Crystal nucleation along an entropic pathway: Teaching liquids how to transition

Caroline Desgranges and Jerome Delhommelle

Department of Chemistry, University of North Dakota, Grand Forks, North Dakota 58202, USA

We combine machine learning (ML) with Monte Carlo (MC) simulations to study the crystal nucleation process. Specifically, we use ML to infer the canonical partition function of the system over the range of densities and temperatures spanned during crystallization. We achieve this on the example of the Lennard-Jones system by training an artificial neural network using, as a reference dataset, equations of state for the Helmholtz free energy for the liquid and solid phases. The accuracy of the ML predictions is tested over a wide range of thermodynamic conditions, and results are shown to provide an accurate estimate for the canonical partition function, when compared to the results from flat-histogram simulations. Then, the ML predictions are used to calculate the entropy of the system during MC simulations in the isothermal-isobaric ensemble. This approach is shown to yield results in very good agreement with the experimental data for both the liquid and solid phases of argon. Finally, taking entropy as a reaction coordinate and using the umbrella sampling technique, we are able to determine the Gibbs free energy profile for the crystal nucleation process. In particular, we obtain a free energy barrier in very good agreement with the results from previous simulation studies. The approach developed here can be readily extended to molecular systems and complex fluids, and is especially promising for the study of entropy-driven processes.

I. INTRODUCTION

Entropy is central to many phenomena in chemistry, physics, biology, and materials science. For instance, entropy is a key concept in our understanding of molecular association processes, such as, e.g., the dimerization of insulin [1], of pattern formation on nanoparticle surfaces [2], of the partitioning behavior of small molecules in lipid bilayers [3], and of self-assembly processes [4,5]. Furthermore, entropy provides a measure of the amount of coded information contained in the canonical genetic code [6], as well as a thermodynamic characterization of supercooled liquids [7,8] and amorphous materials [9,10]. Entropic effects are also crucial for mixture separation during adsorption processes [11,12], for the design of protein geometries to facilitate the active transport of ions [13], and in protein-protein binding for signal transduction and molecular recognition [14,15].

Since entropy provides a measure of the amount of order, or conversely, of disorder, within a system [16], it is often used to characterize the level of organization of this system. As such, entropy-driven processes have been the focus of intense research in recent years and have led to many insightful discoveries. Such phenomena include the entropy-driven self-assembly in charged lock-key particles [17], entropy-driven segregation processes of polymer-grafted nanoparticles [18], and the entropy-driven crystallization of DNA grafted nanoparticles [19], polymers [20], proteins [21], and atomic [22], molecular [23], and polymeric crystals [24]. Furthermore, such studies have shed light on the interplay between competing processes, e.g., entropy-driven segregation and crystallization [25,26] that can lead to complex molecular mechanisms, such as two-step nucleation processes for silicon [27] and clathrate colloidal crystals [28]. Recent work has also highlighted the key role played by entropy during the nucleation of liquid droplets [29,30], the formation of capillary bridges during condensation processes in nanotubes [31–33], and the nucleation of cavitation bubbles [34].

A full characterization of these entropic effects hinges on the determination of entropy along the transition pathway [35]. However, such calculations are not straightforward, since a direct determination of the entropy $S = -k_B \sum_i P_i^B \ln P_i^B$ requires knowing the Boltzmann probability P_i^B, in which i denotes a system configuration [36–39]. Alternatively, free energy calculations can be carried out to determine the entropy change along a given pathway, using either thermodynamic integration [40–43] or flat histogram methods [44–47]. Other possible routes involve entropy expansions in terms of the correlation functions [48,49] or the analysis of the dynamical response of the system, through a Fourier transform of the velocity autocorrelation function [50,51]. In this work, we adopt a different approach and take advantage of machine learning techniques to evaluate on-the-fly the entropy of the system as it undergoes crystal nucleation. Machine learning is now widely used in applications as diverse as modeling molecular atomization energies [52], density functional theory [53], and molecular docking [54]. In particular, artificial neural networks (ANNs) have been shown to give excellent results for the free energy landscape of a wide range of systems, including peptides, biomolecules, and polymers [55–57]. Here, we use ML to predict the canonical partition function of the system over a wide range of temperature and densities spanned during the crystallization process. To achieve this, we train ANNs to model accurately the Helmholtz free energy of the Lennard-Jones system, as given by an equation of state obtained by Johnson *et al.* for the liquid phase [58] and by the equation of state of van der

Hoef [59] for the solid phase. Then, using the predictions provided by the ANN for the Helmholtz free energy, we are able to evaluate the entropy S of the system. The advantages of this approach are threefold, since it allows us to (i) evaluate the canonical partition function and thus the Helmholtz free energy as a function of only two variables, density and temperature, along the crystallization process, (ii) calculate the entropy of a system over a wide range of conditions, and (iii) define entropy as a reaction coordinate to follow the onset of order in the system, here during a crystal nucleation event. To assess the first point, we show that the ML predictions are accurate over a wide range of conditions and yield estimates for the canonical partition function in excellent agreement with the results from flat-histogram simulations. Second, we compare the entropy obtained on the Lennard-Jones system to experimental data available for argon [60] and show that the entropy obtained using our method is in very good agreement with the data on both liquid and solid argon. Third, we simulate the crystal nucleation process in the isothermal-isobaric ensemble using entropy as a reaction coordinate. Specifically, we implement the umbrella sampling technique within NPT Monte Carlo simulations and compare our results with prior work in the field [61] using geometric parameters, such as the bond orientational order parameter Q_6 defined by Steinhardt *et al.* [62].

The paper is organized as follows. In the next section, we present the approach developed in this work, starting with the type of machine learning technique used here. We also discuss how entropy is evaluated and used as a reaction coordinate in Monte Carlo simulations and show how the ML predictions yield accurate results for the partition function, and, in turn, for the thermodynamic properties. We then study the crystal nucleation process from a supercooled liquid of Lennard-Jones particles, determine the free energy barrier of nucleation along the entropic pathway and compare our results to those from previous work. We finally draw the main conclusions from this work in the last section.

II. SIMULATION METHOD

A. Machine learning technique and entropy calculations

The first step in our approach consists of using machine learning (ML) to model accurately the canonical partition function $Q(N, V, T)$, or equivalently the Helmholtz free energy $A = -k_B T \ln Q(N, V, T)$, for densities and temperatures spanned during crystallization. For this purpose, we train an artificial neural network using as reference data the Helmholtz free energy of the Lennard-Jones system from the Johnson, Zollweg and Gubbins equation of state (JZG EOS) for the fluid phase [58] and from the van der Hoef equation of state (VDH EOS) for the face-centered cubic crystal [59]. The structure of the ANN used in this work is given in Fig. 1.

The ANN we employ includes four layers, the input and output layers, as well as two hidden layers. The input layer is composed of two input neurons, corresponding to the two input parameters, temperature T and density ρ, for which we need to evaluate the Helmholtz free energy of the system. The next two layers are hidden layers, for which we have used, respectively, five and four neurons. We add that we include the usual bias node b for the input layer and every hidden
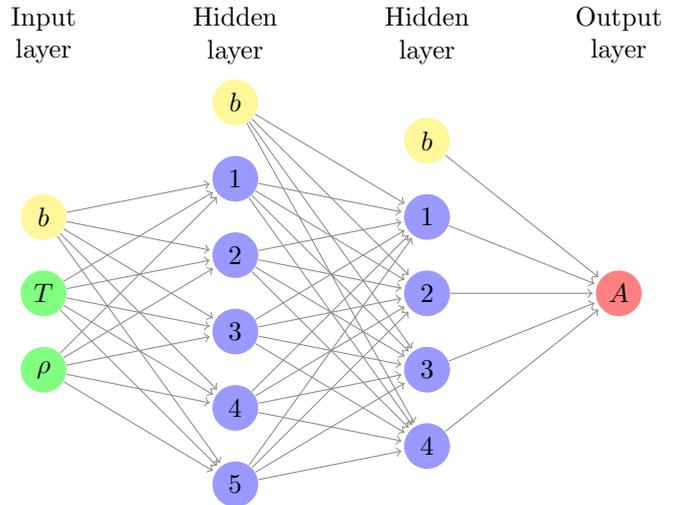

FIG. 1. Structure of the ANN used in this work for the determination of the Helmholtz free energy A for a given temperature T and density ρ, in which each neuron is labeled with a number between 1 and 5 for the first hidden layer, between 1 and 4 for the second hidden layer, and in which b designates a bias node for each of the first three layers.

layer [55–57,63]. Finally, the output layer only contains a single neuron, since we aim here to predict the Helmholtz free energy for any (T, ρ) set. The machine-learned Helmholtz free energy A^{ML} can then be obtained through the following equation [64,65],

$$A^{ML} = f_4[b_3 + \sum_{l=1}^{4} W(3, 4, l, 1) f_3(b_2$$

$$+ \sum_{j=1}^{5} W(2, 3, j, l) f_2[(b_1 + W(1, 2, 1, j)T$$

$$+ W(1, 2, 2, j)\rho])], \tag{1}$$

in which $W(i - 1, i, k_{i-1}, k_i)$ is the weight matrix connecting neuron k_{i-1} from layer $i - 1$ to neuron k_i from layer i, the b_i functions are bias nodes that act as adjustable offsets, the f_i functions are the activation functions, chosen as the tanh function for $i = 1, 2, 3$ and the linear function for $i = 4$. The weights for the ANN are trained during an iterative process, going first through a forward pass and then a back-propagation algorithm. Equation (1) therefore provides a way to estimate the Helmholtz free energy not only for the equilibrium phases, i.e., the solid and liquid which serve to train the ANN, but also for the intermediate, nonequilibrium, states [66] that connect the starting point, the metastable liquid, to the top of the free energy barrier, corresponding to the formation of the critical nucleus.

The entropy of the system S can be calculated from the canonical partition $Q(N, V, T)$ function through the following statistical mechanical expression

$$S = \frac{U + k_B T \ln Q(N, V, T)}{T} \tag{2}$$

in which U is the internal energy of the system.

An important insight brought by this connection between $Q(N, V, T)$ and S is the intrinsically statistical nature of the entropy as defined in Eq. (2). $Q(N, V, T)$ contains all information for a system of a given size, both for homogeneous systems, when the density N/V corresponds to either a liquid or a solid phase, and for inhomogeneous systems, for intermediate densities for which the system may contain nuclei or slabs of the new phase immersed in the parent phase [66]. However, as a result of the finite size of the system, $Q(N, V, T)$ and thus the Helmholtz free energy will vary smoothly and will depend on the system size for these intermediate densities, as shown, e.g., in prior flat histogram simulations of first-order phase transitions [46,67]. In turn, while the value of S, calculated from Eq. (2), is expected to match the bulk values for the solid and liquid phases, its value can depend on the system size for the intermediate nonequilibrium states that span the nucleation pathway.

Equation (2) suggests a route for the determination of entropy during the simulations in a form that can be readily used as a reaction coordinate for the nucleation process. For this purpose, starting from Eq. (2) and using the relation between the Helmholtz free energy and the canonical partition function, we define S_i for each configuration i generated during the (N, P, T) simulations as

$$S_i = \frac{U_i - A^{ML}(\rho_i, T)}{T} \tag{3}$$

in which $U_i = U_i^{\text{pot}} + \frac{3}{2} N k_B T$ is the sum of the interaction (potential) energy and of the ideal gas (kinetic) contribution to the internal energy and $A^{ML}(\rho_i, T)$ are the ML predictions for the Helmholtz free energy at a density $\rho_i = \frac{N}{V_i}$ and a temperature T. We assess in Sec. III the accuracy and reliability of Eq. (3) for the liquid and for the solid by testing it against experimental data for argon, for which the Lennard-Jones model performs well, and against the results from flat-histogram simulations.

B. Exploring an entropic pathway

The next step consists of using the entropy, obtained from our machine learning approach, as a reaction coordinate. Previous work has shown that the umbrella sampling (US) technique [68] is a very successful approach to study the onset of nucleation, both for the vapor-liquid transition [29,69–72] and the solid-liquid transition [26,61,73–86]. We therefore build on our previous work [29,30,33], in which we simulated the nucleation of liquid droplets using US in the grand-canonical $\mu V T$ ensemble. Here, since we focus on crystal nucleation, the very low acceptance rates for the insertions and deletions of atoms preclude the use of $\mu V T$ simulations. We therefore implement the US technique, using the machine-learned entropy as a reaction coordinate, within isothermal-isobaric NPT Monte Carlo simulations. The US method is a non-Boltzmann sampling technique [68,87] that consists in applying a US potential, usually chosen as a harmonic function of a reaction coordinate, to sample configurations with a very low Boltzmann weight. Once the simulation data have been collected, the US potential is then subsequently removed to obtain the free energy profile for the process, specifically here the Gibbs free energy profile since this approach is implemented in

the isothermal-isobaric ensemble [61,68]. The ANN weights allow for the calculation of entropy for a given temperature T and density ρ, making the use of the machine-learned entropy very well suited to serve as a reaction coordinate for processes occurring under isothermal-isobaric conditions, such as crystal nucleation. From a practical standpoint, we start from a metastable supercooled liquid (parent phase) and carry out simulations for gradually decreasing values of entropy to promote the formation of the critical nucleus. In each of these simulations, we use an US potential with the following functional form,

$$U_{\text{bias}} = \frac{1}{2} k (S_i - S_0)^2, \tag{4}$$

in which S_0 is the target value for the entropy, S_i is provided using Eq. (3) for a configuration i of the system, and k is a spring constant. Then, we collect the simulation results under the form of an histogram that yields the probability associated with each value of the entropy once the US potential has been removed [68,87]. Repeating the simulations for steadily decreasing target values for the entropy allows us to build the Gibbs free energy profile for the nucleation process as discussed in prior work [61,84,88].

C. Technical details

The first stage of this work involves training the ANN. This is carried out using a learning rate set to 0.2, as well as a training dataset that includes 11 970 data points, generated by the VDH EOS and the JZG EOS. For the solid (VDH EOS), we include data for reduced densities $\rho^* = \rho \sigma^3$ ranging from 0.94 to 1.2 and reduced temperatures $T^* = k_B T / \epsilon$ from 0.1 to 2 (here, the density and temperature are given in units reduced with respect to the Lennard-Jones parameters ϵ and σ, as discussed by Allen and Tildesley [87]). For the fluid phase (JZG EOS), we include data for temperatures ranging from 0.7 to 6. The ANN training is deemed to be complete when the error estimate becomes less than 5×10^{-7}. Once the ANN has been trained, we perform two types of simulations. First, we test the accuracy of the ML predictions for the Helmholtz free energy, the canonical partition function and for the entropy on simulations carried out on systems of 500 Lennard-Jones particles for the fluid and solid phases. Second, we carry out simulations of the crystal nucleation process using the US technique under isothermal-isobaric conditions and with the entropy as a reaction coordinate and refer to these simulations as NPT-S simulations in the rest of this work. Simulations of crystal nucleation are carried out on system of 3 000 Lennard-Jones particles for a reduced temperature $T^* = 0.86$ and a reduced pressure $P^* = 5.68$, which correspond to a supercooling of 22% [59]. All simulations are carried out within a NPT Monte Carlo (MC) framework. Two types of random moves are implemented, corresponding to the random translation of a single atom (99% of the attempted MC moves) or a random change in the volume of the cubic simulation cell, leading to a rescaling of the positions of all atoms (1% of the attempted MC moves). The maximum displacement and volume change are set such that 50% of the attempted MC moves are accepted. The usual periodic boundary conditions are applied [87], and tail corrections beyond a cutoff radius of 3σ are employed.

III. RESULTS AND DISCUSSION

A. ML predictions for the Helmholtz free energy and the canonical partition function.

We start by presenting the results obtained for the ML predictions of the Helmholtz free energy. Figure 2 shows a plot of the reduced Helmholtz free energy against the reduced density ρ^*. In this graph, we compare the ML predictions, obtained using Eq. (1), to the data from the two equations of state we use for the Lennard-Jones system, namely the JZG EOS for the liquid and the VDH EOS for the solid. Figure 2 shows that there is an excellent agreement between the ML predictions and the EOS data across a wide temperature range, from $T^* = 0.75$ to $T^* = 4$. Furthermore, we see that the weights of the trained ANN allow us to obtain very accurate results for both the liquid phase and the solid phase (face-centered cubic [fcc] crystal). More specifically, focusing, e.g., on the agreement between the VDH EOS and ML predictions for $T^* = 1.35$, we find a root mean square deviation (RMSD) of 0.11. Turning to the agreement between the JZG EOS and the ML predictions for the liquid at $T^* = 3$, we obtain a RMSD of 0.15. Both examples show that ML predictions for the Helmholtz free energy are in excellent agreement with the EOS data.

Second, we analyze how the ML predictions perform for the excess Helmholtz free energy, which is also provided by the EOS data. We show in Fig. 3 the variations of the product of the inverse temperature by the reduced excess

FIG. 3. Product of the reduced excess Helmholtz free energy by $\beta = 1/(k_B T)$ against density for the Lennard-Jones system. Same legend as in Fig. 2.
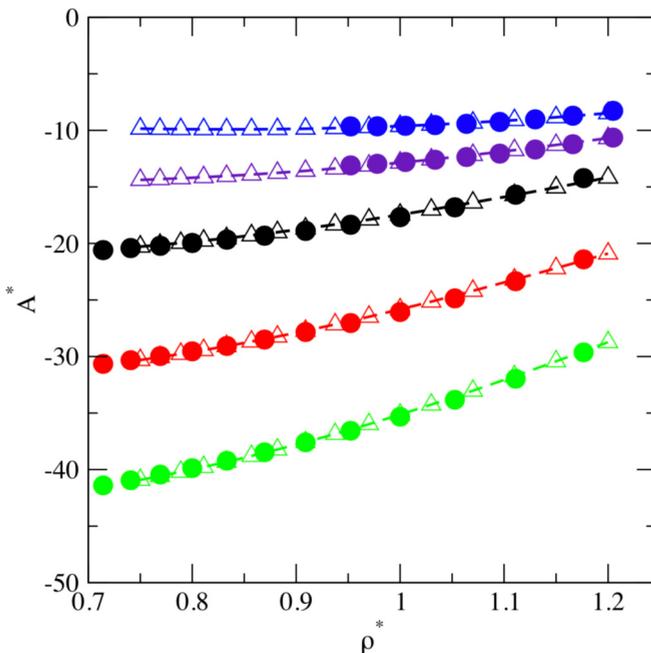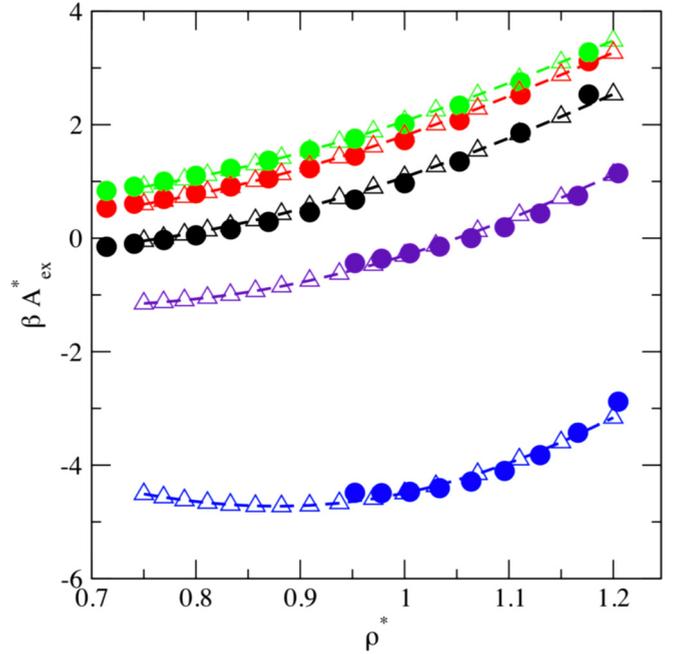

FIG. 2. Reduced Helmholtz free energy as a function of the reduced density. Results from the JZG EOS are shown as filled circles for a reduced temperature of $T^* = 2$ (in black), $T^* = 3$ (in red), and $T^* = 4$ (in green) for the liquid, while results from the VDH EOS are plotted for reduced temperatures of $T^* = 0.75$ (in blue) and $T^* = 1.35$ (in purple). The ML predictions for the corresponding conditions are shown as open triangles connected through dashed lines.

Helmholtz free energy (A_{ex}^*) for both phases, and for reduced temperatures ranging from $T^* = 0.75$ to $T^* = 4$. The quantity βA_{ex}^* behaves in a dramatically different way when compared to the reduced Helmholtz free energy. The largest values for βA_{ex}^* are now obtained for the highest temperature ($T^* = 4$) and the lowest βA_{ex}^* values are reached for $T^* = 0.75$. This is the expected behavior, since the attractive interactions between LJ particles are predominant as the temperature of the system decreases and the solid becomes the stable phase. Furthermore, this effect becomes more pronounced as a result of the scaling by the inverse temperature β at low temperatures. This qualitative trend is well captured by both the EOS data and the ML predictions. From a quantitative standpoint, in line with the results obtained for A^*, we find that the ML predictions for the excess Helmholtz free energy are in very good agreement with the JZG EOS and VDH EOS data. Considering the same state points as for A^*, we find a RMSD between the EOS data and the ML predictions of 0.05 at $T^* = 3$ for the liquid and of 0.08 at $T^* = 1.35$ for the solid. This confirms the ability of the trained ANN to calculate accurately the Helmholtz free energy of both the liquid and solid phases for the Lennard-Jones system.

To assess further the performance of the ML predictions, we now examine their ability to predict the value of the canonical partition function. The partition function can be obtained from the Helmholtz free energy from the following equation:

$$A = -k_B T \ln Q(N, V, T) \tag{5}$$

with

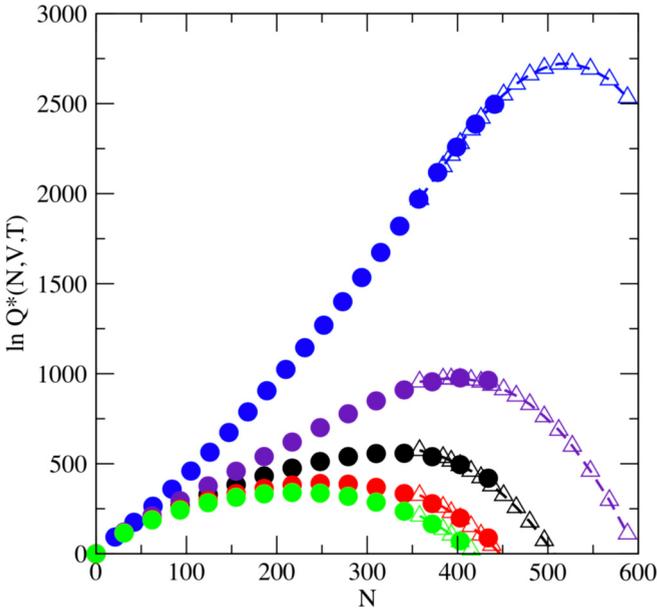$$Q(N, V, T) = \frac{V^N}{N! \Lambda^{3N}} \int \exp[-\beta U(\mathbf{\Gamma})] d\mathbf{\Gamma}, \tag{6}$$

FIG. 4. Logarithm of the reduced canonical partition functions $Q^*(N, V, T)$ against N, the number of LJ particles. Filled circles are the results from prior work using EWL simulations [46], for reduced temperatures $T^* = 0.75$ (in blue), $T^* = 1.35$ (in purple), $T^* = 2$ (in black), $T^* = 3$ (in red), and $T^* = 4$ (in green). Open triangles connected through dashed lines are the ML predictions.
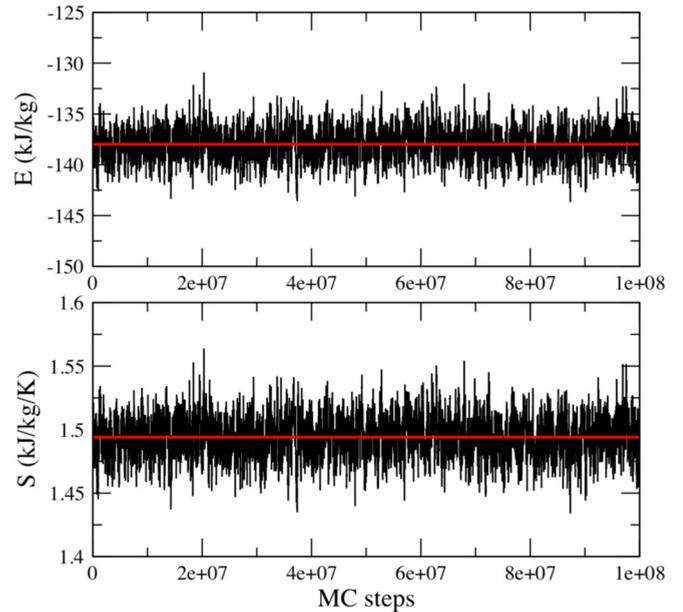

FIG. 5. Monte Carlo NPT simulations of liquid argon at $T = 95$ K and $P = 5$ bar. Variation of the interaction energy (top) and of entropy (bottom) as a function of the number of MC steps. In both plots, the red line indicated the average value.

where Λ is the de Broglie wavelength and the integration is performed over the coordinates of the N atoms of the system (Γ denoting a specific configuration of the system). This relation therefore allows us to obtain $Q^{ML}(N, V, T)$ from the ML predictions for the Helmholtz free energy. We present in Fig. 4 a comparison between the ML-derived partition function $Q^{ML}(N, V, T)$ and the partition function obtained in prior simulation work using the expanded Wang-Landau (EWL) simulation method [46]. Figure 4 shows that the ML predictions and the EWL simulation results are in very good agreement over the whole temperature range from $T^* = 0.75$ to $T^* = 4$, as demonstrated for densities corresponding to the liquid phase in the plot. This corresponds to numbers of particles between about 300 and 450 in the plot, since EWL simulations were performed for a fixed reduced volume of $512\sigma^3$. Interestingly, the ML predictions also provide an insight into how the reduced canonical partition function varies as a function of N for very large densities. This range of densities corresponding to the solid phase is not accessible through EWL simulations. This is because EWL simulations are implemented within the grand-canonical ensemble and, as such, require the insertion and deletion of atoms during the simulation, which becomes very difficult at high densities.

B. Entropy calculations for liquid and solid phases.

Now that the reliability of the ML predictions for the Helmholtz free energy has been examined, we look at how these perform for the determination of the entropy of the system. For this purpose, we carry out the first type of simulation discussed in the Methods section. In order to compare to experimental data for the liquid [60] and for the solid [89], we

present results in real units, using the conventional Lennard-Jones parameters for argon ($\sigma = 3.4$ Å, $\epsilon/k_B = 117.05$ K, and $m = 40$ g/mol). We show in Fig. 5 the results of NPT simulations carried out at $T = 95$ K and $P = 5$ bar for the liquid. In this graph, we plot the variation of the interaction energy during the simulation, as well as the entropy calculated through Eq. (3). These plots exhibit the behavior expected, once the simulation has converged, and provide an estimate for the extent of the fluctuations around the average value during the simulation.

How do the simulation results compare to the experimental data? To assess this in detail, we carry out a series of NPT simulations of the liquid phase for several state points, with pressure ranging from $P = 5$ bar to $P = 100$ bar and temperature ranging from $T = 90$ K to $T = 100$ K. The results are provided in Table I. Liquid densities predicted by the simulations are found to be in good agreement with the experimental data (within 0.03 g/cm^3), which confirms the accuracy of the model for argon under these conditions. Looking closely at the thermodynamic function enthalpy (H), we also see that there is a good agreement between the experimental data and the simulation results with the maximum deviation of 0.6 kJ/kg occurring at the the lowest temperature ($T = 90$ K). This also provides an estimate of the accuracy with which the LJ potential models the thermodynamic properties of real argon. Turning now to the entropy calculations, we find that there is a very good agreement between the experimental data and our calculations, since the maximum deviation is also observed at $T = 90$ K, in line with the findings for H, and is 0.02 kJ/(kg K). Moving on to the Gibbs free energy (G), we obtain a good agreement between the experimental data and the simulation results, with deviations of less than 1% across the entire range of conditions studied.

TABLE I. Liquid argon: Comparison between the experimental data [60] and the simulation results (this work) for various state points. Temperatures are given in K, ρ in g/cm^3, S in kJ/(kg K), and H and G in kJ/kg. Standard deviations are of the order of 3×10^{-3} g/cm^{-3} for the density, of 10^{-2} kJ/(kg K) for S, and of 0.1 kJ/kg for H and G.

T	ρ_{exp}	ρ_{sim}	S_{exp}	S_{sim}	H_{exp}	H_{sim}	G_{exp}	G_{sim}
				$P = 5$ bar				
90	1.377	1.375	1.420	1.442	−113.2	−113.8	−241.0	−243.5
95	1.344	1.342	1.477	1.494	−108.0	−108.0	−248.3	−249.9
100	1.311	1.308	1.532	1.543	−102.6	−102.3	−255.8	−256.6
				$P = 100$ bar				
90	1.405	1.403	1.391	1.414	−109.1	−109.6	−234.3	−236.8
95	1.376	1.374	1.446	1.462	−104.0	−104.0	−241.4	−243.0
100	1.347	1.344	1.498	1.509	−98.9	−98.6	−248.7	−249.6

We now analyze the results obtained for the solid phase. First, we show in Fig. 6 the variation of the interaction energy during a NPT simulation of a fcc crystal at $T = 80$ K and $P = 100$ bar. We observe a behavior similar to that of the liquid (shown in Fig. 5), but with a lower interaction energy as a result of the increased attractive interactions that takes place in the solid phase. Turning to the variations of entropy during the simulation, we also find that the entropy fluctuates around its average value. As expected, we find this average value to be lower than for the liquid, as a result of the greater amount of order found in the crystal.

As for the liquid, we perform a series of NPT simulations on the fcc phase for temperatures ranging from $T = 70$ K to $T = 80$ K and for pressures from $P = 1$ bar to $P = 100$ bar. The simulation results are given in Table II, together with the experimental data for the solid. The results for the densities show that there is an overall good agreement between the experimental data and the simulation results for the density, with deviations of less than 0.02 g/cm^3 for the state points studied here. In line with results for the liquid, we also obtain

FIG. 6. Monte Carlo NPT simulations of solid argon at $T = 80$ K and $P = 100$ bar. Same legend as in Fig. 5.

a good agreement for the thermodynamic functions, with the simulation results for the enthalpy exhibiting a deviation of up to 1% for the enthalpy, of up to 1.1% for the entropy, and up to 1% for the Gibbs free energy. This confirms that the entropy calculations carried out using the ML predictions for the Helmholtz free energy, together with Eq. (3), provide accurate results for both the liquid and the solid phases, as established through this comparison with the experimental data.

C. Crystal nucleation using entropy as reaction coordinate.

We finally turn to the last type of simulations carried out in this work, and perform NPT-S simulations of crystal nucleation in the supercooled liquid of Lennard-Jones particles at $T^* = 0.86$ and $P^* = 5.68$. We start from the supercooled liquid, for which we obtain an average density of $\rho^* = 0.989$. Using the ANN trained in the first part of this work, together with Eq. (1), we have a Helmholtz free energy $A^* = -10.31$, which yields an entropy $S^* = 5.71$ through Eq. (3). From there, we start carrying out NPT-S simulations, through umbrella sampling, to study the onset of crystal nucleation. We perform 13 successive simulations with overlapping windows, each window being associated with a decreasing value for the target entropy. We plot in Fig. 7 the variation of the density during the successive NPT-S simulations, with each data point on the graph corresponding to the average density for a simulation window. We find that, as entropy gradually decreases, the density of the system steadily increases, with an overall increase amounting to 1.3% over the entire range of entropies sampled. We also show in Fig. 7 how the interaction energy is impacted by the decrease in entropy. We observe that the interaction energy steadily decreases with entropy, resulting in a overall change of -2.8% over the whole process. The decrease in interaction energy, combined with an increase in density, point toward a solidification of the system. Furthermore, to characterize the order within the system, we calculate the value taken by the Steinhardt order parameter [62] Q_6 over each window. We plot the evolution of Q_6 as the function of the decreasing entropy in Fig. 7 and find that Q_6 increases as entropy decreases, indicating that crystalline order starts to develop within the system, consistently with the onset of crystal nucleation.

Using the umbrella sampling results, we calculate the free energy profile for the entire process and present it in Fig. 8(a).
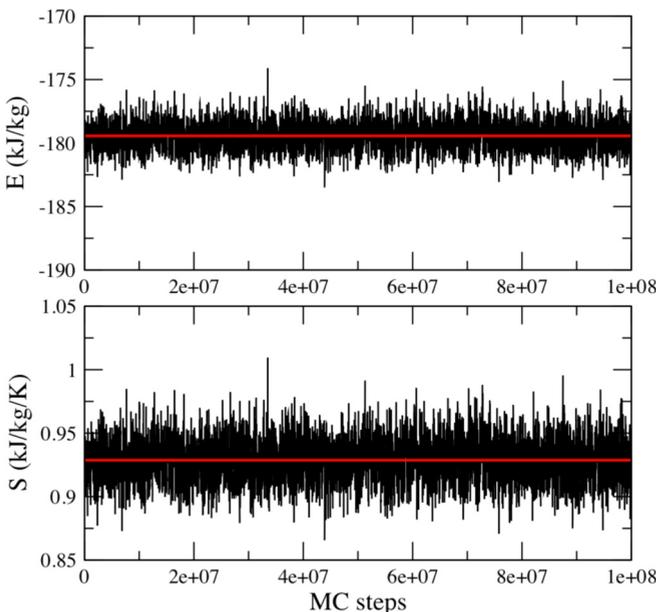
TABLE II. Solid argon: Comparison between the experimental data [89] and the simulation results (this work) for various state points. Temperatures are given in K, ρ in g/cm^3, S in kJ/(kg K), and H and G in kJ/kg (standard deviations are of the same order as for the liquid).

T	ρ_{exp}	ρ_{sim}	S_{exp}	S_{sim}	H_{exp}	H_{sim}	G_{exp}	G_{sim}
				$P = 1\,\mathrm{bar}$				
70	1.664	1.681	0.825	0.823	−160.4	−161.9	−218.2	−219.5
75	1.649	1.662	0.874	0.886	−156.5	−157.6	−222.1	−224.1
80	1.633	1.642	0.932	0.950	−152.4	−153.0	−227.0	−229.0
				$P = 100\,\mathrm{bar}$				
70	1.675	1.693	0.815	0.806	−155.2	−156.9	−212.2	−213.3
75	1.661	1.676	0.867	0.867	−151.4	−152.7	−216.4	−217.4
80	1.645	1.656	0.919	0.929	−147.4	−148.3	−220.9	−222.6

The free energy profile shows that the starting point for the nucleation process is the metastable supercooled liquid with $S^* = 5.71$, which is associated with a local minimum in free energy, as shown in Fig. 8(a). Then, as entropy decreases, the organization within the system starts to develop. During this stage, we observe that the free energy increases, until the top of the free energy barrier is reached for $S^* = 5.48$. At this point, a critical nucleus has formed. We show in Fig. 8(b) a snapshot of the critical nucleus, surrounded by the liquid (solidlike particles are identified using the same criteria as in previous work [61,76,78]). Looking now at the height of the

(a)

(b)

FIG. 8. (a) Crystal nucleation of the Lennard-Jones system at $T^* = 0.86$ and $P^* = 5.68$: Gibbs free energy of nucleation as a function of the reduced entropy S^*. (b) Snapshot of a crystal nucleus (in red) of a critical size formed within a supercooled liquid (in cyan) of Lennard-Jones particles.
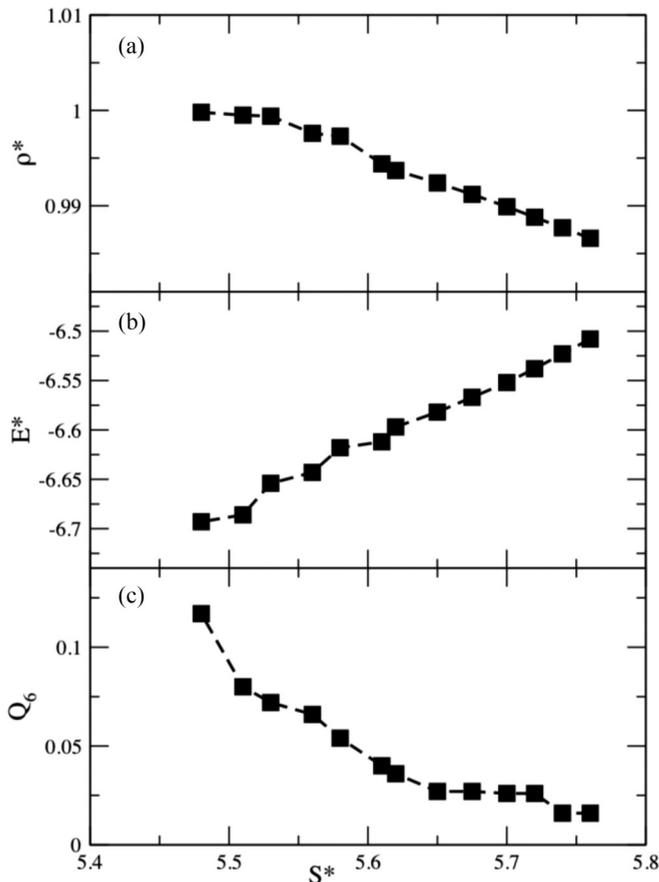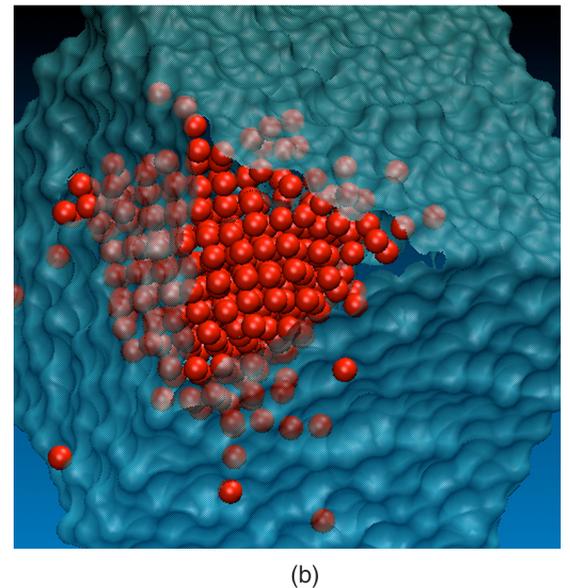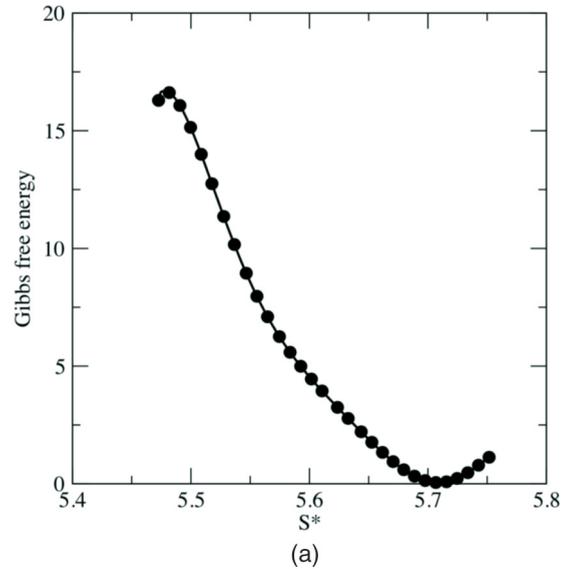

FIG. 7. Variation of (a) the reduced density, (b) the reduced interaction energy, and (c) Q_6 as a function of the reduced entropy for the successive NPT-S simulations.

free energy barrier obtained for this process, we find that it is of $17 \pm 2\, k_B T$. This is consistent with the results from prior simulations on the Lennard-Jones system [61], which estimate a barrier of roughly $20\, k_B T$ for the supercooling considered in this work. These results show that our approach, which relies on the use of entropy as a reaction coordinate, can indeed be used to simulate the onset of crystal nucleation within a supercooled liquid. One of the advantages of using entropy as a reaction coordinate is that it is not structure specific. As such, it does not impose any bias on the type of crystal structure, or polymorph, formed during the crystal nucleation process, either for precritical nuclei and or for the structure of the crystal nucleus. This is of key importance, and in turn, makes the approach implemented here especially promising, since a complete understanding of the polymorph selection process still remains elusive and is one of the outstanding issues in solid-state physics and chemistry [90]. In future work, we will examine this point and compare the structural features of crystallites obtained along the nucleation pathway using either entropy or geometric order parameters, such as, e.g., Q_6.

IV. CONCLUSIONS

In this work, we use machine learning to infer the canonical partition function and the Helmholtz free energy of a system. This, in turn, allows us to calculate its entropy using molecular simulation. This is achieved by training an artificial neural network using, as a training dataset, previously published equations of state for the Lennard-Jones system for both the fluid and solid phases. Once the ANN has been trained, we are able to evaluate the Helmholtz free energy as a function of density and temperature. Our approach is validated by testing the ML predictions over a wide range of thermodynamic conditions. In particular, these predictions for the Helmholtz free energy lead us to calculate the canonical partition function, thereby providing a direct comparison with results from recent flat-histogram simulations on the Lennard-Jones system. The second stage of this work consists of using these ML predictions to calculate the entropy of the system. This is validated by performing MC simulations under NPT conditions. The simulation results are found to be in very good agreement with experimental data on liquid and solid argon, a system that is very well modeled by the Lennard-Jones potential. Finally, we use entropy as a reaction coordinate in umbrella sampling simulations and show that a gradual decrease in entropy results in the formation of a crystal nucleus within a supercooled liquid. The analysis of the crystal nucleation process along the entropic pathway is found to be consistent with the results from previous simulation studies. This approach can be readily extended to nucleation processes for molecular systems and complex fluids and is also especially promising for the study of entropy-driven organization processes. For such systems, the method developed in this work requires the prior knowledge of the canonical partition function for the liquid and solid phases, or alternatively of the dependence of the Helmholtz free energy on temperature and density. Methods have been recently developed to this end [91–99], opening the door for applications of this approach to realistic systems.

[1] B. Tidor and M. Karplus, J. Mol. Biol. **238**, 405 (1994).

[2] C. Singh, P. K. Ghorai, M. A. Horsch, A. M. Jackson, R. G. Larson, F. Stellacci, and S. C. Glotzer, Phys. Rev. Lett. **99**, 226106 (2007).

[3] J. L. MacCallum and D. P. Tieleman, J. Am. Chem. Soc. **128**, 125 (2006).

[4] S. Whitelam, Soft Matter **11**, 8225 (2015).

[5] F. A. Escobedo, J. Chem. Phys. **146**, 134508 (2017).

[6] L. R. Nemzer, J. Theor. Biol. **415**, 158 (2017).

[7] C. Angell, J. Res. Natl. Inst. Stand. Technol. **102**, 171 (1997).

[8] F. W. Starr, C. A. Angell, and H. E. Stanley, Physica A (Amsterdam, Neth.) **323**, 51 (2003).

[9] L. Berthier, P. Charbonneau, D. Coslovich, A. Ninarello, M. Ozawa, and S. Yaida, Proc. Natl. Acad. Sci. USA **114**, 11356 (2017).

[10] L. Yan, Nat. Commun. **9**, 1359 (2018).

[11] M. Schenk, S. Vidal, T. Vlugt, B. Smit, and R. Krishna, Langmuir **17**, 1558 (2001).

[12] R. Krishna, B. Smit, and S. Calero, Chem. Soc. Rev. **31**, 185 (2002).

[13] J. Rubí, A. Lervik, D. Bedeaux, and S. Kjelstrup, J. Chem. Phys. **146**, 185101 (2017).

[14] Z. Sun, Y. N. Yan, M. Yang, and J. Z. Zhang, J. Chem. Phys. **146**, 124124 (2017).

[15] J. A. Caro, K. W. Harpole, V. Kasinath, J. Lim, J. Granja, K. G. Valentine, K. A. Sharp, and A. J. Wand, Proc. Natl. Acad. Sci. USA **114**, 6563 (2017).

[16] D. Frenkel, Nat. Mater. **14**, 9 (2014).

[17] G. Odriozola and M. Lozada-Cassou, J. Phys. Chem. B **120**, 5966 (2016).

[18] R. Zhang, B. Lee, C. M. Stafford, J. F. Douglas, A. V. Dobrynin, M. R. Bockstaller, and A. Karim, Proc. Natl. Acad. Sci. USA **114**, 2462 (2017).

[19] R. V. Thaner, Y. Kim, T. I. Li, R. J. Macfarlane, S. T. Nguyen, M. Olvera de la Cruz, and C. A. Mirkin, Nano Lett. **15**, 5545 (2015).

[20] N. C. Karayiannis, K. Foteinopoulou, and M. Laso, Phys. Rev. Lett. **103**, 045703 (2009).

[21] Z. S. Derewenda and P. G. Vekilov, Acta Crystallogr. D **62**, 116 (2006).

[22] P. M. Piaggi, O. Valsson, and M. Parrinello, Phys. Rev. Lett. **119**, 015701 (2017).

[23] G. Gobbo, M. A. Bellucci, G. A. Tribello, G. Ciccotti, and B. L. Trout, J. Chem. Theory Comput. **14**, 959 (2018).

[24] N. C. Karayiannis, K. Foteinopoulou, and M. Laso, Soft Matter **11**, 1688 (2015).

[25] L. Zha, M. Zhang, L. Li, and W. Hu, J. Phys. Chem. B **120**, 12988 (2016).

[26] C. Desgranges and J. Delhommelle, J. Am. Chem. Soc. **136**, 8145 (2014).

[27] C. Desgranges and J. Delhommelle, J. Am. Chem. Soc. **133**, 2872 (2011).

[28] S. Lee, M. Engel, and S. Glotzer, Bull. Am. Phys. Soc. **K57**, 00005 (2018).

[29] C. Desgranges and J. Delhommelle, J. Chem. Phys. **145**, 204112 (2016).

[30] C. Desgranges and J. Delhommelle, J. Chem. Phys. **145**, 234505 (2016).

[31] A. Vishnyakov and A. V. Neimark, J. Chem. Phys. **119**, 9755 (2003).

[32] R. C. Remsing, E. Xi, S. Vembanur, S. Sharma, P. G. Debenedetti, S. Garde, and A. J. Patel, Proc. Natl. Acad. Sci. USA **112**, 8181 (2015).

[33] C. Desgranges and J. Delhommelle, J. Chem. Phys. **146**, 184104 (2017).

[34] G. Menzl and C. Dellago, J. Chem. Phys. **145**, 211918 (2016).

[35] H. Meirovitch, Curr. Opinion Struct. Biol. **17**, 181 (2007).

[36] M. Karplus and J. N. Kushick, Macromolecules **14**, 325 (1981).

[37] H. Schäfer, A. E. Mark, and W. F. van Gunsteren, J. Chem. Phys. **113**, 7809 (2000).

[38] M. S. Shell, J. Chem. Phys. **129**, 144108 (2008).

[39] T. T. Foley, M. S. Shell, and W. Noid, J. Chem. Phys. **143**, 243104 (2015).

[40] N. Lu and D. A. Kofke, J. Chem. Phys. **114**, 7303 (2001).

[41] R. P. White and H. Meirovitch, Proc. Natl. Acad. Sci. USA **101**, 9235 (2004).

[42] S. Cheluvaraja and H. Meirovitch, Proc. Natl. Acad. Sci. USA **101**, 9241 (2004).

[43] Z. Preisler and M. Dijkstra, Soft Matter **12**, 6043 (2016).

[44] J. K. Singh, D. A. Kofke, and J. R. Errington, J. Chem. Phys. **119**, 3405 (2003).

[45] M. J. Pond, J. R. Errington, and T. M. Truskett, J. Chem. Phys. **134**, 081101 (2011).

[46] C. Desgranges and J. Delhommelle, J. Chem. Phys. **136**, 184107 (2012).

[47] C. Desgranges and J. Delhommelle, J. Chem. Phys. **136**, 184108 (2012).

[48] A. Baranyai and D. J. Evans, Phys. Rev. A **40**, 3817 (1989).

[49] N. Jakse and A. Pasturel, Sci. Rep. **6**, 20689 (2016).

[50] S.-T. Lin, M. Blanco, and W. A. Goddard III, J. Chem. Phys. **119**, 11792 (2003).

[51] S.-T. Lin, P. K. Maiti, and W. A. Goddard III, J. Phys. Chem. B **114**, 8191 (2010).

[52] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012).

[53] A. P. Bartók, M. J. Gillan, F. R. Manby, and G. Csányi, Phys. Rev. B **88**, 054104 (2013).

[54] P. J. Ballester and J. B. Mitchell, Bioinformatics **26**, 1169 (2010).

[55] A. Z. Guo, E. Sevgen, H. Sidky, J. K. Whitmer, J. A. Hubbell, and J. J. de Pablo, J. Chem. Phys. **148**, 134108 (2018).

[56] H. Sidky and J. K. Whitmer, J. Chem. Phys. **148**, 104111 (2018).

[57] R. A. Mansbach and A. L. Ferguson, J. Chem. Phys. **142**, 105101 (2015).

[58] J. K. Johnson, J. A. Zollweg, and K. E. Gubbins, Mol. Phys. **78**, 591 (1993).

[59] M. A. Van der Hoef, J. Chem. Phys. **113**, 8142 (2000).

[60] N. B. Vargaftik, Y. K. Vinoradov, and V. S. Yargin, *Handbook of Physical Properties of Liquids and Gases* (Begell House, New York, 1996).

[61] P. Rein ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, J. Chem. Phys. **104**, 9932 (1996).

[62] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Phys. Rev. B **28**, 784 (1983).

[63] C. Desgranges and J. Delhommelle, J. Chem. Phys. **149**, 044118 (2018).

[64] J. Behler, J. Chem. Phys. **145**, 170901 (2016).

[65] P. Geiger and C. Dellago, J. Chem. Phys. **139**, 164105 (2013).

[66] P. Koß, A. Statt, P. Virnau, and K. Binder, Phys. Rev. E **96**, 042609 (2017).

[67] J. R. Errington, Phys. Rev. E **67**, 012102 (2003).

[68] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).

[69] P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, J. Chem. Phys. **110**, 1591 (1999).

[70] M. J. McGrath, J. N. Ghogomu, N. T. Tsona, J. I. Siepmann, B. Chen, I. Napari, and H. Vehkamäki, J. Chem. Phys. **133**, 084106 (2010).

[71] C. Desgranges and J. Delhommelle, Chem. Phys. Lett. **695**, 194 (2018).

[72] E. Xi, R. C. Remsing, and A. J. Patel, J. Chem. Theory Comput. **12**, 706 (2016).

[73] S. Auer and D. Frenkel, Nature (London) **713**, 411 (2001).

[74] R. Radhakrishnan and B. L. Trout, J. Am. Chem. Soc. **125**, 7743 (2003).

[75] R. Blaak, S. Auer, D. Frenkel, and H. Löwen, Phys. Rev. Lett. **93**, 068303 (2004).

[76] C. Desgranges and J. Delhommelle, J. Am. Chem. Soc. **128**, 10368 (2006).

[77] C. Desgranges and J. Delhommelle, J. Am. Chem. Soc. **128**, 15104 (2006).

[78] C. Desgranges and J. Delhommelle, Phys. Rev. Lett. **98**, 235502 (2007).

[79] C. Desgranges and J. Delhommelle, J. Am. Chem. Soc. **129**, 7012 (2007).

[80] S. Punnathanam and P. Monson, J. Chem. Phys. **125**, 024508 (2006).

[81] E. Sanz, C. Valeriani, D. Frenkel, and M. Dijkstra, Phys. Rev. Lett. **99**, 055501 (2007).

[82] P. Yi and G. C. Rutledge, J. Chem. Phys. **131**, 134902 (2009).

[83] L. Filion, M. Hermes, R. Ni, and M. Dijkstra, J. Chem. Phys. **133**, 244115 (2010).

[84] C. Desgranges and J. Delhommelle, Phys. Rev. Lett. **120**, 115701 (2018).

[85] S. Ketzetzi, J. Russo, and D. Bonn, J. Chem. Phys. **148**, 064901 (2018).

[86] S. Prestipino, J. Chem. Phys. **148**, 124505 (2018).

[87] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, UK, 1987).

[88] S. Auer and D. Frenkel, Nature (London) **409**, 1020 (2001).

[89] V. A. Rabinovich, A. Vasserman, V. Nedostup, and L. S. Veksler, *Thermophysical Properties of Neon, Argon, Krypton, and Xenon*, National Standard Reference Data Service of the USSR Vol. 10, (Hemisphere, Washington, DC, 1988).

[90] J. Bernstein, *Polymorphism in Molecular Crystals* (Clarendon Press, Oxford, UK, 2002).

[91] D. Frenkel and A. J. Ladd, J. Chem. Phys. **81**, 3188 (1984).

[92] B. B. Laird and A. Haymet, Mol. Phys. **75**, 71 (1992).

[93] L. A. Báez and P. Clancy, Mol. Phys. **86**, 385 (1995).

[94] P. Monson and D. Kofke, Adv. Chem. Phys. **115**, 113 (2000).

[95] J. R. Errington, J. Chem. Phys. **120**, 3130 (2004).

[96] E. A. Mastny and J. J. de Pablo, J. Chem. Phys. **122**, 124109 (2005).

[97] T.-Q. Yu and M. E. Tuckerman, Phys. Rev. Lett. **107**, 015701 (2011).

[98] S. O. Nielsen, J. Chem. Phys. **139**, 124104 (2013).

[99] C. Desgranges and J. Delhommelle, J. Chem. Phys. **144**, 124510 (2016).