

# Maximum Entropy-Mediated Liquid-to-Solid Nucleation and Transition

Lars Dammann,\* Richard Kohns, Patrick Huber, and Robert H. Meißner\*



Cite This: <https://doi.org/10.1021/acs.jctc.4c01621>



Read Online

ACCESS |



Metrics & More

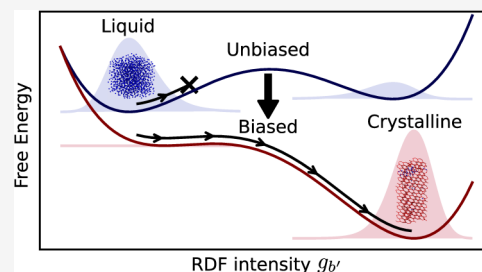


Article Recommendations



Supporting Information

**ABSTRACT:** Molecular dynamics (MD) simulations are a powerful tool for studying matter at the atomic scale. However, to simulate solids, an initial atomic structure is crucial for the successful execution of MD simulations but can be difficult to prepare due to insufficient atomistic information. At the same time, wide-angle X-ray scattering (WAXS) measurements can determine the radial distribution function (RDF) of atomic structures. However, the interpretation of RDFs is often challenging. Here, we present an algorithm that can bias MD simulations with RDFs by combining the information on the MD atomic interaction potential and the RDF under the principle of maximum relative entropy. We show that this algorithm can be used to adjust the RDF of one liquid model, e.g., the TIP3P water model, to reproduce the RDF and improve the angular distribution function (ADF) of another model, such as the TIP4P/2005 water model. In addition, we demonstrate that the algorithm can initiate crystallization in liquid systems, leading to both stable and metastable crystalline states defined by the RDF, e.g., crystallization of water to ice and liquid  $\text{TiO}_2$  to rutile or anatase. Finally, we discuss how this method can be useful for improving interaction models, studying crystallization processes, interpreting measured RDFs, or training machine-learned potentials.



## 1. INTRODUCTION

Understanding the structure of solid matter at the atomic level is an important task in materials science and a prerequisite for the development of new tailored materials. To this end, Molecular Dynamics (MD) simulations are an important method to study materials at the atomic scale. A crucial component to perform a MD simulation are suitable models of the interatomic potentials. However, knowledge of the interatomic potential alone is not sufficient to predict the structure of stable or metastable solid atomic structures. Therefore, the atomic structure of interest must be provided as another essential requirement to perform a physically meaningful MD simulation.

Sometimes it is not possible to provide the required atomic structure *ab initio*, because it is difficult to generate, e.g., in the case of amorphous structures, or not enough details are known about the exact atomic configuration. In this case an alternative option is to generate the target structure during the MD simulation from an initially unstable random or liquid phase. For this strategy to be successful, it is sometimes necessary to induce a phase transition in the initial system. This requires the system to move from one local free energy minimum to another stable free energy minimum by overcoming a free energy barrier. To date, several methods have been developed that can be used to overcome the free energy barrier and generate realistic solid atomic configurations during a MD simulation.

Some examples of these methods include crystal seed insertion,<sup>1–5</sup> simulated annealing,<sup>6,7</sup> the introduction of a bias

potential in one form or another,<sup>8–15</sup> e.g., umbrella sampling<sup>16</sup> or well-tempered metadynamics,<sup>17</sup> or replica exchange methods.<sup>18–23</sup> However, each strategy has its own challenges. The introduction of a crystal seed requires a seed of the correct crystalline structure, which is problematic if the correct crystalline structure is not available. Introducing a meaningful bias potential requires the choice of a set of collective variables that summarize the position of the system in the free energy landscape. The collective variables usually have to be developed and chosen individually for each problem. In contrast to these methods, simulated annealing and replica exchange methods are structure agnostic. However, the methods cannot be used to target a specific stable or metastable state of interest, since simulated annealing is designed to find the state of the global energy minimum and replica exchange methods are undirected with respect to the sampled configurations. In addition, it can be quite challenging to find the right parameters for the successful application of simulated annealing, biasing methods, or replica exchange methods.

**Received:** November 28, 2024

**Revised:** January 17, 2025

**Accepted:** January 27, 2025

A common approach to the determination of unknown atomic structures is an experimental investigation with Wide Angle X-ray Scattering (WAXS). WAXS allows the determination of radial distribution functions (RDF)  $g(r)$  of an atomic structure by Fourier transforming the measured isotropic total scattering structure function  $S(q)$ , which depends on the magnitude of the reciprocal scattering vector  $q = \frac{2\pi}{r}$ .<sup>24</sup>

However, the interpretation of the atomic structure function  $S(q)$  or the derived RDF  $g(r)$  without any prior structural knowledge is an ill-posed inverse problem, since the generated RDF could originate from different atomic structures all resembling the same  $S(q)$ .

Considering the potential struggle to prepare atomic structures in MD simulations and to determine unknown atomic structures from experimental approaches, the question arises whether the information from the RDF can be combined with the information contained in the atomic interaction potentials used in MD simulations to generate atomic structures. For example, the Reverse Monte Carlo (RMC) method<sup>25–27</sup> attempts to reconstruct the atomic structure from the atomic structure function  $S(q)$  alone with unquantifiable fidelity. The Hybrid Reverse Monte Carlo (HRMC) method and its extension incorporating MD sampling<sup>28,29</sup> on the other hand aims to improve the RMC algorithm by combining it with a molecular sampling scheme to reproduce the RDF  $g(r)$  or the structure factor function  $S(q)$ . In HRMC, the RMC algorithm is added as a bias to the potential energy of the system. However, the mathematical structure of the introduced bias not only constrains the ensemble mean of the system to reproduce the RDF or the structure factor function, but it also constrains the variance of how the target function  $g(r)$  or  $S(q)$  can be reproduced from the system ensemble.<sup>30</sup> Since information about the variance of the objective function from an atomistic point of view is usually not available from experiments, this additional constraint is not supported by the experimental data. Furthermore, the application of the HRMC method requires the specification of a coupling parameter that determines the strength of the bias on the original potential energy of the system. The choice of the coupling parameter can drastically alter the simulation results and should be based on the level of confidence in the original atomic potential or the influence of the experimentally determined data. The proper estimation of this trade-off makes the choice of the appropriate coupling parameter a difficult task. Therefore, a method that can incorporate the information contained in an RDF into MD simulations in a way that avoids inappropriate biases of the original force fields and facilitates the choice of appropriate coupling parameters would be an important step forward.

A suitable approach to combine information from measured observables with a thermodynamic system is the application of the principle of maximum entropy, first introduced by Pitera and Chodera<sup>30</sup> and subsequently reviewed for the related principle of maximum relative entropy in the context of atomistic modeling by Cesari et al.<sup>31</sup> Prior to our work, relative entropy applications have been used in the development of coarse-graining potentials<sup>32–34</sup> and the improvement of existing force fields.<sup>35,36</sup> Cilloco<sup>37</sup> demonstrated that it is possible to reproduce a simulated target RDF of a Lennard-Jones fluid in a Monte Carlo simulation from a maximum entropy approach. Subsequently, the resulting interaction bias was found to be in agreement with the Lennard-Jones

interactions that were used to create the target RDF. In more recent work, White and Voth<sup>38</sup> biased molecular dynamics simulations to reproduce ensemble averages using a maximum relative entropy approach. In this study, the authors indirectly reproduced the target RDF of liquid systems by biasing the moments of the atomic coordination numbers. In a subsequent publication, White et al.<sup>39</sup> presented a method for designing free energy surfaces that align with experimental data under the principle of maximum entropy with metadynamics.

In contrast to these prior works, the approach presented here, directly reproduces RDFs from the individual bins of a target RDF, thereby allowing the reproduction of not only partial RDFs but also total RDFs consisting of multiple partial weighted RDFs. Additionally, it offers greater control over the introduced virial pressure into the system. Furthermore, we show that with the presented approach, it can be sufficient to bias liquid systems with RDFs from stable or metastable crystalline states to induce a liquid–solid phase transition to a corresponding state that is defined by the target RDF.

## 2. THEORY

**2.1. Radial Distribution Functions.** We can describe the radial distribution function of a macroscopic physical system with a large number of identical atoms  $\tilde{N} \sim 10^{23}$  as a histogram  $\tilde{g}_b$  with bins  $b$  ranging from 1, ...,  $B$ . Each bin extends over a range of  $\pm \Delta/2$  around the radial position  $r_b$ .  $\rho_b$  describes the position in bin  $b$  and is defined on the interval  $\left[r_b - \frac{\Delta}{2}, r_b + \frac{\Delta}{2}\right]$ . Considering the atom coordinates  $\mathcal{R} = \{\mathbf{r}^{\tilde{N}}\}$  the intensity of the radial distribution function  $\tilde{g}_b$  in bin  $b$  can be calculated by counting the weighted number of distances  $r_{ij} = \|\mathbf{r}_{ij}\| = \|\mathbf{r}_i - \mathbf{r}_j\|$  between the individual particles  $i$  and  $j$ . Mathematically, this can be represented as an integration over all delta distributions  $\delta(\rho_b - r_{ij})$  positioned at each distance  $r_{ij}$  situated within the bin boundaries from  $\left(r_b - \frac{\Delta}{2}\right)$  to  $\left(r_b + \frac{\Delta}{2}\right)$ . Hence,  $\tilde{g}_b$  reads as

$$\tilde{g}_b = \frac{\tilde{V}}{\tilde{N}(\tilde{N} - 1)\delta V_b} \times \sum_{i=1}^{\tilde{N}} \sum_{j \neq i}^{\tilde{N}-1} \int_{(r_b - \frac{\Delta}{2})}^{(r_b + \frac{\Delta}{2})} d\rho_b \delta(\rho_b - r_{ij}) \quad (1)$$

Here,  $\tilde{V}$  is the system volume and

$$\delta V_b = \frac{4}{3}\pi \left[ \left( r_b + \frac{\Delta}{2} \right)^3 - \left( r_b - \frac{\Delta}{2} \right)^3 \right] \quad (2)$$

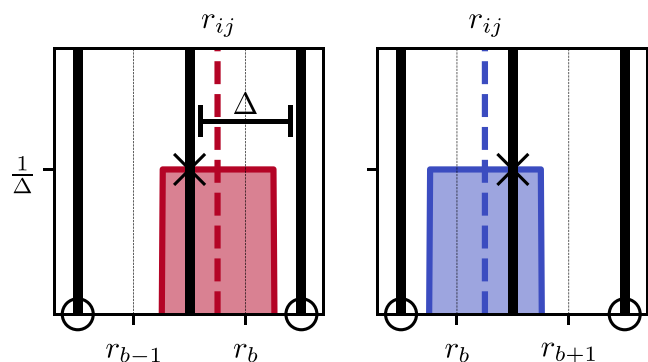
the volume of a spherical bin. If the RDF is determined in a MD simulation, the number of atoms in the simulation  $N$  and the volume  $V$  of the simulated system will be significantly smaller, i.e.,  $N \ll \tilde{N}$  and  $V \ll \tilde{V}$ , than in a real macroscopic system. To simulate the RDF of a macroscopic system, the instantaneous RDFs as determined from the MD simulation  $g_b$  are averaged  $\langle \cdot \rangle$  over several time steps to approximate the macroscopic ensemble in the ergodic limit

$$\begin{aligned} \tilde{g}_b &\approx \langle g_b \rangle \\ &= \left\langle \frac{V}{N(N-1)\delta V_b} \times \sum_{i=1}^N \sum_{j \neq i}^{N-1} \int_{(r_b - \frac{\Delta}{2})}^{(r_b + \frac{\Delta}{2})} d\rho_b \delta(\rho_b - r_{ij}) \right\rangle \end{aligned} \quad (3)$$

To use the RDF as a bias potential in MD simulations, it has to be differentiable with respect to the interatomic distance  $r_{ij}$ . The nondifferentiable  $\delta$ -distributions are thus replaced with differentiable kernel functions  $\Theta(\rho_b - r_{ij})$  as an approximation to the original RDF. The kernel function is centered around  $r_{ij}$  and extends over the entire radial bin range  $\pm \Delta/2$ . Furthermore, the integral over the kernel function has to be equal to one. In order to avoid any unjustified variations in the density of the contributions from  $r_{ij}$  to the RDF over the bin range  $\Delta$ , we choose a simple rectangular function as kernel function. Thus,  $\Theta(\rho_b - r_{ij})$  is defined as

$$\Theta(\rho_b - r_{ij}) = \begin{cases} \frac{1}{\Delta} & \text{if } \left| \frac{\rho_b - r_{ij}}{\Delta} \right| < \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Due to the extent  $\Delta$  of the kernel function, the contribution of  $r_{ij}$  to the histogram is spread over two adjacent bins, if  $r_{ij}$  is not positioned in the exact center  $r_b$  of the bin  $b$ . The position of the adjacent bin to which the kernel function contributes to is dependent on the position of  $r_{ij}$  in the bin. If  $r_b - \frac{\Delta}{2} < r_{ij} < r_b$  the kernel function contributes to the bins  $b - 1$  and  $b$ , while if  $r_b < r_{ij} < r_b + \frac{\Delta}{2}$  the kernel function contributes to the bins  $b$



**Figure 1.** Sketch of the contribution of each distance  $r_{ij}$  to the RDF  $g_b$ . The colored dashed lines represent the original delta distribution  $\delta(\rho_b - r_{ij})$  at  $r_{ij}$ . We replace the delta distributions with rectangular kernel functions  $\Theta(\rho_b - r_{ij})$  centered around  $r_{ij}$ . The contribution of the distance  $r_{ij}$  to the bin  $b$  is defined by the area under the rectangular function in the bin  $b$  (lighter color). Due to the extent of the rectangular kernel function, the distance  $r_{ij}$  contributes to the RDF intensity of two adjacent bins. If it is situated to the left of the center of the bin  $b$ , the affected bins are  $b - 1$  and  $b$  (red). If it is situated to the right of the center of bin  $b$ , the affected bins are  $b$  and  $b + 1$  (blue). The black cross indicates the nonzero value of  $\Theta\left(r_b - \frac{\Delta}{2} - r_{ij}\right) = \Theta\left(r_{b-1} + \frac{\Delta}{2} - r_{ij}\right) = \frac{1}{\Delta}$  (red) and  $\Theta\left(r_b + \frac{\Delta}{2} - r_{ij}\right) = \Theta\left(r_{b+1} - \frac{\Delta}{2} - r_{ij}\right) = \frac{1}{\Delta}$  (blue) at the bin boundaries, while the kernel function is zero at the other boundaries of the bins (black circle).

and  $b + 1$  (see Figure 1). After replacing the delta distributions with the rectangular functions eq 3 reads as

$$\begin{aligned} \tilde{g}_b &\approx \langle g_b \rangle \\ &= \left\langle \frac{V}{N(N-1)\delta V_b} \times \sum_{i=1}^N \sum_{j \neq i}^{N-1} \int_{(r_b - \frac{\Delta}{2})}^{(r_b + \frac{\Delta}{2})} d\rho_b \Theta(\rho_b - r_{ij}) \right\rangle \end{aligned} \quad (5)$$

where the approximation holds for sufficiently large sample sizes of the  $r_{ij}$  and small bin sizes  $\Delta$ .

Often a system contains several atom types. In this case, the total RDF can be described as the sum of the individual partial RDFs. Considering the atom types  $\alpha$  and  $\beta$ , the partial radial distribution function, which describes the radial distribution function restricted to the atom pairs of type  $\alpha$  and  $\beta$ , is written as  $g_b^{\alpha\beta}$ . To compute the total radial distribution function, the partial RDFs  $g_b^{\alpha\beta}$  have to be weighted by the relative frequency  $w^{\alpha\beta}$  of the  $\alpha$  and  $\beta$  pairs, resulting in

$$g_b = \sum_{\alpha\beta} w_{\alpha\beta} g_b^{\alpha\beta} \quad (6)$$

If the total radial distribution function is measured by WAXS, the atom type specific structure factor  $f$  has to be added to calculate the total RDF from its partial constituents. To mitigate the dependence of the structure factor from the length of the reciprocal scattering vector  $q$ , the Warren–Krutter–Morningstar approximation<sup>40</sup> is used throughout this work. The total RDF is then given by

$$g_b = \sum_{\alpha\beta} f_{\alpha} f_{\beta} w_{\alpha\beta} g_b^{\alpha\beta} \quad (7)$$

**2.2. Maximum Relative Entropy Bias.** Equilibrium MD simulations are designed to sample particle coordinates  $\mathcal{R}$  from the generalized Boltzmann distribution<sup>41</sup>

$$P_0(\mathcal{R}) = e^{-\beta(U(\mathcal{R}) - \sum_m X_m x_m)} \quad (8)$$

with the potential energy  $U(\mathcal{R})$ , the thermodynamic  $\beta = \frac{1}{k_B T}$ ,  $M$  generalized forces  $X_m$  (e.g., the pressure), and  $M$  conjugate generalized coordinates  $x_m$  (e.g., the volume). The goal of this work is to bias the original generalized Boltzmann distribution  $P_0$  of a MD simulation to obtain an updated ensemble distribution  $P$  that reproduces a target RDF

$$\langle g_b \rangle_P = \tilde{g}_b^{\text{target}} \quad (9)$$

in the MD simulation. An infinite set of probability distributions  $P$  exists that satisfy eq 9. However, the useful physical information they contain is markedly different. Thus, an additional requirement to decide for the best of the new distributions  $P$  is necessary. This requirement is chosen such that the average probability of encountering a specific atomic configuration  $\mathcal{R}$  under the new ensemble distribution  $P$  should be as similar as possible to encounter the same atomic configuration  $\mathcal{R}$  under the original distribution  $P_0$ . The relative entropy  $S[P||P_0]$ , also known as the Kullback–Leibler divergence, is the statistical distance between the distributions  $P$  relative to  $P_0$ , denoted as  $P||P_0$ , that quantifies this similarity requirement. It is defined as

$$\begin{aligned} S[P||P_0] &= - \int d\mathcal{R} P(\mathcal{R}) \ln \frac{P(\mathcal{R})}{P_0(\mathcal{R})} \\ &= - \int d\mathcal{R} P(\mathcal{R}) [\ln P(\mathcal{R}) - \ln P_0(\mathcal{R})] \end{aligned} \quad (10)$$

Therefore, the closer eq 10 is to zero, the smaller the discrepancy between the distributions of  $P$  and  $P_0$ . Consequently, the desired biased distribution can be identified by maximizing the relative entropy (because  $S \leq 0$ )<sup>30,42</sup>

$$P_{\text{ME}} = \arg \max_P S[P||P_0] \quad (11)$$

under the two constraints

$$\int d\mathcal{R} g_b(\mathcal{R}) P(\mathcal{R}) = \langle g_b \rangle_P = \tilde{g}_b^{\text{target}} \quad (12)$$

$$\int d\mathcal{R} P(\mathcal{R}) = 1 \quad (13)$$

Here,  $\int d\mathcal{R}$  implies an integral over the whole phase space. This approach is called the principle of maximum relative entropy. A review on biasing with the maximum relative entropy approach is provided by Cesari et al.<sup>31</sup> Solving the constrained optimization problem with the method of Lagrange multipliers leads to a posterior probability distribution that reads as

$$P_{\text{ME}} \propto e^{-\sum_{b=1}^B \lambda_b g_b} P_0 \quad (14)$$

where  $\lambda_b$  are the Lagrange parameters, that need to be determined. Thus, the generalized Boltzmann distribution is updated to be

$$\begin{aligned} P_{\text{ME}} &= e^{-\beta(U - \sum_m X_m x_m) + \sum_{b=1}^B \lambda_b g_b} \\ &= e^{-\beta(U - \sum_m X_m x_m + \sum_{b=1}^B \hat{\lambda}_b g_b)} \\ &= e^{-\beta(U - \sum_m X_m x_m + V)} \end{aligned} \quad (15)$$

with the linear bias potential

$$V = \sum_{b=1}^B \hat{\lambda}_b g_b \quad (16)$$

and the definition that

$$\hat{\lambda}_b = \frac{1}{\beta} \lambda_b \quad (17)$$

The similarity of the term  $\sum_m X_m x_m$  to  $\sum_{b=1}^B \hat{\lambda}_b g_b$  in eq 15 is due to the fact that thermodynamic systems reproduce an average of the generalized coordinates  $\langle x_m \rangle$  specified by  $X_m$  while maximizing the entropy of the distribution of generalized coordinates of the system. In this sense, the values  $g_b(\mathcal{R})$  can be compared to “generalized coordinates” and  $\hat{\lambda}_b$  to “generalized forces”.

In MD simulations, the bias potential  $V(\mathcal{R})$  adds a force  $\mathbf{F}_i$  to the atom  $i$ , which can be expressed as

$$\mathbf{F}_i = \nabla_i V(\mathcal{R}) = \sum_{b=1}^B \hat{\lambda}_b \nabla_i g_b(\mathcal{R}) \quad (18)$$

The derivative of the RDF contribution  $\nabla_i g_b(\mathcal{R})$  with respect to the coordinate  $\mathbf{r}_i$  of atom  $i$  is

$$\begin{aligned} \nabla_i g_b(\mathcal{R}) &= \frac{V}{N(N-1)\delta V_b} \times \sum_{j \neq i} \nabla_{\mathbf{r}_i} \int_{(r_b - \frac{\Delta}{2})}^{(r_b + \frac{\Delta}{2})} d\rho_b \Theta(\rho_b - r_{ij}) \\ &= \frac{V}{N(N-1)\delta V_b} \frac{\mathbf{r}_{ij}}{r_{ij}} \times \sum_{j \neq i}^{N-1} \left[ \Theta\left(r_b + \frac{\Delta}{2} - r_{ij}\right) - \Theta\left(r_b - \frac{\Delta}{2} - r_{ij}\right) \right] \end{aligned} \quad (19)$$

By inserting eq 19 into eq 18 and rearranging, we get

$$\mathbf{F}_i = \sum_{j \neq i}^{N-1} \mathbf{F}_{ij} = \sum_{j \neq i}^{N-1} \frac{V \hat{\mathbf{F}}_{ij}}{N(N-1)} \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (20)$$

with

$$\mathbf{F}_{ij} = \frac{V \hat{\mathbf{F}}_{ij}}{N(N-1)} \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (21)$$

and

$$\begin{aligned} \hat{\mathbf{F}}_{ij} &= \sum_{b=1}^B \hat{\mathbf{F}}_{ij}^b \\ &= \sum_{b=1}^B \frac{\hat{\lambda}_b}{\delta V_b} \left[ \Theta\left(r_b + \frac{\Delta}{2} - r_{ij}\right) - \Theta\left(r_b - \frac{\Delta}{2} - r_{ij}\right) \right] \end{aligned} \quad (22)$$

Note that  $\Theta\left(r_b - \frac{\Delta}{2} - r_{ij}\right)$  is the value of the kernel function centered at  $r_{ij}$  at the lower boundary of bin  $b$  and  $\Theta\left(r_b + \frac{\Delta}{2} - r_{ij}\right)$  the value of the same kernel function at the upper boundary of bin  $b$ . We can simplify eq 22 by considering that each kernel function  $\Theta$  around  $r_{ij}$  contributes to only two adjacent bins, the bin  $\hat{b}$  in which  $r_{ij}$  is located and a neighboring bin  $\hat{b}'$ . Thus, the kernel function crosses the bin boundaries only at one bin boundary position (see Figure 1 at the location of the black cross). Since eq 22 depends only on the values of the kernel function  $\Theta$  at the bin boundaries, only two terms, stemming from  $b = \hat{b}$  and  $b = \hat{b}'$  in the sum  $\sum_{b=1}^B$  over all bins in eq 22 are nonzero, or more precisely  $1/\Delta$  (black cross in Figure 1), and contribute to the term eq 22. At the position of all other bin boundaries, the kernel function is zero and does not contribute to the term eq 22 (black circles in Figure 1). Whether  $r_{ij}$  is positioned to the left or right of the bin center  $r_b$  determines whether the adjacent bin is  $\hat{b}' = \hat{b} - 1$  (red kernel function in Figure 1) or  $\hat{b}' = \hat{b} + 1$  (blue kernel function in Figure 1). So eq 22 can be simplified to

$$\hat{\mathbf{F}}_{ij} = \begin{cases} \frac{1}{\Delta} \left[ \frac{\hat{\lambda}_{\hat{b}}}{\delta V_{\hat{b}}} - \frac{\hat{\lambda}_{\hat{b}-1}}{\delta V_{\hat{b}-1}} \right] & \text{if } r_{\hat{b}} - \frac{\Delta}{2} < r_{ij} < r_{\hat{b}} \\ \frac{1}{\Delta} \left[ \frac{\hat{\lambda}_{\hat{b}+1}}{\delta V_{\hat{b}+1}} - \frac{\hat{\lambda}_{\hat{b}}}{\delta V_{\hat{b}}} \right] & \text{if } r_{\hat{b}} < r_{ij} < r_{\hat{b}} + \frac{\Delta}{2} \end{cases} \quad (23)$$

If partial RDF compositions (eq 6) or RDFs measured by X-ray scattering (eq 7) are considered, the weights of the partial RDFs have to be included in the forces and eq 21 is altered to

$$\mathbf{F}_{ij} = w_{\alpha\beta} \frac{V \hat{\mathbf{F}}_{ij}}{N(N-1)} \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (24)$$

or

$$\mathbf{F}_{ij} = f_{\alpha\beta} w_{\alpha\beta} \frac{V \hat{\mathbf{F}}_{ij}}{N(N-1)} \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (25)$$

respectively. Given that the bias force is generated by a pairwise interaction, the implementation of the bias is analogous in computational complexity to the incorporation of an additional pairwise potential into the MD simulation. Additionally, the bias will contribute to the original virial pressure of the system,  $T_0$ , by adding a term,  $T_{\text{ME}}$ .

Consequently, the new total virial pressure of the system,  $T_{\text{total}}$ , is given by

$$T_{\text{total}} = T_0 + T_{\text{ME}} \quad (26)$$

The virial resulting from the bias can be calculated as<sup>43</sup>

$$T_{\text{ME}} = \frac{1}{Vd} \sum_i^N \sum_{j \neq i}^{N-1} \mathbf{F}_{ij} \cdot \mathbf{r}_{ij} \quad (27)$$

where  $d$  is the dimensionality of the system, e.g., three for a three-dimensional system. By substituting eqs 21 and 22, 27 can be transformed to

$$T_{\text{ME}} = \sum_{b=1}^B T_{\text{ME}}^b \quad (28)$$

with

$$T_{\text{ME}}^b = \frac{1}{dN(N-1)} \sum_i^N \sum_{j \neq i}^{N-1} \hat{F}_{ij}^b r_{ij} \quad (29)$$

where  $T_{\text{ME}}^b$  represents a measure for the bias-induced virial per bin  $b$ .

**2.3. Determination of Lagrange Multiplier  $\lambda_b$ .** To bias the original probability distribution  $P_0(\mathcal{R})$ , the parameters  $\lambda_b$  need to be determined. Following Pitera and Chodera<sup>30</sup> the function

$$\Gamma(\lambda) = \ln \left[ \int d\mathcal{R} P_0(\mathcal{R}) e^{-\sum_{b=1}^B \lambda_b g_b(\mathcal{R})} \right] + \sum_{b=1}^B \lambda_b \tilde{g}_b^{\text{target}} \quad (30)$$

is introduced, which combines eq 11 with eqs 12 and 13 by a Legendre transformation.<sup>44</sup> The set of parameters  $\lambda$  that solve the combined eqs 11–13 can be found by minimizing eq 30.

In this study, we use gradient descent with a step size of  $\gamma$  to iteratively update the Lagrange multipliers during the simulation to minimize  $\Gamma(\lambda)$ . The gradient of  $\Gamma(\lambda)$  with respect to  $\lambda_b$  is defined as

$$\begin{aligned} \frac{\partial \Gamma}{\partial \lambda_b} &= \tilde{g}_b^{\text{target}} - \frac{\int d\mathcal{R} P_0(\mathcal{R}) e^{-\sum_{b=1}^B \lambda_b g_b(\mathcal{R})} g_b(\mathcal{R})}{\int d\mathcal{R} P_0(\mathcal{R}) e^{-\sum_{b=1}^B \lambda_b g_b(\mathcal{R})}} \\ &= \tilde{g}_b^{\text{target}} - \langle g_b \rangle \end{aligned} \quad (31)$$

and thus can be determined by calculating the difference of the target RDF  $\tilde{g}_b^{\text{target}}$  at bin  $b$  to the average RDF  $\langle g_b \rangle$  at bin  $b$  measured in the (biased) simulation.

In order for gradient descent to consistently identify an optimal set of Lagrange parameters, it is imperative that the optimization does not become trapped in local minima of  $\Gamma(\lambda)$ . This can only be guaranteed if  $\Gamma(\lambda)$  is convex. That is to say, the Hessian matrix of  $\Gamma(\lambda)$  must be positive-definite. The Hessian matrix of  $\Gamma(\lambda)$  is<sup>30</sup>

$$\frac{\partial^2 \Gamma}{\partial \lambda_b \partial \lambda_{b'}} = \langle g_b g_{b'} \rangle - \langle g_b \rangle \langle g_{b'} \rangle \quad (32)$$

eq 32 demonstrates that the positive-definiteness of  $\Gamma$  necessitates the individual bins of the RDFs to be statistically uncorrelated; however, given that the intensities in the RDF bins are typically statistically correlated, the function  $\Gamma(\lambda)$  is not convex. Consequently, gradient descent may not identify an optimal set of parameters  $\lambda$  but rather converge to a local minimum. This suggests that a set of parameters  $\lambda$  determined

by gradient descent may represent a local rather than a global minimum solution of the objective function  $\Gamma(\lambda)$ .

As discussed at the end of Section 2.2, the bias force leads to a change in the virial of the system (eq 27). Depending on the target RDF, this can lead to a significant change in volume in systems simulated as isobaric ensembles. This ultimately impedes the induction of phase transitions in the simulations. To counteract an undesired large change in volume, we introduce an optional scaling parameter  $\kappa_b$  into the gradient descent algorithm. The scaling parameter  $\kappa_b$  adjusts the update of the Lagrange parameters  $\lambda_b$  based on the average contribution of the induced virial per bin  $\langle T_{\text{ME}}^b \rangle$ . To counteract excessive contraction or expansion of the system, we reduce the update rate by setting  $\kappa_b < 1.0$  for bins  $b$  for which the average virial per bin  $\langle T_{\text{ME}}^b \rangle$  contributes negatively or positively to the total virial, respectively. The update rates for the other bins remain unchanged with  $\kappa_b = 1.0$ . With the additional parameter  $\kappa_b$  we observed satisfactory results in minimizing  $\Gamma(\lambda)$  with gradient descent for all considered systems.

As the discrepancy between the measured RDF  $\langle g_b \rangle$  and the target RDF  $\tilde{g}_b^{\text{target}}$  is reduced during the gradient descent process, the adjustments to the parameters  $\lambda_b$  per update cycle would become increasingly small. This would result in a deceleration of the convergence of the mean RDF  $\langle g_b \rangle$  toward the target RDF  $\tilde{g}_b^{\text{target}}$ . Consequently, the update is normalized by the sum of the absolute values of the RDF differences across all bins  $\sum_{b=1}^B |\tilde{g}_b^{\text{target}} - \langle g_b \rangle|$ . This normalization guarantees that the updates of the Lagrange multipliers  $\lambda_b$  remain constant, irrespective of the proximity of the measured RDF  $\langle g_b \rangle$  to the target RDF  $\tilde{g}_b^{\text{target}}$ . This facilitates the selection of a meaningful update parameter  $\gamma$ , that can be maintained constant throughout the simulation. Hence, the update rule for the parameters  $\lambda_b$  is calculated as

$$\begin{aligned} \lambda_b' &= \lambda_b + \gamma \kappa_b \frac{\frac{\partial \Gamma}{\partial \lambda_b}}{\sum_{b=1}^B \left| \frac{\partial \Gamma}{\partial \lambda_b} \right|} \\ &= \lambda_b + \gamma \kappa_b \frac{\tilde{g}_b^{\text{target}} - \langle g_b \rangle}{\sum_{b=1}^B |\tilde{g}_b^{\text{target}} - \langle g_b \rangle|} \end{aligned} \quad (33)$$

The magnitude of change in the Lagrange parameters in each update cycle is determined by the selection of the update parameter  $\gamma$ . An increase in the update parameter results in a larger adjustment of the Lagrange parameters  $\lambda_b$  and subsequently requires a smaller number of updates for  $\langle g_b \rangle$  to converge to its target value  $\tilde{g}_b^{\text{target}}$ . However, simulations that employed excessive update parameters  $\gamma$  exhibited instability due to abrupt and substantial alterations in forces. Furthermore, to induce liquid-to-solid nucleation, it is crucial to avoid setting excessively high values, as the formation of a nucleation seed remains a stochastic process, albeit with a noticeably increased probability. In our experimental investigations, we found that values of  $\gamma$  ranging from 10 to 5 and from 5 to 1 were effective in reproducing RDFs in liquids or in inducing liquid-to-solid transitions, respectively.

In addition to the update parameter  $\gamma$ , the number of bins  $B$ , or equivalently the size of the bins  $\Delta$ , is a parameter that must be provided in the presented algorithm. This determines the number of Lagrange parameters  $\lambda_b$  that need to be found through the gradient descent method. While it is not feasible to establish a universal guideline for the binning of the target RDF, the following factors should be taken into account when

determining a bin size: The binning should be sufficiently small to capture the essential features of the RDF being reproduced, while avoiding bins that are excessively small. The latter results in insufficient statistics to determine the average number of atoms in the bins. This, in turn, leads to noisy parameter updates and consequently large biasing forces, resulting in instabilities of the simulation.

**2.4. Potential of Mean Force.** The free energy or potential of mean force as a function of the instantaneous RDF at all bins obtained in the MD simulation  $\mathbf{g}$  forms a hypersurface

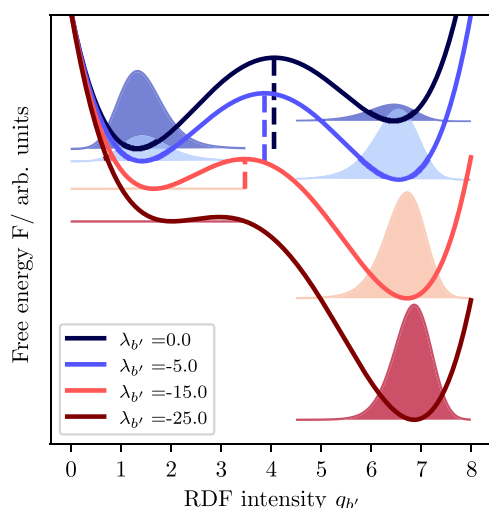
$$F_0(\mathbf{g}) = -\frac{1}{\beta} \log \int d\mathcal{R} \delta(\mathbf{g}'(\mathcal{R}) - \mathbf{g}) P_0(\mathcal{R}) \quad (34)$$

with as many dimensions as there are biased bins  $B$ . This hypersurface is tilted by the applied linear bias

$$\begin{aligned} F(\mathbf{g}) &= -\frac{1}{\beta} \log \int d\mathcal{R} \delta(\mathbf{g}'(\mathcal{R}) - \mathbf{g}) P_{\text{ME}}(\mathcal{R}) \\ &= -\frac{1}{\beta} \log \int d\mathcal{R} \delta(\mathbf{g}'(\mathcal{R}) - \mathbf{g}) P_0(\mathcal{R}) e^{\lambda \mathbf{g}'(\mathcal{R})} \\ &= F_0(\mathbf{g}) + k_B T \lambda \mathbf{g} \end{aligned} \quad (35)$$

as discussed in Cesari et al.<sup>31</sup>

Figure 2 shows the hypothetical free energy surface for an exemplary one-dimensional case of a single bin  $b'$ . The



**Figure 2.** Hypothetical free energy surface as a function of the RDF intensity  $g_{b'}$  at one specific bin position  $r_{b'}$ . Two local free energy minima are visible that are separated by a free energy barrier. The resulting Boltzmann probability distribution is indicated by slightly lighter distributions in the energy minima. An increase of the bias is expressed by the decrease of the Lagrange multiplier  $\lambda_{b'}$  from  $\lambda_{b'} = 0.0$  (dark blue) to  $\lambda_{b'} = -25.0$  (dark red).

unbiased system  $\lambda_{b'} = 0.0$  (dark blue) has two local free energy minima. The state at  $g_{b'} \approx 1.5$  is the lowest free energy state and thus stable, while the state at about  $g_{b'} \approx 6.5$  is a metastable state of higher energy that is separated from the stable state by a free energy barrier indicated by the dashed line. The resulting Boltzmann probability distribution is indicated by slightly lighter distributions in the energy minima. The stepwise decrease of the Lagrange multiplier down to  $\lambda_{b'} = -25.0$  (from dark blue to dark red) causes the free energy profile to tilt downward with increasing intensity of  $g_{b'}$ . The

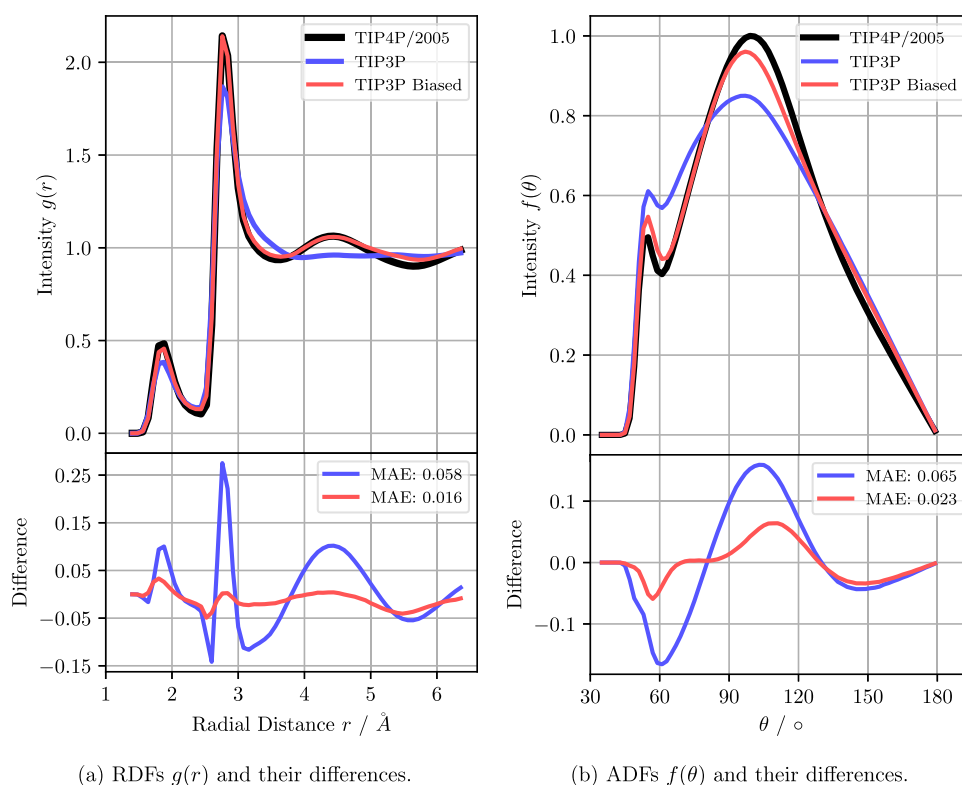
induced tilt has two important effects on the free energy surface. First, the free energy of the metastable state is reduced to the point where it becomes the new minimum energy state of the system (in Figure 2 the new global minimum changes from  $g_{b'} \approx 1.5$  to  $g_{b'} \approx 7$ ). Thus, the former metastable state becomes the new stable state and vice versa. As a result, crystalline phases may become stable in the simulation at temperatures and/or pressures where the phase was not stable prior to the induced bias. Second, the free energy barrier is lowered. Reducing the barrier increases the probability of a state transition up to the point where, by increasing the bias, the free energy barrier disappears and a state transition is certain to occur. On multidimensional free energy surfaces with multiple bins  $b$ , the direction of tilt of the free energy surface determines the amount of energy reduction of the target state and the barrier. The tilt angle between the different dimensions defined by the bins  $b$  in the presented algorithm is implicitly chosen by the step size  $\gamma$  and the virial weighting  $\kappa_b$  at which each individual Lagrange multiplier  $\lambda_b$  is updated.

### 3. RESULTS AND DISCUSSION

We implemented the maximum relative entropy formalism as an extension to the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS).<sup>45</sup> To incorporate the atomic structure factor dependence of the RDFs, we build on the code of Coleman et al.,<sup>46</sup> which in turn uses an analytical approximation of the atomic structure factors<sup>47</sup> parametrized by Peng et al.<sup>48</sup> We then evaluated the implemented algorithm in two different applications. In the first application, we biased a liquid water system modeled with the TIP3P<sup>49</sup> model to reproduce the RDF of another system modeled with the TIP4P/2005<sup>50</sup> water model. In the second application, we investigated the potential of the algorithm to moderate liquid–solid phase transitions. Therefore, we biased a liquid water system to reproduce the hexagonal phase  $I_h$  and a liquid  $\text{TiO}_2$  system to reproduce the stable polymorphs rutile or anatase depending on of the applied target RDF. All examples presented were simulated under the assumption that the RDFs were generated from WAXS measurements. This means that atomic form factors had to be used to weight the partial RDF contributions in the RDF and the resulting forces (eq 7 and eq 25). We chose WAXS target RDFs because they are the most complex to reproduce with the additional weighting factors. Therefore, we expect the formalism to work similarly for target RDFs generated by identical particles or partial RDF compositions. In the following sections, we will present the simulated RDFs  $\langle g_b \rangle$  with small bin sizes  $\Delta$ . Thus, the bin positions  $r_b$  and the RDFs  $\langle g_b \rangle$  are represented as continuous  $\{r_1, \dots, r_B\} \approx r$  and  $\{\langle g_1 \rangle, \dots, \langle g_B \rangle\} \approx g(r)$  for clarity.

**3.1. Biasing of Liquid Systems: Water.** A wide variety of atomic interaction potentials have been developed to reproduce different properties of water, each with a different level of computational effort.<sup>51</sup> Among other properties, MD simulations with different water models tend to have different radial distribution functions. Here, we demonstrate that our method can be used to fit the RDF of a liquid system modeled with one atomic interaction potential to the RDF of a liquid system modeled with a different atomic interaction potential. As a demonstration, we biased the atomic interaction potential of TIP3P<sup>49</sup> to reproduce the RDF of the TIP4P/2005<sup>50</sup> model.

The TIP3P water model is a three-site model where the water molecule is simulated as three individual atoms. In contrast, the TIP4P/2005 water model is a four-site model that



**Figure 3.** Upper panel: RDF (left) and ADF (right) of liquid water systems modeled with different interatomic potentials. Black is the RDF/ADF of the TIP4P/2005 water model from which the target RDF was extracted. The RDF/ADF of the unbiased TIP3P water model is depicted in blue, while the biased TIP3P water model is depicted in red. Bottom panel: Difference in the RDF/ADF between the target TIP4P/2005 water model and the unbiased (blue) and biased (red) TIP3P water model. The two respective mean absolute errors (MAE) between the functions are given in the plot legend, with smaller values indicating better agreement.

adds an additional dummy atom to the water molecule. This dummy atom has no mass, but carries a charge, allowing for a more accurate representation of the charge distribution among water molecules. This makes the TIP4P/2005 model more suitable for reproducing a variety of bulk water properties, including density, viscosity, and diffusion coefficient, as well as the local water structuring.<sup>52</sup> However, this advantage is offset by the increased computational complexity.

The goal of this demonstration is to use the maximum entropy algorithm to improve the agreement between the TIP3P and TIP4P/2005 models by biasing the TIP3P model. For this purpose, we simulated a liquid water system with the TIP4P/2005 model at a temperature of 300 K and a pressure of 1 atm. From the equilibrated system, we calculated a target RDF  $g_b^{\text{target}}$  ranging from 0 to 8  $\text{\AA}$  with a bin size of 0.08  $\text{\AA}$ . This target RDF was then used to bias a TIP3P equilibrium system according to the maximum relative entropy formalism. The RDF  $\langle g_b \rangle$  was calculated every ten time steps and averaged over 100 time steps to conduct the update of the Lagrange multiplier  $\lambda_b$  with a step size of  $\gamma = 10.0$  and no virial specific weighting, i.e.,  $\kappa_b = 1$  for all bins. The incorporation of the bias resulted in a reduction of the simulation's computational speed from 3.2 simulated nanoseconds per day for the unbiased TIP3P water model to 2.7 simulated nanoseconds per day. Therefore, this modification caused a decrease of approximately 16% in the simulation's computational speed.

Figure 3a shows the resulting RDF from the TIP3P model before (blue) and after (red) the application of the bias compared to the target RDF from the TIP4P/2005 model (black) in the upper part of the figure. The lower part of the

figure depicts the difference between the TIP4P/2005 target RDF and the unbiased and biased RDFs of the TIP3P model in blue and red, respectively. After applying the bias, the initial mean absolute error (MAE) between the RDFs was reduced from 0.058 to 0.016.

In addition to the effect of the bias on the generated RDFs, we also investigated the influence of the introduced bias on another characteristic of the model unrelated to the RDF, i.e., the angular distribution function (ADF)  $f(\theta)$  of the oxygen atoms within a distance of 0  $\text{\AA}$  to 3.4  $\text{\AA}$  over an angle  $\theta$  of 180 $^\circ$  in 2 $^\circ$  bins, as shown in Figure 3b. Similar to Figure 3a, in the upper part of the figure the ADF of the target model TIP4P/2005 is shown in black, the unbiased TIP3P in blue and the biased TIP3P in red. In the lower part of the figure the difference between the ADF of the TIP4P/2005 model and the unbiased and biased TIP3P models is shown in blue and red, respectively. The ADF of the TIP3P model shows a clear change with the introduction of the bias. Similar to the RDF generated by the original TIP3P model, the unbiased ADF was less structured than the ADF generated by the TIP4P/2005 target model. With the application of the bias, the difference between the TIP4P/2005 and the biased TIP3P ADF decreased from a MAE of 0.065 to 0.023. In addition to the RDF and ADF, which are structural properties, the diffusion coefficients  $D$  were determined from the mean squared distance of the oxygen atoms in the water molecules (see SI for more information). Table 1 contains the measured values of the diffusion coefficients for the TIP4P/2005 water system, the unbiased TIP3P water system, and the biased TIP3P water system. The results indicate that water modeled with the

**Table 1. Diffusion Coefficients  $D$  of Oxygen Atoms Bound in the Water Molecules for the Considered Water Models at 300 K and 1 atm**

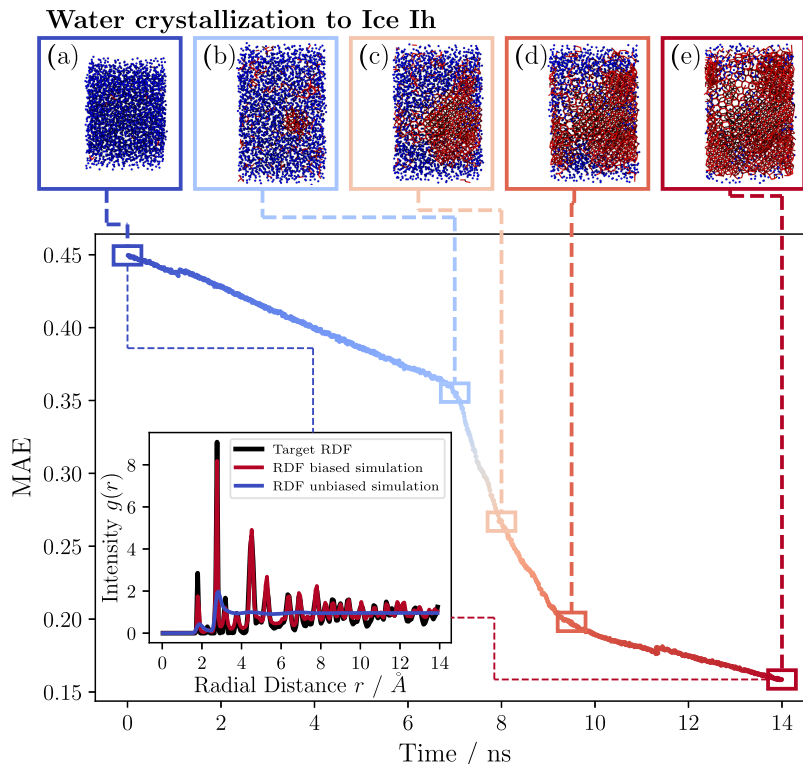
water model	diffusion coeff. $D/\text{m}^2 \text{s}^{-1}$
TIP4P/2005	$2.40 \times 10^{-9}$
TIP3P	$1.28 \times 10^{-9}$
Biased TIP3P	$0.31 \times 10^{-9}$

TIP4P/2005 model exhibits increased diffusivity compared to water modeled with TIP3P. The diffusion coefficient determined for the TIP4P/2005 model lies close to the diffusion coefficient of  $2.1 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$  reported by Markland et al.<sup>53</sup> However, the determined diffusion coefficient for the TIP3P model deviates from  $\sim 5.8 \times 10^{-9} \text{ m}^2$  reported by Mark and Nilsson.<sup>54</sup> The reason for this discrepancy is not fully understood, but it may be related to the strong system-property dependence of the diffusion coefficient, such as the assumed water density. Since Mark and Nilsson<sup>54</sup> determined the TIP3P diffusion coefficients from constant volume and energy (*NVE*) ensemble simulations, the properties may not be directly comparable to the constant pressure and temperature (*NpT*) ensemble simulations under consideration here. Therefore, the diffusion coefficients determined here will be used moving forward. To more accurately represent the TIP4P/2005 model, the diffusion coefficient would need to be increased; however, the applied bias results in a decrease of diffusivity to approximately one-fourth of the original diffusivity. This reduction in diffusivity is likely attributable

to the additional constraints imposed on the dynamics of the water molecules to reproduce the target RDF.

While the scientific value of biasing the TIP3P liquid model with data from RDFs obtained from the TIP4P/2005 model may be limited, a potential application could be to bias other atomic interaction models to structurally align more closely with RDFs derived from experiments. Furthermore, it could be beneficial to align simpler, more lightweight models to reproduce structural properties of complex, computationally intensive models to obtain structures for the training of machine learning potentials. In the context of water, for instance, the training of coarse-grained models has been demonstrated to be of interest for studying water.<sup>55,56</sup> However, it is imperative to acknowledge that this approach may result in a diminution of atomic mobility, which renders the biased model unfit to investigate dynamic properties of a system.

**3.2. Crystallization of Water.** Homogeneous crystallization of water is a process of great interest, but is challenging to study with MD simulations. As predicted by classical nucleation theory (CNT), homogeneous crystallization requires the formation of a nucleation seed. The nucleation seed creates a surface against the surrounding liquid. This surface makes the existence of the nucleation seed energetically unfavorable for small seed volumes. Therefore, in order for homogeneous crystallization to occur, the system has to overcome a free energy barrier. Consequently, the formation of a nucleus of sufficient size happens with low probability in unbiased MD simulations. This makes the study of



**Figure 4.** Bias-induced liquid–solid transition of a liquid water system modeled with TIP4P/ICE at 360 K and 1 atm. The mean absolute error (MAE) between the RDF of the biased water system and the target RDF of the simulated hexagonal ice  $I_h$  is depicted. Several snapshots of the biased simulation show the state of the oxygen atoms in the system from left to right at 0, 7, 8, 9.5, and 14 ns. The atoms in the liquid and hexagonal ice phase are colored blue and red, respectively. The inset shows the RDF of the unbiased liquid simulation (blue) and the final state of the system (red) compared to the target RDF (black).

crystallization processes difficult and often computationally infeasible.

In this section, we show that the maximum relative entropy formalism can be used to mediate a crystallization of the liquid water system into the hexagonal ice phase  $I_h$ . To this end, we simulated an isothermal–isobaric system of water molecules in the hexagonal ice phase  $I_h$  at 10 K and 1 atm. The TIP4P/ICE model was used as water potential.<sup>57</sup> Based on this system, the target RDF  $\tilde{g}_b^{\text{target}}$  was calculated from 0 to 14 Å with a bin size of 0.07 Å. By using the computed RDF as input to the maximum relative entropy algorithm, a simulation of liquid water at 360 K and 1 atm was biased.

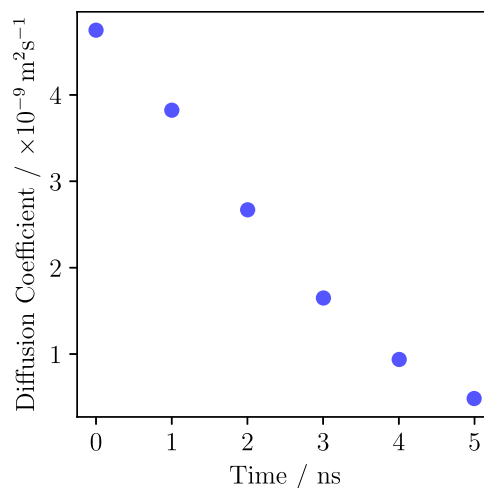
A temperature above the freezing point of water at 273 K was chosen to ensure sufficient mobility of water molecules in the system. Nevertheless, a phase transition to the hexagonal ice phase was possible in the simulation, since the free energy of the crystalline state is lowered by the induced tilt of the free energy surface (see Sec. 2.4). The Lagrange parameters were updated every 1100 time steps with a step size of 1 fs. The step size was  $\gamma = 2.5$  and no virial weighting ( $\kappa = 1.0$ ) was applied. The RDF of the system was computed every 10 time steps and averaged over 1000 time steps to determine the difference between the target RDF and the instantaneous system RDF for the gradient descent update. After the bias update, the system was equilibrated for 100 time steps before next samples of the instantaneous RDF  $g_b$  were taken from the simulation. The incorporation of the bias resulted in a reduction of the simulation's computational speed from 3.2 simulated nanoseconds per day for the unbiased TIP4P/ICE water model to 1.7 simulated nanoseconds per day. Therefore, this modification caused a decrease of approximately 47% in the simulation's computational speed. The substantial additional computation time requirement is likely associated with the extensive RDF range of 14 Å.

After a certain amount of induced bias strength, a gradual increase in temperature in the system was observed during the simulation as the bias was increased further. This rise in temperature is attributed to the sharp peaks of crystalline RDFs, such as those observed for water ice at 10 K. These sharp peaks lead to significant differences in Lagrange parameters  $\lambda_b$  of adjacent bins. Consequently, strong forces, that are not adequately damped by the thermostat, are created by the applied bias (refer to eq 23). These forces cause a rise in temperature. We were able to mitigate this effect by increasing the damping parameter of the applied thermostat or by decreasing the time steps. However, for moderate temperature increases, we did not observe any problems with reproducing the target RDF.

The upper part of Figure 4 shows snapshots of the oxygen atoms contained in the water molecules generated with VMD<sup>58</sup> during the simulation at time steps 0 ns (a), 7 ns (b), 8 ns (c), 9.5 ns (d), and 14 ns (e). The oxygen atoms are colored in dependence of a calculated average local bond order parameter  $\bar{q}_6$ ,<sup>59</sup> which is a modification of the local bond order parameter,<sup>60</sup> with a cutoff value of 3.5 Å. Blue oxygen atoms have an average local bond order parameter  $\bar{q}_6 \leq 0.07$ , while red oxygen atoms are  $\bar{q}_6 > 0.07$  (see SI for more information on bond order parameters).

At the beginning of the simulation, the dissimilarity of the RDF of the liquid water and the hexagonal ice system was relatively large, with an MAE of about 0.45. The dissimilarity is also evident from the inset, which shows the RDF of the unbiased liquid system (blue) and the target RDF of the

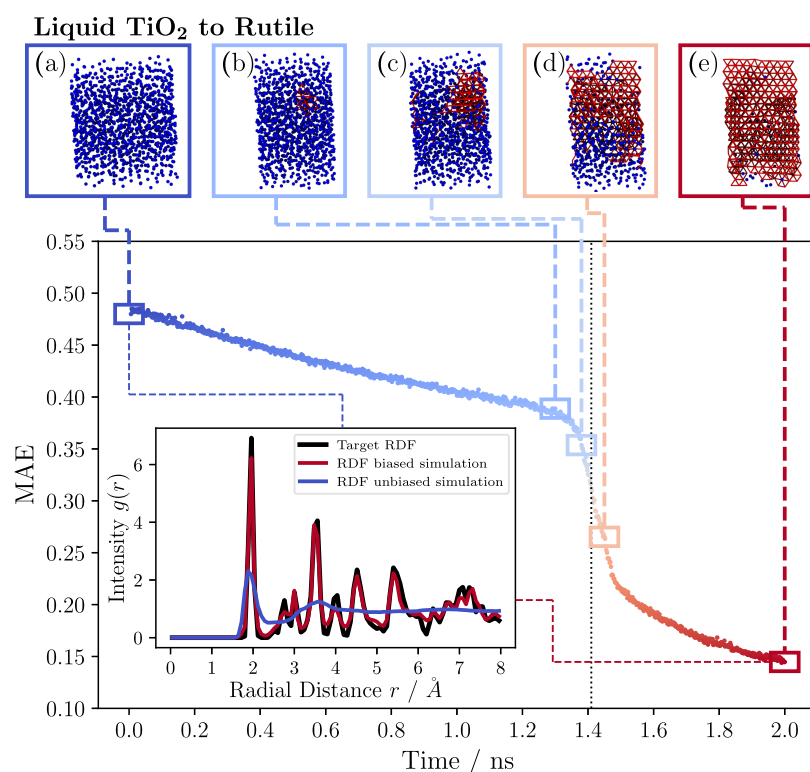
hexagonal ice (black). During the initial time period from 0 ns (a) to 7 ns (b), the biased simulation exhibited a decrease in density with increasing bias strength. At the same time, the MAE between the target RDF and the RDF originating from the biased system decreased linearly. The diffusion coefficients of oxygen atoms in water molecules were measured for the water system prior to the water-ice phase transition at approximately 1, 2, 3, 4, and 5 ns. The result is depicted in Figure 5. It is evident from the figure that as the bias strength



**Figure 5.** Diffusion coefficients of the oxygen atoms bound in the water molecules in dependence of the applied bias at different points in time before the water-ice phase transition occurred.

increased, the diffusivity of the water molecules decreased due to the increasing constraint to the water structure. After approximately 7 ns (b), a nucleation seed formed in the biased system. This nucleation seed grew in the following simulation steps. After approximately 10 ns, the whole system transitioned into a crystalline phase of  $I_h$  ice (e). The phase transition is visible as a steep drop in MAE between the target and instantaneous RDF. The presence of the  $I_h$  ice phase is corroborated by the similarity between the measured RDF and the target RDF of the biased system at the end of the simulation, as demonstrated in the inset. During the phase transition, the system did not form a single hexagonal ice crystal; rather, it formed a composite of crystal domains with varying crystal plane orientations. The thermal stability of the resulting ice structure was assessed by applying the unbiased TIP4P/ICE force field to the ice structure formed after 10 ns of simulation. To this end, a  $NpT$  ensemble of 260 K and 1 atm was simulated for approximately 3 ns. During the simulation, the ice system was observed to melt into a liquid system. The results of this study indicate that the crystal domains formed during the phase transition were inadequate in size to surmount the free energy barrier and establish a stable crystallization seed. The underlying reason for the formation of ice crystals with domains instead of a single crystal under the application of bias remains unclear and is a subject to be explored in subsequent research.

**3.3. Polymorphs of  $\text{TiO}_2$ : Rutile and Anatase.**  $\text{TiO}_2$  is an important material with various applications, such as photocatalytic water splitting,<sup>61,62</sup> photodynamic therapy for cancer treatment, inactivation of antibiotic resistant bacteria,<sup>63</sup> and gas sensing,<sup>64</sup> to name a few examples.



**Figure 6.** Crystallization of liquid  $\text{TiO}_2$  to rutile at 2500 K and 1 atm. The mean absolute error (MAE) between the RDF of the biased  $\text{TiO}_2$  system and the target RDF of the simulated rutile is depicted. The time at which the crystallization seed formed a critical nucleus is marked by the vertical dotted black line. Several snapshots of the biased simulation, show the state of the titanium atoms in the system from left to right at 0, 1.3, 1.38, 1.45, and 2 ns. The atoms in the liquid and rutile phases are colored blue and red, respectively. The inset shows the RDF of the unbiased liquid simulation (blue) and the final state of the system (red) against the target RDF (black).

At atmospheric pressure, there are three different polymorphs of  $\text{TiO}_2$ , namely rutile, anatase, and brookite, with rutile and anatase being the most important polymorphs for applications. While rutile is the lowest energy configuration under ambient conditions, anatase and brookite are metastable polymorphs.

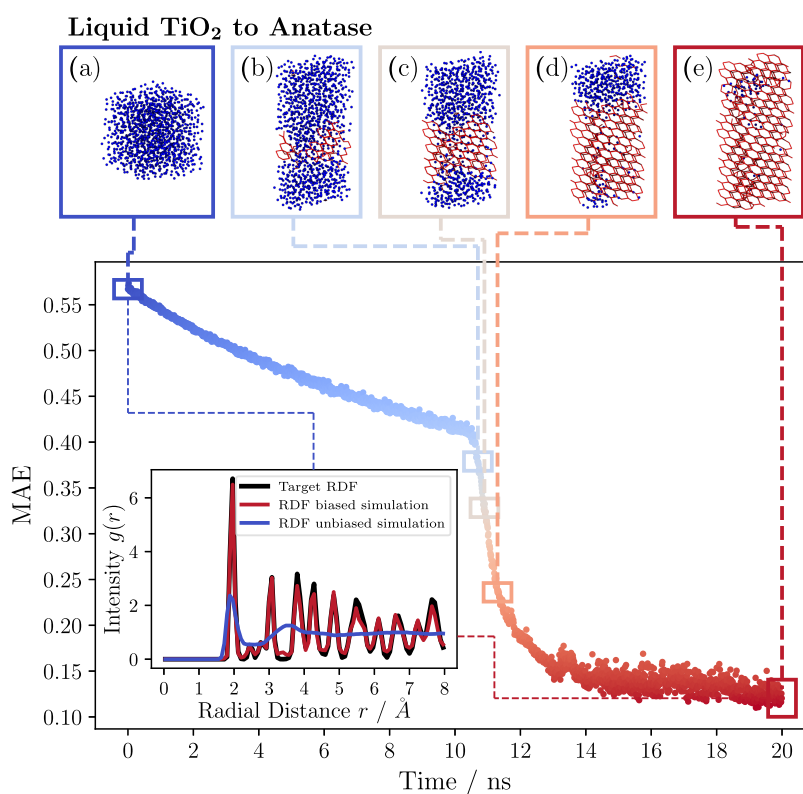
In this section, we demonstrate the crystallization of liquid  $\text{TiO}_2$  into the corresponding stable rutile or metastable anatase polymorph by biasing a liquid  $\text{TiO}_2$  system with the RDF extracted from a rutile or anatase system. In particular, the crystallization into the metastable anatase phase would not be feasible using conventional methods developed to find the configuration of the global energy minimum, such as simulated annealing or replica exchange methods.

To bias the liquid  $\text{TiO}_2$  system, two target RDFs were created for rutile and anatase, ranging from 0 to 8 Å with a bin size of 0.08 Å. For this purpose, rutile and anatase were modeled as an isothermal–isobaric ensemble at a temperature of 300 K and a pressure of 1 atm with the atomic interaction potential developed by Matsui and Akaogi<sup>65</sup> and provided by the Knowledgebase of Interatomic Models (KIM).<sup>66</sup> The calculated RDFs were used as inputs to the maximum relative entropy algorithm to bias the two identical initial systems of liquid  $\text{TiO}_2$  at 2500 K and 1 atm to reproduce the RDF of either rutile or anatase. A time step size of 1 fs was set for both simulations. The Lagrange parameters were updated every 1000 time steps with a step size of  $\gamma = 5.0$  for rutile and  $\gamma = 1.25$  for anatase.

Similar to the crystallization of water, we observed an increase in temperature at high bias forces. It is possible to

mitigate the temperature increase by increasing the thermostat damping or by using smaller simulation time steps. However, in the present case, the temperature increase did not have a negative effect on the simulations. To counteract the excessive contraction of the system due to the virial contribution  $T_{\text{ME}}$  of the bias (eq 27), the update rate of all bins contributing negatively to the total virial contribution was reduced by a factor of  $\kappa_b = 0.25$ . The system RDF was computed at 10 time step intervals and averaged after 1000 time steps to calculate the difference between the target RDF and the instantaneous system RDF for the gradient descent update. It was determined that an equilibration period following the Lagrangian parameter update was unnecessary, as the change in bias potential was sufficiently small in each iteration. The incorporation of the bias resulted in a reduction of the simulation's computational speed from 8.7 simulated nanoseconds per day for the unbiased  $\text{TiO}_2$  model to 6.9 simulated nanoseconds per day. Therefore, this modification caused a decrease of approximately 21% in the simulation's computational speed.

Figures 6 and 7 show the results for the simulations biased with the RDF of rutile and anatase, respectively. In both figures, snapshots of the Ti atoms in the biased  $\text{TiO}_2$  systems are depicted in the upper part of the figures. The titanium atoms are colored as a function of the averaged local bond order parameter values. In Figure 6, blue atoms in the liquid phase correspond to an averaged local bond order parameter  $\bar{q}_8$ , calculated over the 12 nearest neighbors, with a value  $\leq 0.073$ , while red atoms in the rutile phase have a value  $>0.073$  (see SI for more information). Similarly, in Figure 7, blue



**Figure 7.** Crystallization of liquid  $\text{TiO}_2$  to anatase at 2500 K and 1 atm. The mean absolute error (MAE) between the RDF of the biased  $\text{TiO}_2$  system and the target RDF of the simulated anatase is depicted. Several snapshots of the biased simulation, show the state of the titanium atoms in the system from left to right at 0, 10.7, 10.9, 11.3, and 20 ns. The atoms in the liquid and anatase phases are colored blue and red, respectively. The inset shows the RDF of the unbiased liquid simulation (blue) and the final state of the system (red) against the target RDF (black).

atoms correspond to Ti atoms in the liquid phase with an averaged local bond order parameter of degree  $\bar{q}_{10}$ , calculated over the 12 nearest neighbors, with values  $\leq 0.085$ . Red atoms with  $>0.085$  belong to the anatase phase (see SI for more information). Both figures show the MAE between the respective target RDF and the instantaneous RDF in the biased system. Additionally, the RDF of the liquid  $\text{TiO}_2$  at the beginning of the biased simulations (blue), the final RDF at the end of the biased simulations after 2 ns for rutile and 20 ns for anatase (red), and the respective target RDFs (black) are shown in the inset.

Overall, the figures indicate that both biased simulations proceeded in a similar manner. During the initial phase of the biased simulation, the simulation box stretched in one direction, but the overall density of the system increased. The initial phase of the biased simulation was characterized by a linear reduction of the MAE in both systems. After the initial phase, a nucleation seed formed in the biased simulations after approximately 1.3 ns for the rutile system and 10.7 ns for the anatase system. From this point on, the system entered a phase transition process and complete crystallization occurred from the nucleation seed. The phase transition was characterized by a steep decrease of the MAE. After about 1.5 and 12.0 ns, respectively, the transition was complete and no further significant changes in atomic configuration occurred. Except for some visible defects, the systems crystallized completely. The defects were responsible for the small deviations from the final system RDF (red) and the target RDF (black) in the inset. Overall, the target RDF and the instantaneous system RDF exhibited strong agreement.

The comparative examples of rutile and anatase crystallization demonstrate that the maximum relative entropy formalism can be effectively applied to reproduce different crystal configurations of polymorphs. In contrast to the phase transitions of liquid water to hexagonal ice, as discussed in Sec. 3.2, both crystallized systems form a single crystal with only minor defects, as evidenced by the snapshots (e) in Figures 6 and 7, respectively. To assess the thermal stability of the intermediate and final structures of rutile, 16 structures were extracted from the simulation trajectory at simulation times ranging from 1.3 to 1.6 ns with a temporal distance of 0.02 ns. These structures were utilized as initial configurations to simulate the  $\text{TiO}_2$  system as an  $NpT$  ensemble at 2500 K and 1 atm with an unbiased force field. The continuation of the crystallization process from the crystallization seed was dependent on the precise time of the biased simulation from which the initial structure was acquired. For all initial structures obtained from the biased simulation prior to the 1.42 ns time threshold, the crystallization seed underwent a phase transition to a liquid state, resulting in the return of the  $\text{TiO}_2$  to its liquid phase. Conversely, every initial structure obtained from the biased simulation at or later than 1.42 ns underwent a complete transformation into the stable crystalline rutile state, suggesting that these structures have surmounted the free energy barrier. The derived point in time at which the crystallization seed consequently formed a critical nucleus is indicated in Figure 6 by the vertical dotted black line. These results demonstrate the efficacy of utilizing biased simulation trajectories to determine committor probabilities of atomic structures and critical nucleus sizes. In this case, the crystallization process was feasible with only minor changes

in the simulation or algorithm parameters. More specifically, to reproduce anatase instead of rutile, it was only necessary to change the step size,  $\gamma$ , and the target RDF.

#### 4. CONCLUSIONS AND OUTLOOK

An algorithm was developed to bias MD simulations to reproduce target radial distribution functions. This was done based on the principle of maximum relative entropy. For this purpose, the original atomic interaction potential was linearly biased with a target RDF. The linear functional form of the applied bias, as determined by the maximum entropy approach, is advantageous because it constrains the reproduced ensemble average solely by the information contained in the target RDF.

The magnitude of the bias is determined by the values of the Lagrangian multipliers, which are calculated during the simulation using the gradient descent method. Since the  $g_b$  are not uncorrelated observables of the system, it cannot be guaranteed that the optimization function has a global minimum. However, this has not been found to affect the systems studied. The algorithm allows individual atomic contributions to the RDFs to be weighted by atomic scattering factors. This allows the bias of MD simulations with RDFs derived from WAXS techniques. The algorithm has been implemented in LAMMPS and evaluated for two use cases.

In one use case, the RDF of a TIP4P/2005 liquid water system was reproduced in a TIP3P liquid water system. The method proved to be effective in accurately reproducing the RDF of the TIP4P/2005 water system while improving the agreement between the angular distribution of the oxygen molecules in the bias and target systems. The application of bias was found to result in a reduction of the diffusivity of the water molecules. Based on these results, it can be concluded that the approach represents a theoretically sound methodology for updating an interatomic potential using prior knowledge of a system RDF to investigate structural properties of the liquid. At the same time, studies investigating the dynamical properties of an atomic system should be treated with caution.

As a second application, the method was used to induce homogeneous crystallization in liquid systems. To achieve this, liquid water was biased to crystallize into hexagonal ice and liquid  $\text{TiO}_2$  was biased to crystallize into rutile or anatase. The process of crystallization of liquid water gave rise to crystalline structures of hexagonal ice. However, the crystallization process itself did not produce a single ice crystal, but rather multiple domains of crystals with varying crystal plane directions. The absence of a single, thermally stable crystal domain for temperatures below the melting point of ice, resulted in the melting of the crystal structure once the bias was alleviated from the system. The underlying reasons for the formation of these crystal domains under the influence of the bias will be the focus of future research. In contrast, the crystallization of liquid  $\text{TiO}_2$  into rutile and anatase produced stable single crystals. Furthermore, the simulation trajectory of the rutile phase transition demonstrated that the observed transition states are stable and could be used to investigate the crystallization process, e.g., in the form of the calculation of committer probabilities. During the execution of the biased simulations with target RDFs derived from crystalline structures, a temperature increase was observed in strongly biased systems. This increase can be attributed to the sharp peaks of the crystalline RDFs, which in turn lead to high

difference in values of the Lagrange parameters  $\lambda_b$  of adjacent bins. Consequently, strong forces are added to the simulated system. For moderate temperature increases, no resulting problems were found in the reproduction of the target RDF. Additionally, the effect can be mitigated by increasing the damping parameter of the used thermostat or by downscaling the time steps to shorter lengths.

The method presented here has a variety of potential applications. One such application could be the improvement of liquid interaction potentials by biasing the potential with a RDF, as demonstrated for the TIP3P water model. Another application could be the study of liquid-to-solid phase transitions. Through a sophisticated combination of biased and unbiased simulations, it may be possible to prepare states along a crystallization transition path and continue with unbiased simulations from these states, or use the prepared states as starting points for advanced sampling methods, such as umbrella sampling.

In addition, the algorithm described here could assist in the interpretation of experimentally obtained RDFs from WAXS measurements of liquid or solid states. While the atomic states considered in this work are well-known, this algorithm could reveal other atomic configurations by combining measured RDFs with MD simulations. This is particularly true for metastable states, as the method presented here is able to reproduce such metastable states, in contrast to other methods such as simulated annealing. General purpose machine-learned atomic interaction potentials, which can extrapolate energies of stable atomic configurations to which they have not been directly trained,<sup>67,68</sup> in combination with the algorithm presented here, offer the possibility to interpret WAXS measurements of atomic structures that are still undiscovered.

In the future, the method may also be applied to more complex systems, such as crystallization processes of confined matter, where only the confined atoms would be distorted by the measured RDFs.

The method can also be used to provide atomic configurations for training machine learning potentials. For a machine learning potential to interpolate effectively, a diverse range of atomic configurations must be provided as under-sampled regions of the configuration space, which can lead to unsatisfactory generalization of the models. However, generating a sufficient number of atomic configuration samples, especially in the transition regions of a crystallization process, is a challenging task. In this way, our method could contribute to the training of machine learning potential specialized for the study of specific phase transitions.

In addition to the discussed applications, the presented method offers many directions for further research. With respect to the interpretation of WAXS data, it might be interesting to develop a method that uses the principle of maximum relative entropy to bias MD simulations with the directly measurable structure factor function  $S(q)$ . A bias derived directly from the structure factor function  $S(q)$  would avoid an initial Fourier transform of the measurement data, which in turn would avoid additional noise. In general, the application of the principle of maximum relative entropy can be used in conjunction with any system observable derived from the ensemble average. Therefore, the use of local bond order parameters to bias MD systems would be of interest, as these parameters are designed to discriminate between different atomic configurations.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

The simulation code is available under: <https://collaborating.tuhh.de/m-29/software/maxentrdf>

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.4c01621>

Explanatory content regarding the used average local bond order parameters and determination of diffusion coefficients (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Lars Dammann** – Institute of Surface Science, Helmholtz-Zentrum Hereon, 21502 Geesthacht, Germany; Institute for Soft Matter Modeling, Hamburg University of Technology, 21073 Hamburg, Germany; Institute for Materials and X-Ray Physics, Hamburg University of Technology, 21073 Hamburg, Germany; Centre for X-ray and Nano Science CXNS, Deutsches Elektronen-Synchrotron DESY, 22607 Hamburg, Germany; [orcid.org/0000-0001-5759-3896](https://orcid.org/0000-0001-5759-3896); Email: [lars.dammann@tuhh.de](mailto:lars.dammann@tuhh.de)

**Robert H. Meißner** – Institute of Surface Science, Helmholtz-Zentrum Hereon, 21502 Geesthacht, Germany; Institute for Soft Matter Modeling, Hamburg University of Technology, 21073 Hamburg, Germany; [orcid.org/0000-0003-1926-114X](https://orcid.org/0000-0003-1926-114X); Email: [robert.meissner@tuhh.de](mailto:robert.meissner@tuhh.de)

### Authors

**Richard Kohns** – Institute for Materials and X-Ray Physics, Hamburg University of Technology, 21073 Hamburg, Germany; Centre for X-ray and Nano Science CXNS, Deutsches Elektronen-Synchrotron DESY, 22607 Hamburg, Germany; [orcid.org/0000-0002-8153-8376](https://orcid.org/0000-0002-8153-8376)

**Patrick Huber** – Institute for Materials and X-Ray Physics, Hamburg University of Technology, 21073 Hamburg, Germany; Centre for X-ray and Nano Science CXNS, Deutsches Elektronen-Synchrotron DESY, 22607 Hamburg, Germany; [orcid.org/0000-0002-2126-9100](https://orcid.org/0000-0002-2126-9100)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.4c01621>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We acknowledge the support by DASHH (Data Science in Hamburg - Helmholtz Graduate School for the Structure of Matter) with the Grant-No. HIDSS-0002. We also acknowledge the scientific exchange and support of the Centre for Molecular Water Science CMWS and the research initiative BlueMat: Water-Driven Materials, Hamburg (Germany). This research was supported in part through the Maxwell computational resources operated at Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 192346071; 390794421.

## ■ REFERENCES

(1) Piaggi, P. M.; Weis, J.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Car, R. Homogeneous ice nucleation in an ab initio machine-

learning model of water. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, No. e2207294119.

(2) Pedevilla, P.; Fitzner, M.; Sosso, G. C.; Michaelides, A. Heterogeneous seeded molecular dynamics as a tool to probe the ice nucleating ability of crystalline surfaces. *J. Chem. Phys.* **2018**, *149*, No. 072327.

(3) Sosso, G. C.; Tribello, G. A.; Zen, A.; Pedevilla, P.; Michaelides, A. Ice formation on kaolinite: Insights from molecular dynamics simulations. *J. Chem. Phys.* **2016**, *145*, No. 211927.

(4) Sanz, E.; Vega, C.; Espinosa, J. R.; Caballero-Bernal, R.; Abascal, J. L. F.; Valeriani, C. Homogeneous Ice Nucleation at Moderate Supercooling from Molecular Simulation. *J. Am. Chem. Soc.* **2013**, *135*, 15008–15017.

(5) Espinosa, J. R.; Vega, C.; Valeriani, C.; Sanz, E. Seeding approach to crystal nucleation. *J. Chem. Phys.* **2016**, *144*, No. 034501.

(6) Biswas, R.; Hamann, D. R. Simulated annealing of silicon atom clusters in Langevin molecular dynamics. *Phys. Rev. B* **1986**, *34*, 895–901.

(7) Zhang, S.; Wu, Y. Investigating the crystallization behavior of TiO<sub>2</sub> during annealing: Molecular dynamics simulations. *AIP Adv.* **2023**, *13*, No. 085226.

(8) Piaggi, P. M.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Car, R. Phase Equilibrium of Water with Hexagonal and Cubic Ice Using the SCAN Functional. *J. Chem. Theory Comput.* **2021**, *17*, 3065–3077.

(9) Cheng, B.; Tribello, G. A.; Ceriotti, M. The Gibbs free energy of homogeneous nucleation: From atomistic nuclei to the planar limit. *J. Chem. Phys.* **2017**, *147*, No. 104707.

(10) Niu, H.; Piaggi, P. M.; Invernizzi, M.; Parrinello, M. Molecular dynamics simulations of liquid silica crystallization. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 5348–5352.

(11) Giberti, F.; Salvalaglio, M.; Parrinello, M. Metadynamics studies of crystal nucleation. *IUCrJ.* **2015**, *2*, 256–266.

(12) Bonati, L.; Parrinello, M. Silicon Liquid Structure and Crystal Nucleation from Ab Initio Deep Metadynamics. *Phys. Rev. Lett.* **2018**, *121*, No. 265701.

(13) Gobbo, G.; Bellucci, M. A.; Tribello, G. A.; Ciccotti, G.; Trout, B. L. Nucleation of Molecular Crystals Driven by Relative Information Entropy. *J. Chem. Theory Comput.* **2018**, *14*, 959–972.

(14) Zhang, Y.-Y.; Niu, H.; Piccini, G.; Mendels, D.; Parrinello, M. Improving collective variables: The case of crystallization. *J. Chem. Phys.* **2019**, *150*, No. 094509.

(15) Nada, H. Pathways for the formation of ice polymorphs from water predicted by a metadynamics method. *Sci. Rep.* **2020**, *10*, No. 4708.

(16) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(17) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, No. 020603.

(18) Coasne, B.; Jain, S. K.; Gubbins, K. E. Freezing of Fluids Confined in a Disordered Nanoporous Structure. *Phys. Rev. Lett.* **2006**, *97*, No. 105702.

(19) Hung, F. R.; Coasne, B.; Santiso, E. E.; Gubbins, K. E.; Siperstein, F. R.; Sliwiska-Bartkowiak, M. Molecular modeling of freezing of simple fluids confined within carbon nanotubes. *J. Chem. Phys.* **2005**, *122*, No. 144706.

(20) Scalfi, L.; Coasne, B.; Rotenberg, B. On the Gibbs–Thomson equation for the crystallization of confined fluids. *J. Chem. Phys.* **2021**, *154*, No. 114711.

(21) Malolepsza, E.; Secor, M.; Keyes, T. Isobaric Molecular Dynamics Version of the Generalized Replica Exchange Method (gREM): Liquid–Vapor Equilibrium. *J. Phys. Chem. B* **2015**, *119*, 13379–13384.

(22) Malolepsza, E.; Kim, J.; Keyes, T. Generalized ensemble method applied to study systems with strong first order transitions. *J. Phys.: Conf. Ser.* **2015**, *640*, No. 012003.

(23) Malolepsza, E.; Keyes, T. Water Freezing and Ice Melting. *J. Chem. Theory Comput.* **2015**, *11*, 5613–5623.

- (24) Egami, T.; Billinge, S. J. L. *Underneath the Bragg Peaks: Structural Analysis of Complex Materials*; Elsevier Science, 2012.
- (25) McGreevy, R. L.; Pusztai, L. Reverse Monte Carlo Simulation: A New Technique for the Determination of Disordered Structures. *Mol. Simul.* **1988**, *1*, 359–367.
- (26) McGreevy, R. L. Reverse Monte Carlo modelling. *J. Phys.: Condens. Matter* **2001**, *13*, R877.
- (27) Keen, D. A.; Tucker, M. G.; Dove, M. T. Reverse Monte Carlo modelling of crystalline disorder. *J. Phys.: Condens. Matter* **2005**, *17*, S15.
- (28) Bousige, C.; Boğan, A.; Ulm, F.-J.; Pellenq, R. J.-M.; Coasne, B. Optimized molecular reconstruction procedure combining hybrid reverse Monte Carlo and molecular dynamics. *J. Chem. Phys.* **2015**, *142*, No. 114112.
- (29) Farmahini, A. H.; Bhatia, S. K. Hybrid Reverse Monte Carlo simulation of amorphous carbon: Distinguishing between competing structures obtained using different modeling protocols. *Carbon* **2015**, *83*, 53–70.
- (30) Pitera, J. W.; Chodera, J. D. On the Use of Experimental Observations to Bias Simulated Ensembles. *J. Chem. Theory Comput.* **2012**, *8*, 3445–3451.
- (31) Cesari, A.; Reißer, S.; Bussi, G. Using the Maximum Entropy Principle to Combine Simulations and Solution Experiments. *Computation* **2018**, *6*, 15.
- (32) Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **2008**, *129*, No. 144108.
- (33) Shell, M. S. *Advances in Chemical Physics*; John Wiley & Sons, Ltd, 2016; pp 395–441 DOI: 10.1002/9781119290971.ch5.
- (34) Chaimovich, A.; Shell, M. S. Coarse-graining errors and numerical optimization using a relative entropy framework. *J. Chem. Phys.* **2011**, *134*, No. 094112.
- (35) Cesari, A.; Gil-Ley, A.; Bussi, G. Combining Simulations and Solution Experiments as a Paradigm for RNA Force Field Refinement. *J. Chem. Theory Comput.* **2016**, *12*, 6192–6200.
- (36) Fröhlich, T.; Bernetti, M.; Calonaci, N.; Bussi, G. Toward empirical force fields that match experimental observables. *J. Chem. Phys.* **2020**, *152*, No. 230902.
- (37) Cilloco, F. Relative entropy minimization and the determination of pair interactions. *J. Mol. Struct.* **1993**, *296*, 253–258.
- (38) White, A. D.; Voth, G. A. Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data. *J. Chem. Theory Comput.* **2014**, *10*, 3023–3030.
- (39) White, A. D.; Dama, J. F.; Voth, G. A. Designing Free Energy Surfaces That Match Experimental Data with Metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2451–2460.
- (40) Warren, B. E.; Krutter, H.; Morningstar, O. FOURIER ANALYSIS OF X-RAY PATTERNS OF VITREOUS SiO<sub>2</sub> AND B<sub>2</sub>O<sub>3</sub>. *J. Am. Ceram. Soc.* **1936**, *19*, 202–206.
- (41) Gao, X. The mathematics of the ensemble theory. *Results Phys.* **2022**, *34*, No. 105230.
- (42) Caticha, A. Relative Entropy and Inductive Inference. *AIP Conf. Proc.* **2004**, *707*, 75–96.
- (43) de Miguel, E.; Jackson, G. The nature of the calculation of the pressure in molecular simulations of continuous models from volume perturbations. *J. Chem. Phys.* **2006**, *125*, No. 164109.
- (44) Mead, L. R.; Papanicolaou, N. Maximum entropy in the problem of moments. *J. Math. Phys.* **1984**, *25*, 2404–2417.
- (45) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, No. 108171.
- (46) Coleman, S. P.; Sichani, M. M.; Spearot, D. E. A Computational Algorithm to Produce Virtual X-ray and Electron Diffraction Patterns from Atomistic Simulations. *JOM* **2014**, *66*, 408–416.
- (47) Prince, E. *International tables for crystallography. C: Mathematical, physical and chemical tables/Hrsg*, 3rd ed.; Prince, E., Ed.; Springer Netherland: Berlin, 2007 DOI: 10.1107/97809553602060000103.
- (48) Peng, L. M.; Ren, G.; Dudarev, S. L.; Whelan, M. J. Robust Parameterization of Elastic and Absorptive Electron Atomic Scattering Factors. *Acta Crystallogr.* **1996**, *52*, 257–276.
- (49) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (50) Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, No. 234505.
- (51) Kadaoluwa Pathirannahalage, S. P.; Meftahi, N.; Elbourne, A.; Weiss, A. C. G.; McConville, C. F.; Padua, A.; Winkler, D. A.; Costa Gomes, M.; Greaves, T. L.; Le, T. C.; Besford, Q. A.; Christofferson, A. J. Systematic Comparison of the Structural and Dynamic Properties of Commonly Used Water Models for Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2021**, *61*, 4521–4536.
- (52) Dorrani, H.; Mohebbi, A. A Comparative Study of TIP4P-2005, SPC/E, SPC, and TIP3P-Ew Models for Predicting Water Transport Coefficients Using EMD and NEMD Simulations. *J. Eng. Thermophys.* **2023**, *32*, 138–161.
- (53) Markland, T. E.; Habershon, S.; Manolopoulos, D. E. Quantum diffusion of hydrogen and muonium atoms in liquid water and hexagonal ice. *J. Chem. Phys.* **2008**, *128*, No. 194506.
- (54) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960. Publisher: American Chemical Society.
- (55) Chan, H.; Cherukara, M. J.; Narayanan, B.; Loeffler, T. D.; Benmore, C.; Gray, S. K.; Sankaranarayanan, S. K. R. S. Machine learning coarse grained models for water. *Nat. Commun.* **2019**, *10*, No. 379.
- (56) Dhabal, D.; Kumar, R.; Molinero, V. Liquid–liquid transition and ice crystallization in a machine-learned coarse-grained water model. *Proc. Natl. Acad. Sci. U.S.A.* **2024**, *121*, No. e2322853121.
- (57) Abascal, J. L. F.; Sanz, E.; García Fernández, R.; Vega, C. A potential model for the study of ices and amorphous water: TIP4P/Ice. *J. Chem. Phys.* **2005**, *122*, No. 234511.
- (58) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (59) Lechner, W.; Dellago, C. Accurate determination of crystal structures based on averaged local bond order parameters. *J. Chem. Phys.* **2008**, *129*, No. 114707.
- (60) Steinhardt, P. J.; Nelson, D. R.; Ronchetti, M. Bond-orientational order in liquids and glasses. *Phys. Rev. B* **1983**, *28*, 784–805.
- (61) Miyoshi, A.; Nishioka, S.; Maeda, K. Water Splitting on Rutile TiO<sub>2</sub>-Based Photocatalysts. *Chem. - Eur. J.* **2018**, *24*, 18204–18219.
- (62) Hanaor, D. A. H.; Sorrell, C. C. Review of the anatase to rutile phase transformation. *J. Mater. Sci.* **2011**, *46*, 855–874.
- (63) Ziental, D.; Czarczynska-Goslinska, B.; Mlynarczyk, D. T.; Glowacka-Sobotta, A.; Stanisz, B.; Goslinski, T.; Sobotta, L. Titanium Dioxide Nanoparticles: Prospects and Applications in Medicine. *Nanomaterials* **2020**, *10*, 387.
- (64) Bai, J.; Zhou, B. Titanium Dioxide Nanomaterials for Sensor Applications. *Chem. Rev.* **2014**, *114*, 10131–10176.
- (65) Matsui, M.; Akaogi, M. Molecular Dynamics Simulation of the Structural and Physical Properties of the Four Polymorphs of TiO<sub>2</sub>. *Mol. Simul.* **1991**, *6*, 239–244.
- (66) Tadmor, E. B.; Elliott, R. S.; Sethna, J. P.; Miller, R. E.; Becker, C. A. The potential of atomistic simulations and the knowledgebase of interatomic models. *JOM* **2011**, *63*, 17.
- (67) Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2022**, *2*, 718–728.
- (68) Batatia, I.; Benner, P.; Chiang, Y.; Elena, A. M.; Kovács, D. P.; Riebesell, J.; Advincula, X. R.; Asta, M.; Avaylon, M.; Baldwin, W. J.; Berger, F.; Bernstein, N.; Bhowmik, A.; Blau, S. M.; Cărare, V.; Darby,

J. P.; De, S.; Pia, F. D.; Deringer, V. L.; Elijošius, R.; El-Machachi, Z.; Falcioni, F.; Fako, E.; Ferrari, A. C.; Genreith-Schriever, A.; George, J.; Goodall, R. E. A.; Grey, C. P.; Grigorev, P.; Han, S.; Handley, W.; Heenen, H. H.; Hermansson, K.; Holm, C.; Jaafar, J.; Hofmann, S.; Jakob, K. S.; Jung, H.; Kapil, V.; Kaplan, A. D.; Karimitari, N.; Kermode, J. R.; Kroupa, N.; Kullgren, J.; Kuner, M. C.; Kuryla, D.; Liepuoniute, G.; Margraf, J. T.; Magdău, I.-B.; Michaelides, A.; Moore, J. H.; Naik, A. A.; Niblett, S. P.; Norwood, S. W.; O'Neill, N.; Ortner, C.; Persson, K. A.; Reuter, K.; Rosen, A. S.; Schaaf, L. L.; Schran, C.; Shi, B. X.; Sivonxay, E.; Stenczel, T. K.; Svahn, V.; Sutton, C.; Swinburne, T. D.; Tilly, J.; Oord, C. v. d.; Varga-Umbrich, E.; Vegge, T.; Vondrák, M.; Wang, Y.; Witt, W. C.; Zills, F.; Csányi, G. A foundation model for atomistic materials chemistry. 2024; <http://arxiv.org/abs/2401.00096>.