; Helfferich, J.; Liu, Y. M.; et al. Efficient crystal structure prediction for structurally related molecules with accurate and transferable tailor-made force fields. *J. Chem. Theory Comput.* **2022**, *18* (9), 5725–5738.
- (8) Beran, Gregory JO Frontiers of molecular crystal structure prediction for pharmaceuticals and functional organic materials. *Chemical Science* **2023**, *14* (46), 13290–13312.
- (9) Burcham, Christopher L; Doherty, Michael F; Peters, Baron G; Price, Sarah L; Salvalaglio, Matteo; Reutzel-Edens, Susan M; Price, Louise S; Kumar Reddy Addula, Ravi; Francia, Nicholas; Khanna, Vikram; et al. Pharmaceutical digital design: From chemical structure through crystal polymorph to conceptual crystallization process. *Cryst. Growth Des.* **2024**, *24* (13), 5417–5438.
- (10) Price, Sarah L Why don't we find more polymorphs? *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2013**, *69* (4), 313–328.
- (11) Reilly, Anthony M; Cooper, Richard I; Adjiman, Claire S; Bhattacharya, Saswata; Boese, A Daniel; Brandenburg, Jan Gerit; Bygrave, Peter J; Bylsma, Rita; Campbell, Josh E; Car, Roberto; et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2016**, *72* (4), 439–459.
- (12) Hoja, Johannes; Ko, Hsin-Yu; Neumann, Marcus A; Car, Roberto; DiStasio, Robert A, Jr; Tkatchenko, Alexandre Reliable and practical computational description of molecular crystal polymorphs. *Science Advances* **2019**, *5* (1), No. eaau3338.
- (13) Dybeck, Eric C; Abraham, Nathan S; Schieber, Natalie P; Shirts, Michael R Capturing entropic contributions to temperature-mediated polymorphic transformations through molecular modeling. *Cryst. Growth Des.* **2017**, *17* (4), 1775–1787.
- (14) Hunnisett, Lily M.; Nyman, Jonas.; Francia, Nicholas.; Abraham, Nathan S.; Adjiman, Claire S.; Aitipamula, Srinivasulu.; Alkhalid, Tamador.; Almehairbi, Mubarak.; Anelli, Andrea.; Anstine, Dylan M. The seventh blind test of crystal structure prediction: structure generation methods. *Acta crystallographica Section B, Structural science, crystal engineering and materials* **2024**, *80*, 517.
- (15) Hunnisett, Lily M.; Francia, Nicholas.; Nyman, Jonas.; Abraham, Nathan S.; Aitipamula, Srinivasulu.; Alkhalid, Tamador.; Almehairbi, Mubarak.; Anelli, Andrea.; Anstine, Dylan M. The seventh blind test of crystal structure prediction: structure ranking methods. *Acta Crystallographica Section B* **2024**, *80* (6), 548.
- (16) Chipot, C.; Pohorille, A. *Free energy calculations*, volume 86. Springer.
- (17) Frenkel, D.; Smit, B. *Understanding molecular simulation: From algorithms to applications*. Academic Press, 1996.

- (18) Li, Lunna; Totton, Tim.; Frenkel, Daan. Computational methodology for solubility prediction: Application to the sparingly soluble solutes. *J. Chem. Phys.* **2017**, *146* (21), 214110.
- (19) Khanna, Vikram.; Anwar, Jamshed.; Frenkel, Daan.; Doherty, Michael F.; Peters, Baron. Free energies of crystals computed using Einstein crystal with fixed center of mass and differing spring constants. *J. Chem. Phys.* **2021**, *154* (16), 164509.
- (20) Bruce, A. D.; Jackson, A. N.; Ackland, G. J.; Wilding, N. B. Lattice-switch monte carlo method. *Phys. Rev. E* **2000**, *61* (1), 906.
- (21) Kamat, Kartik.; Peters, Baron. Gibbs free-energy differences between polymorphs via a diabat approach. *J. Chem. Phys.* **2018**, *149* (21), 214106.
- (22) Kamat, Kartik.; Guo, Rui.; Reutzel-Edens, Susan M.; Price, Sarah L.; Peters, Baron. Diabat method for polymorph free energies: Extension to molecular crystals. *J. Chem. Phys.* **2020**, *153* (24), 244105.
- (23) Piaggi, Pablo M.; Valsson, Omar; Parrinello, Michele Enhancing entropy and enthalpy fluctuations to drive crystallization in atomistic simulations. *Physical review letters* **2017**, *119* (1), 015701.
- (24) Gimondi, Ilaria.; Salvalaglio, Matteo. CO₂ packing polymorphism under pressure: Mechanism and thermodynamics of the i-iii polymorphic transition. *J. Chem. Phys.* **2017**, *147* (11), 114502.
- (25) Giberti, Federico.; Salvalaglio, Matteo.; Mazzotti, Marco; Parrinello, Michele Insight into the nucleation of urea crystals from the melt. *Chem. Eng. Sci.* **2015**, *121*, 51–59.
- (26) Wirnsberger, Peter; Ballard, Andrew J.; Papamakarios, George; Abercrombie, Stuart; Racanière, Sébastien; Pritzel, Alexander; Rezende, Danilo Jimenez; Blundell, Charles Targeted free energy estimation via learned mappings. *J. Chem. Phys.* **2020**, *153* (14), 144112.
- (27) Wirnsberger, Peter; Papamakarios, George; Ibarz, Borja; Racanière, Sébastien; Ballard, Andrew J.; Pritzel, Alexander; Blundell, Charles Normalizing flows for atomic solids. *Machine Learning: Science and Technology* **2022**, *3* (2), 025009.
- (28) Wirnsberger, Peter.; Ibarz, Borja.; George, Papamakarios. Gibbs free energies via isobaric-isothermal flows. May 2023. *arXiv:2305.13233* [cond-mat, physics:physics, stat].
- (29) Ahmad, Rasool; Cai, Wei Free energy calculation of crystalline solids using normalizing flow. *Modell. Simul. Mater. Sci. Eng.* **2022**, *30* (6), 065007.
- (30) Köhler, Jonas.; Invernizzi, Michele.; de Haan, Pim.; Noé, Frank. Rigid body flows for sampling molecular crystal structures. January 2023. *arXiv:2301.11355* [physics, stat].
- (31) Schebek, M.; Invernizzi, Michele.; Noé, Frank.; Rogal, Jutta Efficient mapping of phase diagrams with conditional normalizing flows, June 2024. *arXiv:2406.12378* [cond-mat].
- (32) Hahn, A. M.; Then, H. Using bijective maps to improve free energy estimates. *Phys. Rev. E* **2009**, *79* (1), 011113.
- (33) Paliwal, H.; Shirts, Michael R. Multistate reweighting and configuration mapping together accelerate the efficiency of thermodynamic calculations as a function of molecular geometry by orders of magnitude. *J. Chem. Phys.* **2013**, *138* (11), 154108.
- (34) Noé, Frank.; Olsson, S.; Köhler, Jonas.; Hao, Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **2019**, *365* (6457), eaaw1147.
- (35) Ding, X.; Zhang, B. Computing Absolute Free Energy with Deep Generative Models. *The Journal of Physical Chemistry B* **2020**, *124* (45), 10166–10172.
- (36) Ding, X.; Zhang, B. DeepBAR: A Fast and Exact Method for Binding Free Energy Computation. *The Journal of Physical Chemistry Letters* **2021**, *12* (10), 2509–2515.
- (37) Rizzi, Andrea; Carloni, Paolo; Parrinello, Michele Targeted Free Energy Perturbation Revisited: Accurate Free Energies from Mapped Reference Potentials. *J. Phys. Chem. Lett.* **2021**, *12* (39), 9449–9454.
- (38) Rizzi, Andrea; Carloni, Paolo; Parrinello, Michele Multimap targeted free energy estimation. *Proc. Natl. Acad. Sci. U. S. A.* **2023**, *120* (46), e2304308120.
- (39) Olehnovics, Edgar; Liu, Yifei Michelle; Mehio, Nada; Sheikh, Ahmad Y.; Shirts, Michael R.; Salvalaglio, Matteo Assessing the Accuracy and Efficiency of Free Energy Differences Obtained from Reweighted Flow-Based Probabilistic Generative Models. *J. Chem. Theory Comput.* **2024**, *20* (14), 5913–5922.
- (40) Durkan, Conor.; Bekasov, Artur.; Murray, Iain.; Papamakarios, George. Neural Spline Flows. December 2019. *arXiv:1906.04032* [cs, stat].
- (41) Dinh, Laurent.; Krueger, David.; Bengio, Yoshua. NICE: Non-linear Independent Components Estimation. April 2015. *arXiv:1410.8516* [cs].
- (42) Gao, Christina; Isaacson, Joshua; Krause, Claudius i-flow: High-dimensional Integration and Sampling with Normalizing Flows. *Machine Learning: Science and Technology* **2020**, *1* (4), 045023.
- (43) Khanna, Vikram.; Anwar, Jamshed.; Frenkel, Daan.; Doherty, Michael F.; Peters, Baron Free energies of crystals computed using Einstein crystal with fixed center of mass and differing spring constants. *J. Chem. Phys.* **2021**, *154* (16), 164509.
- (44) Parsons, Jerod; Bradley Holmes, J.; Maurice Rojas, J.; Tsai, Jerry; Strauss, Charlie E. M. Practical conversion from torsion space to Cartesian space for in silico protein synthesis. *J. Comput. Chem.* **2005**, *26* (10), 1063–1068.
- (45) GitHub - noegroup/rigid-flows.
- (46) Lynch, Kevin M.; Park, Frank C. *Modern Robotics: Mechanics, Planning, and Control*. Cambridge University Press, 2017.
- (47) Gallier, J. The Quaternions and the Spaces S³, SU(2), SO(3), and P³. In Gallier, J., editor, *Geometric Methods and Applications: For Computer Science and Engineering*, pages 248–266. Springer, New York, NY, 2001.
- (48) Alcorn, Michael A. AQuaMaM: An Autoregressive, Quaternion Manifold Model for Rapidly Estimating Complex SO(3) Distributions, January 2023. *arXiv:2301.08838* [cs].
- (49) Gilitshenski, Igor.; Sahoo, Roshni.; Schwarting, Wilko.; Amini, Alexander.; Karaman, Sertac.; Rus, Daniela. Deep Orientation Uncertainty Learning based on a Bingham Loss. September 2019.
- (50) Alcorn, Michael A. AQuaMaM: An Autoregressive, Quaternion Manifold Model for Rapidly Estimating Complex SO(3) Distributions, January 2023. *arXiv:2301.08838* [cs].
- (51) pymbar, April 2023. original-date: 2013–05–10T22:45:44Z.
- (52) Kuhs, W. F.; Bliss, D. V.; Finney, J. L. High-resolution neutron powder diffraction study of ice Ic. *Le Journal de Physique Colloques* **1987**, *48* (C1), C1–636.
- (53) Leadbetter, A. J.; Ward, R. C.; Clark, J. W.; Tucker, P. A.; Matsuo, T.; Suga, H. The equilibrium low-temperature structure of ice. *J. Chem. Phys.* **1985**, *82* (1), 424–428.
- (54) Macrae, C. F.; Sovago, I.; Cottrell, S. J.; Galek, P. T. A.; McCabe, P.; Pidcock, E.; Platings, M.; Shields, G. P.; Stevens, J. S.; Towler, M.; Wood, P. A. Mercury 4.0: from visualization to analysis, design and prediction. *J. Appl. Crystallogr.* **2020**, *53* (1), 226–235.
- (55) Eastman, Peter; Swails, Jason; Chodera, John D.; McGibbon, Robert T.; Zhao, Yutong; Beauchamp, Kyle A.; Wang, Lee-Ping; Simmonett, Andrew C.; Harrigan, Matthew P.; Stern, Chaya D.; Wiewiora, Rafal P.; Brooks, Bernard R.; Pande, Vijay S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology* **2017**, *13* (7), e1005659.