Analyzing and Driving Cluster Formation in Atomistic Simulations

Analyzing and Driving Cluster Formation in Atomistic Simulations

Gareth A. Tribello,[*,†] Federico Giberti,[‡] Gabriele C. Sosso,[¶] Matteo Salvalaglio,[§] and Michele Parrinello[‡]

Atomistic Simulation Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast, BT7 1NN, Computational Science, Department of Chemistry and Applied Biosciences, ETH Zurich, USI-Campus, Via Giuseppe Buffi 13, C-6900 Lugano, Switzerland, Thomas Young Centre, London Centre for Nanotechnology and Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom, and Department of Chemical Engineering, University College London,Torrington Place,London WC1E 7JE,United Kingdom

E-mail: g.tribello@qub.ac.uk

Abstract

In this paper a set of computational tools for identifying the phases contained in a system composed of atoms/molecules is introduced. The method is rooted in graph theory and combines atom centered symmetry functions, adjacency matrices

---

[*]To whom correspondence should be addressed
[†]Atomistic Simulation Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast, BT7 1NN
[‡]Computational Science, Department of Chemistry and Applied Biosciences, ETH Zurich, USI-Campus, Via Giuseppe Buffi 13, C-6900 Lugano, Switzerland
[¶]Thomas Young Centre, London Centre for Nanotechnology and Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom
[§]Department of Chemical Engineering, University College London,Torrington Place,London WC1E 7JE,United Kingdom

and clustering algorithms to identify regions of space where the properties of the system constituents can be considered uniform. We show how this method can be used to define collective variables and how these collective variables can be used to enhance the sampling of nucleation events. We then show how this method can be used to analyze simulations of crystal nucleation and growth by using it to analyze simulations of the nucleation of the molecular crystal urea and simulations of nucleation in a semiconducting alloy. The semiconducting alloy example we discuss is particular challenging as multiple nucleation centers are formed. We show, however, that our algorithm is able to detect the grain boundaries in the resulting polycrystal.

1 Introduction

Many interesting phenomena in materials science occur at the interface between different phases. For example, when a material melts or freezes an interface between the solid and liquid phases must form. When describing this scenario at the macroscopic level the two phases are identified by finding the regions of space throughout which the physical properties of the system are uniform both in terms of structure and composition. As such, whenever a boundary between two phases is crossed, the local physical properties of the system change discontinuously. This description works well when the volumes of the two phases are large as the the ratio between the number of surface atoms and the number of bulk atoms is then guaranteed to be exceedingly small. When the extents of the two phases involved are on the order of nanometers, however, this assumption breaks down as the surface atoms now constitute a substantial part of the system. When studying problems at these tiny length scales atomistic molecular dynamics (MD) is an invaluable tool as it allows one to monitor the static and dynamic properties of the atoms/molecules. One difficulty with this method is extracting the pertinent information from the vast amount of data contained in a molecular dynamics trajectory. Doing this effectively requires the use of a certain degree chemical/physical intuition about the problem in question. So, for example, when studying

an interface between the solid and liquid phases of a material one would really like to begin by defining the extent of the two phases and the location of the interface between them. If, as in MD, one is given just the positions of all the atoms in the system this is a non-trivial problem, particularly for dense systems.

In this paper, we discuss a set of computational tools we have developed that can be used both to detect clusters of one phase within a second phase and to enhance the rate at which phase separation processes occur. Our method is based on a set of differentiable Collective Variables (CV) that have been implemented in PLUMED 2.0[1,2] and that are therefore freely available online. In what follows we first briefly discuss the essential theoretical background of our approach in section 2. We then demonstrate how these CVs can be used in enhanced sampling simulations to drive the condensation of a Lennard-Jones vapor into liquid droplets. Next, we show the various ways we have used these tools to quickly implement new CVs that can be used to analyze molecular dynamics trajectories in which phase separation events are observed. In the last of these examples, we demonstrate that these methods can even identify complex interfaces such as three-dimensional grain boundaries. This example is particularly problematic as the phases in this particular system possess the same crystalline structure with different orientations.

2   Methodology

As discussed in the introduction, a phase is a region of space in which local physical properties of the system are reasonably uniform. This uniformity occurs because the local symmetry around each of the atoms/molecules in any given phase is similar. There are differences in the symmetry around each of the molecules due to thermal motions, but as these fluctuations are small we can exploit the differences in the local symmetries around atoms/molecules in the various phases when designing CVs to understand nucleation. In what follows we will begin by discussing the formation of a solid from the melt in order to make our explanation

3

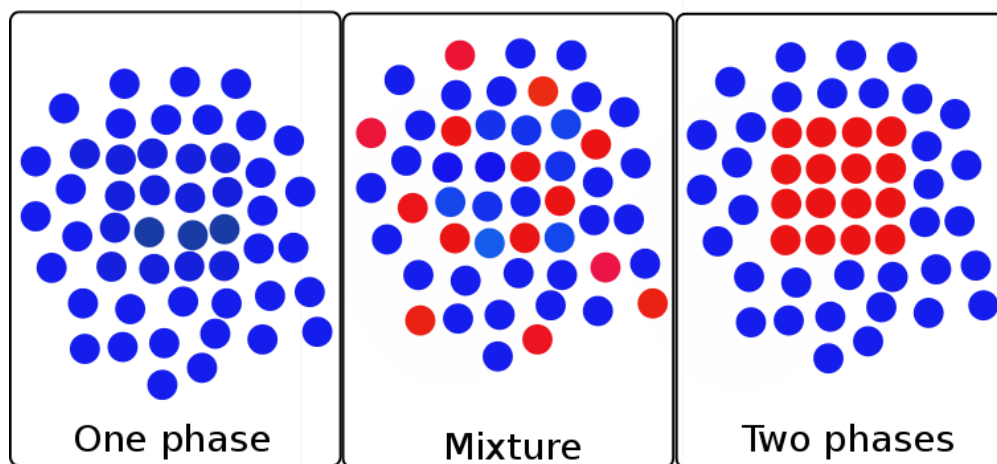clearer. We will then generalize the ideas and explain how these methods can be used in other contexts.

Figure 1: Figure illustrating what a CV that establishes the extent of a solid-portion of a two phase system must accomplish. In the figures above the red atoms differ from the blue atoms in some way. They red atoms might be a different elemental type to the blue atoms or they might just have different values for some symmetry function. As discussed in the text a good CV should be able to differentiate the red atoms from the blue atoms. It must then determine the largest subset of red atoms that are clustered together in a single continuous region. If the CV only does the first of these tasks then it is not guaranteed that it will distinguish the two phase system from the mixture shown above. Similarly if it only determines if atoms are in a contiguous region it cannot distinguish a one phase system from a two phase system.

Any CV that establishes the extent of the solid portion of a two phase system from the positions of all the constituent atoms/molecules must do two things:

1. It must establish which of the atoms are solid-like.

2. It must determine the subsets of atoms/molecules from this set of solid-like atoms that are clustered together into contiguous regions.

As shown in figure 1 if only the first of these steps is performed there is no guarantee that the CV will be able to differentiate between a true two phase system and a mixture. Within our CVs these two tasks are performed by two separate sets of functions. We determine the set of solid-like atoms by calculating atom-centered symmetry functions for each of the atoms

in our system. These symmetry functions are calculated based on the position of the central atom and the positions of the atoms its first few coordination spheres. Similar functions are used in a number of different computational frameworks, a particularly notable example being the high-dimensional neural networks devised by Behler *et al.*[3]

The simplest symmetry function is the coordination number, which is frequently calculated using:

$$c_i = \sum_{j \neq i} \frac{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^n}{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^m} \tag{1}$$

where r_{ij} is the distance between atom i and atom j and r_0, d_0, n and m are parameters that determine how rapidly this function decays to zero. More complicated symmetry functions include the cubic harmonics,[4,5] directional order parameters that measure the relative orientations of molecules,[6] the Steinhardt order parameters[7,8] and the local Steinhardt order parameters[8,9] all of which are available in PLUMED 2.0[2] and all of which can be used within our method.
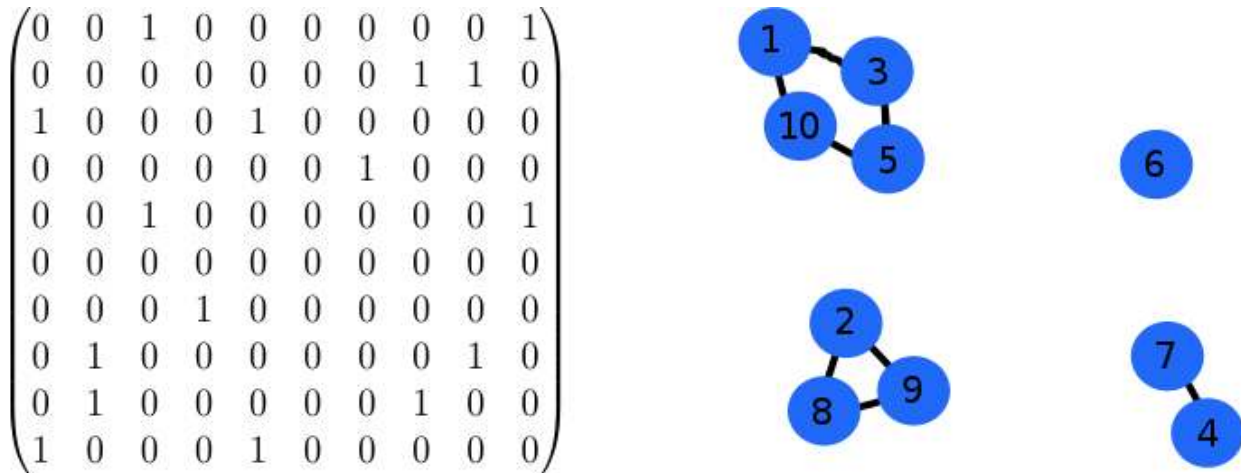
Figure 2: A figure illustrating how the graph reduction algorithms that are at the heart of this work operate. A 10×10 adjacency matrix is constructed based on the positions of the atoms shown in the left hand side. The i, j element of this matrix tells you whether or not atoms i and j are connected. The DFS algorithm takes this matrix as input and returns the list of atoms that are in each of the four connected components of the corresponding graph shown on the right.

The second element of our method is a graph reduction algorithm.[10] The input for this

algorithm is a symmetric adjacency matrix whose (i, j) element tells us whether or not the ith and jth particles are connected. The elements of this matrix can be calculated by using something as simple as a Heaviside step function acting on the distance, r_{ij}, between the centers of mass of molecules i and j. In general, however, because each of these elements equals either one or zero, we can illustrate the connectivity structure encoded in our matrix using a graph as shown in figure 2. There are numerous algorithms[11–14] that can then be used to divide the nodes of this graph into the subsets of connected components shown in figure 2. We choose to use the depth first search (DFS)[10] graph reduction algorithm although other clustering algorithms would work equally well.

The final step in our prescription is to calculate some function, g, of the sum of the values of the symmetry functions for the atoms in each of the connected components we find. In other words, we compute:

$$s = \sum_{i=1}^{M} g \left(\sum_{j=1}^{N_i} c_j \right) \tag{2}$$

where M is the number of connected components. The second sum runs over the N_i atoms in the ith connected component and c_j is the value of the symmetry function for the jth atom that forms part of the ith connected component. The value of s that we extract using this formula is not continuous. However, as discussed in the next section, we have found that we can use such quantities as the CV in umbrella sampling,[15] steered MD[16] or metadynamics[17,18] simulations and that we can use these variable to enhance the sampling of rare events such as nucleation.[19] The only caveat being that, when doing such calculations, special care is required as the system must be prevented from entering regions of phase space in which the CV defined by equation 2 changes discontinuously. With these technicalities aside we will, in the following sections, demonstrate how this method can be used to study nucleation of liquid from gas, nucleation of crystals from solution and to study nucleation from the melt.

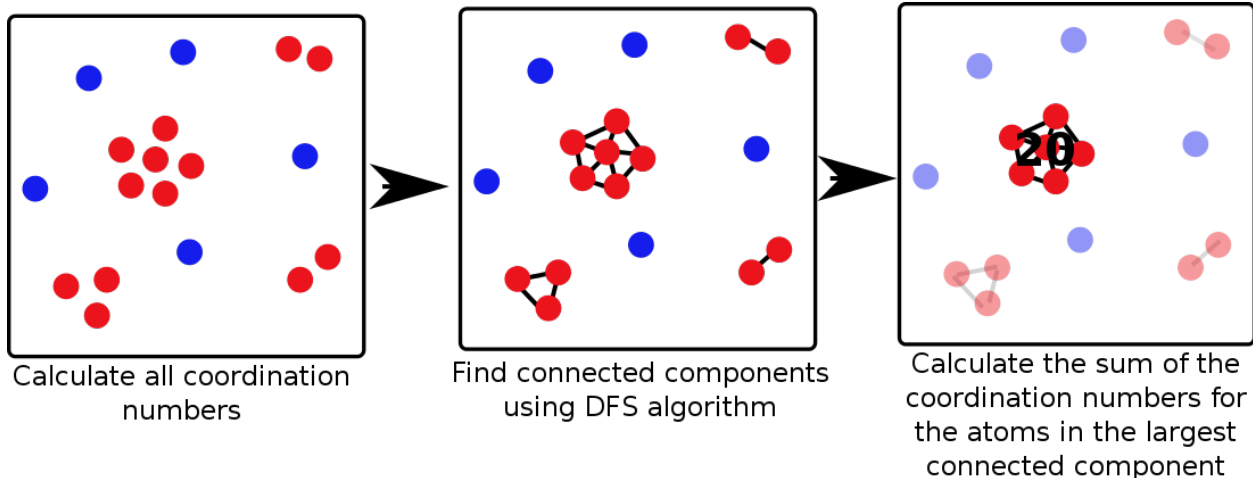# 3 Driving nucleation of liquid droplets from gas

Figure 3: Figure showing the manner in which the CV is calculated in section 3. The first step involves computing the coordination numbers of all the atoms in the system. For clarity, we have colored the atoms with a high coordination number in red in this figure. In the second step, an adjacency matrix is constructed in which element i, j tells you whether or not atoms i and j are within some cutoff distance. The connected components in the graph corresponding to this matrix are then computed using the DFS algorithm. In the final step, only the atoms in the largest connected component are considered and the sum of the coordination numbers for these atoms is evaluated.

The first transformation we chose to analyze with our method is the nucleation of argon droplets from supersaturated argon vapor. This phenomenon has been simulated using the Lennard-Jones potential in previous works[20] and the coordination number of the atoms has been used to distinguish those atoms in the gaseous phase from those in the nascent droplet.[21] In this work, we thus chose to model this process using the Lennard Jones potential that was used in these previous work and to use the coordination number as the symmetry function in the scheme in section 2. The (i, j) element of the adjacency matrix, M_{ij}, was calculated using the following function:

$$
M_{ij} = \begin{cases} 1 & \text{for} \quad r_{ij} < 4.5 \\ [r_{ij} - 5.5]^2 \, [r_{ij} - 3.5] & \text{for} \quad 4.5 \leq r_{ij} \leq 5.5 \\ 0 & \text{otherwise} \end{cases} \tag{3}
$$

where r_{ij} is the distance between atom i and atom j. The function above goes smoothly to zero at $r_{ij} = 5.5$ Å. We thus state that atoms i and j are only connected if they are within 5.5 Å of each other. As illustrated in the middle panel of figure 3 we can analyze this adjacency matrix using the DFS algorithm, find the connected components and thus identify any droplets of liquid that have nucleated in our simulation. Furthermore, if we take the sum of any row of the above adjacency matrix, $\sum_{i=1}^{N} = M_{ij}$, we get the coordination number, c_i, for the ith atom. Our final CV in this section is thus the sum of the coordination numbers for the atoms in the largest cluster:

$$s = \sum_{i=1}^{N_1} \sum_{j=1}^{N} M_{ij} \tag{4}$$

where, to reiterate, the first sum here runs over those atoms in the largest cluster identified by the DFS algorithm and the second sum runs over all the atoms in the system.

When evaluating the adjacency matrix, we use the continuous function described in equation 3 rather than a simpler Heavyside function because we would like to be able to calculate derivatives for the final collective variable s. We need these derivatives because we would like to use a bias that is a function of this coordinate to accelerate sampling. This bias will introduce additional forces on the atoms the values of which will depend on the derivative of the CV with respect to the atomic positions. The fact that our CV is not differentiable is thus a concern. However, we have found that these issues resolve themselves if the parameters are set sensibly. In particular, when using derivatives, it is important to ensure that a smooth and differentiable function is used within the definition of the symmetry function. If this is done, and if the cutoff that determines whether or not two atoms are connected for the DFS algorithm is set at a value where this continuous switching function has decayed to zero, the value of the final function will not change discontinuously when one further atoms is added to the cluster. Even if these cutoffs are set in this way the value of the function will still change discontinuously if two smaller clusters merge to

form one much larger cluster. Fortunately, such events were not observed during the liquid nucleation trajectories described in this section. This not surprising because, at these low supersaturations, classical nucleation theory predicts that the concurrent formation of two large nuclei is extremely unlikely.

As discontinuous changes in the value of equation 4 are not observed we can safely use this function as a collective variable in a metadynamics simulation.[17] This method uses a history-dependent bias to force droplets to grow and dissolve rapidly. Furthermore, at the end of a metadynamics simulation, we can recover the free energy as a function of our collective variable from the bias.[17,22] The system we investigated was composed of 100 Argon atoms at 80.7 K. The temperature was kept constant using the velocity-rescale thermostat[23] and the volume was fixed at 180.36 nm^3. The equations of motion were propagated with a 2 fs time-step for 200 ns using gromacs-4.6.5[24–27] and the following PLUMED input file:

```
lq: COORDINATIONNUMBER SPECIES=1-100 SWITCH={CUBIC D_0=0.45  D_MAX=0.55} LOWMEM

cm: CONTACT_MATRIX ATOMS=lq  SWITCH={CUBIC D_0=0.45  D_MAX=0.55}

dfs: DFSCLUSTERING MATRIX=cm

clust1: CLUSTER_PROPERTIES CLUSTERS=dfs CLUSTER=1 SUM

mt: METAD ...
 ARG=clust1.sum SIGMA=10. HEIGHT=2. PACE=500
 TEMP=80.7 BIASFACTOR=50 GRID_MIN=0 GRID_MAX=6000
... mt: METAD
```

Figure 4 shows that the CV is able to drive a cluster with a larger number of bonds to form and dissolve multiple times over the course of this short simulation. As clusters form and dissolve a large number of times during the metadynamics simulation we are able to say with confidence that any estimate of the free energy we extract from this simulation will be converged. We therefore calculated the estimate of the free energy as a function of the number of atoms in the largest of the clusters that was identified using the DFS algorithm that is shown in figure 5 using the reweighting algorithm developed by Bonomi *et al.*[28] We
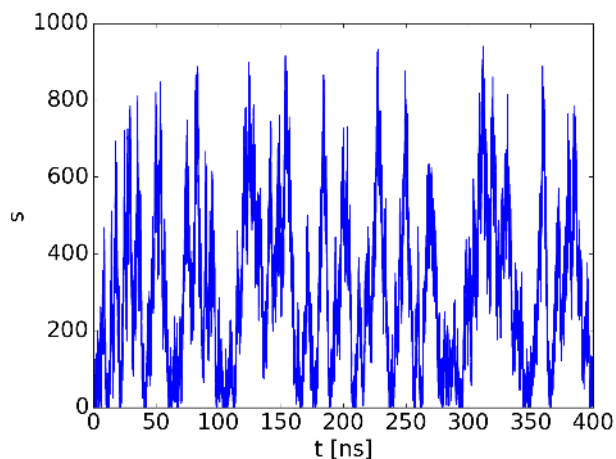
9

Figure 4: Behavior of the CV employed to drive the nucleation of a droplet in a LJ gas as a function of the simulation time.
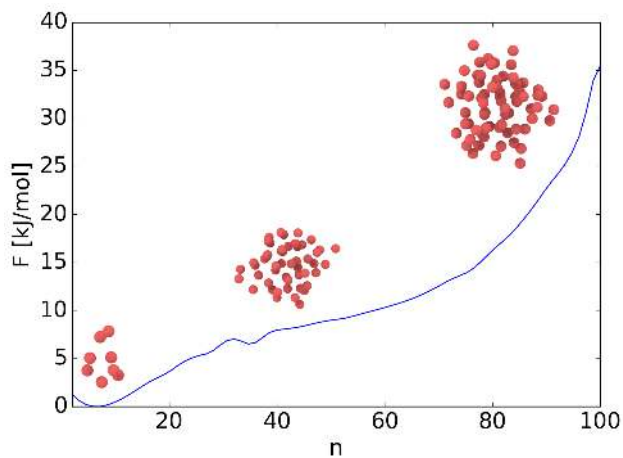

Figure 5: The free energy for a supersaturated Lenard Jones gas as a function of the number of atoms in the liquid nucleus. Snapshots of the bubble obtained during the metadynamics calculations are also shown.

can calculate the quantity on the x-axis of this figure by adding the following instructions to the PLUMED input that was used for the metadynamics simulation and which was given above:

```
ss: CLUSTER_NATOMS CLUSTERS=dfs CLUSTER=1
PRINT ARG=ss FILE=colvar
```

~~This converged free energy surface is shown in figure 5.~~ It is important to note that we have to use reweighting to extract the free energy as a function of cluster size as we cannot use the number of atoms in the largest cluster as a CV for a metadynamics simulation as this quantity can change discontinuously.

Our metadynamics simulations of the nucleation of the droplets were run with a constant number of atoms, a constant temperature, and a constant volume so there is a coupling between the size of the droplet and the chemical potential of the surrounding gas.[29–32] The free energy shown in figure 5 thus increases as the cluster grows because of this finite size effect.

A visual inspection of the trajectories led us to believe that clusters became more spherical as they grow. To investigate this more thoroughly we introduce the shape anisotropy k as a measure of the sphericity of the clusters:

$$k = \frac{3}{2}\frac{(\lambda_1^4 + \lambda_2^4 + \lambda_3^4)}{(\lambda_1^2 + \lambda_1^2 + \lambda_1^2)^2} - \frac{1}{2} \tag{5}$$

To compute k we evaluated the inertia tensor for the largest cluster of atoms in our system:

$$T_{\alpha\beta} = \sum_{i=1}^{N_1} m_i(x_{\alpha\beta}^{(i)} - \hat{x}_{\alpha\beta})$$

In this formula the sums run over the atoms in the largest cluster and m_i, $x^{(i)}$ and $\hat{x}$ are used to denote the mass of the ith atom in this cluster, the position of this atom and the position of the center of mass for the atom in the cluster. $T_{\alpha\beta}$ is a symmetric, 3×3 matrix

11

and the λ_1, λ_2 and λ_3 values in equation 5 are the eigenvalues of this matrix. Consequently, k is a unit-less quantity that takes a value between 0 and 1. This quantity is zero when the cluster is spherical and one when the atoms are arranged in a line.[33]

The free energy surface shown in figure 6 was again calculated using the reweighting algorithm developed by Bonomi *et al.*[28] The free energy surface here is shown as a function of the number of atoms in the largest cluster, n, and the shape anisotropy, k. Figure 6 shows clearly that small clusters are very flexible and that for these small clusters there is no barrier for going from a spherical shape with $k \approx 0$ to a less isotropic shape with with $k \approx 0.25$. A visual inspection of the trajectory shows that these anisotropic clusters have a flattened, pancake-like shape with $\lambda_1 < \lambda_2 \approx \lambda_3$. When the droplets grow larger these non-spherical shapes disappear and the droplet becomes more spherical so as to minimize the surface-volume ratio. This behavior at large sizes is completely in accordance with the predictions of classical nucleation theory. Furthermore, the fact that the smallest nuclei are non-spherical is similar to the behavior that has been observed in simulations of the nucleation of crystalline Lennard-Jones nuclei from the melt.[34]

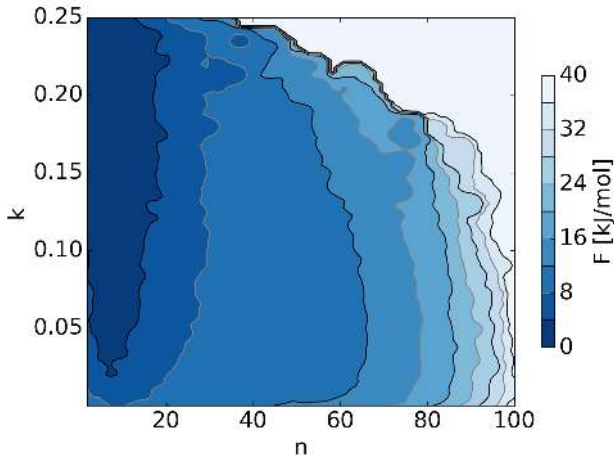Figure 6: The free energy surface for droplet formation from the gas as a function of the size of the cluster n and the shape anisotropy k. This free energy surface was obtained by reweighting the biased metadynamics trajectory. It is clear from this figure that small droplets are flexible and can adopt a range of shapes. However, as the droplets grow this flexibility disappears and the clusters become spherical.
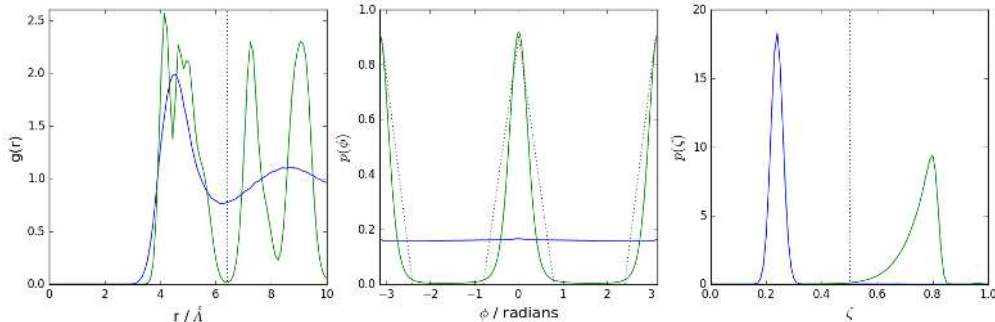
4 Nucleation of solid urea from the melt

Figure 7: Histograms of symmetry functions for pure crystalline urea (green lines) and liquid urea (blue lines). These probability densities were calculated from 300 ns long trajectories of 300 urea molecules, which were run at 450 K. The left-most panel shows the radial distribution function for the carbon atoms of urea together with a dashed line that indicates where the first coordination sphere is considered to end in this work. The central panel shows the histograms of torsional angles between the vectors connecting the carbon and oxygen atoms on neighboring urea molecules. The dashed line in this figure shows the shape of the function, $K(\phi_{ij})$, that we use to determine whether or not neighboring urea molecules are arranged as they would be in the solid. The rightmost figure shows a histogram for the quantity ζ that is defined in equation 7 and that was used to distinguish those molecules in the solid from those in the liquid. The dashed line in this final figure shows where we locate the dividing surface between the solid and the liquid in this work.

For this second example we chose to re-examine the well-tempered metadynamics trajectories from a recent article on the formation of solid urea from the melt.[35] The particular simulations we chose were the ones performed in the NVT ensemble on a system of 300 urea molecules at the melting temperature. The methods described in the previous section can be used when the atoms or molecules come closer together when nucleating the new phase. They can thus be applied to study the formation of liquid droplets from vapor or to examine the precipitation of a solute from solution.[36–40] They are clearly not appropriate when examining nucleation of crystals from the melt, however, because, as can be seen clearly in the radial distribution functions shown in the left-most panel of figure 7, the molecules in the new (solid) phase are no closer together than they were in the liquid.

The difference between the solid and liquid phases is illustrated in the central panel of figure 7 which shows histograms of torsional angles between the vectors connecting the
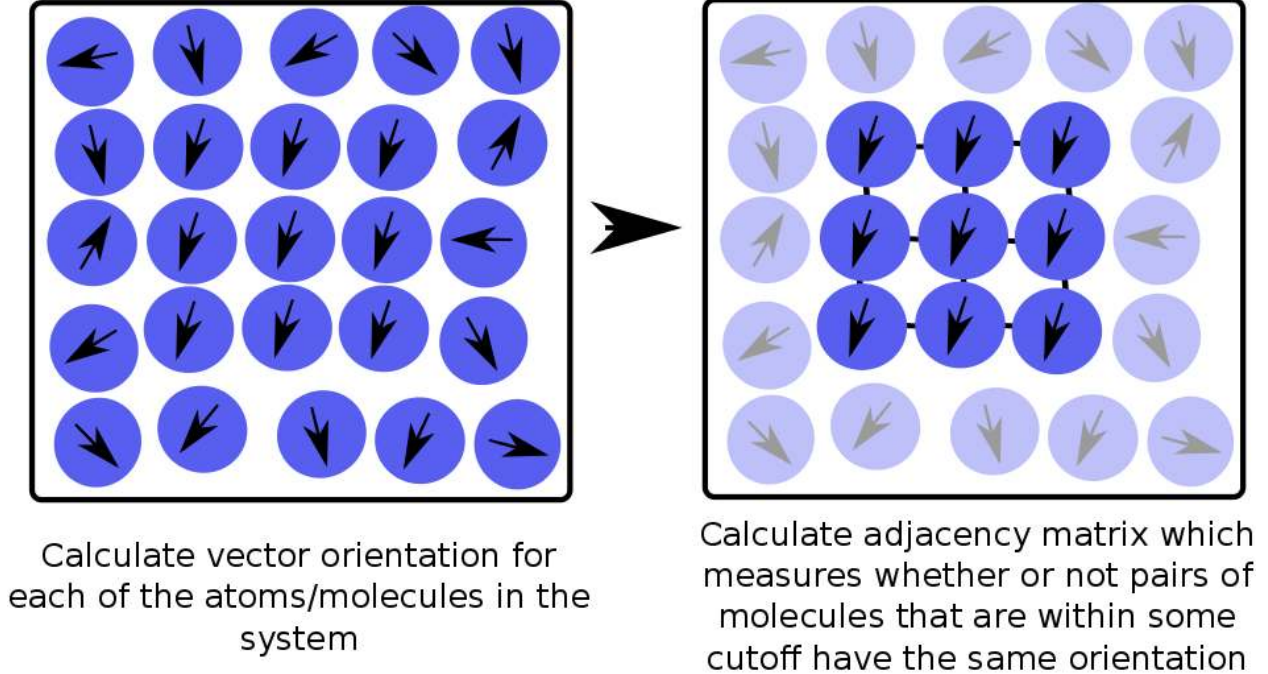
Figure 8: Figure illustrating the manner in which the CV is calculated in the early parts of section 4. Each of the blue circles in the panel on the left hand side represents a molecule. The black arrows represent the orientations of the molecules, which can be calculated using PLUMED. In any solid portions of the system the molecules will be aligned and hence these orientation vectors will point in the same direction. As shown in the second panel we can construct an adjacency matrix with elements that measure whether or not the molecules that are within a cutoff distance of each other have the same orientation. This matrix can then be inserted into the DFS clustering algorithm and the crystalline regions can thus be discovered.

carbon and oxygen atoms on all urea molecules whose central carbon atoms are within 6.4 Å of each other. In the solid phase this distribution has peaks around 0 and π radians indicating that adjacent urea molecules are aligned. In the liquid by contrast the distribution of angles is close to uniform and the structure is thus less ordered. These observations suggest that solid clusters forming from the melt can be identified using a strategy like that shown in figure 8. Essentially the elements of the adjacency matrix are calculated using:

$$A_{ij} = \sigma(r_{ij})K(\phi_{ij}) \qquad \text{where} \qquad K(\phi_{ij}) = \left[1 - \left|\frac{\phi_{ij}}{b_1}\right|\right] + \left[1 - \left|\frac{\phi_{ij} - \pi}{b_2}\right|\right] \qquad (6)$$

Here $\sigma(r_{ij})$ is a switching function on the distance between the carbon atoms of molecules i

and j that is one when $r_{ij} < 6.4$ and zero otherwise. ϕ_{ij}, meanwhile, is the torsional angle between the vectors connecting the carbon atoms to the oxygen atoms in urea molecules i and j. The function $K(\theta_{ij})$ is shown as a dashed line in the middle panel of figure 7 - the parameters b_1 and b_2 were set equal to 0.8 and 0.7 throughout this work. $K(\theta_{ij})$ converts torsional angles close to zero and π to a number close to one, while converting other values of the torsion to numbers close to zero. Consequently, the i, j element of the adjacency matrix is only large, and the corresponding urea molecules are thus only connected, if they are close together and if the orientations of their carbon oxygen bonds are close to parallel or anti parallel. When this connectivity matrix is analyzed using the DFS algorithm the largest cluster of solid molecules should thus be identified. This is not what is observed in practice, however. The experiments we ran using this method suggested that all the molecules in the system were part of the largest crystalline cluster at all times. In other words, the algorithm found that every molecule in the system was connected into one contiguous crystalline mass at all times despite the fact that multiple conversions between solid and liquid were clearly seen in the trajectory. It would appear that a small number of molecules are aligned when the system is in the liquid phase simply because, as shown in figure 7, the distribution of relative orientations of molecules is close to uniform. When the adjacency matrix is calculated as described above molecules thus appear connected even though they are not crystalline.

To remedy this problem an average value for $K(\phi_{ij})$ was calculated for each of the molecules in the system using:

$$\zeta_i = \frac{\sum_j \sigma(r_{ij}) K(\phi_{ij})}{\sum_j \sigma(r_{ij})} \tag{7}$$

This function, which is similar to that used in a number of other recent articles,[37–39] measures whether or not the C-O bonds on the urea molecules in the first coordination sphere around molecule i point in the same direction as the bond in molecule i. As shown in the right-most panel of figure 7 the value of this order parameter effectively distinguishes molecules within the solid from molecules within the liquid. The crystalline clusters that are present in the system can thus be identified using the strategy illustrated in figure 9. DFS clustering is
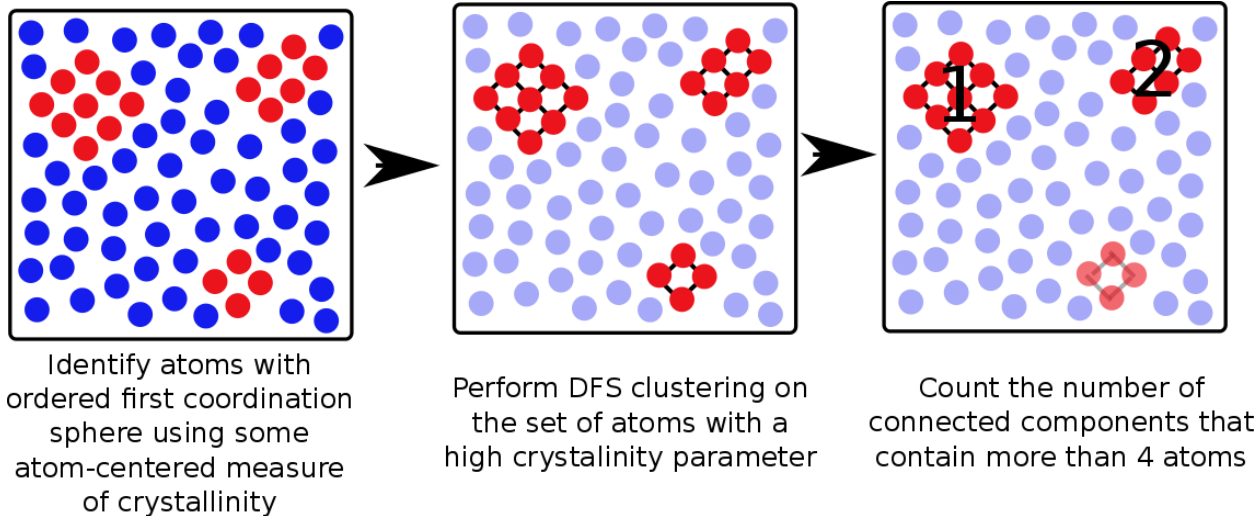
Figure 9: Figure illustrating the manner in which the CV is calculated in the later parts of section 4 and in section 5. For urea the value of the ζ_i order parameter is calculated for each of the molecules in the system while for GeTe the value of c_i is calculated for each atom. These quantities allow us to differentiate those atoms/molecules whose environment resembles that in the solid from those atoms/molecules whose environment resembles that in the liquid phase. In the leftmost panel these duly-identified, solid-like atoms/molecules are shown in red, while liquid-like atoms/molecules are shown in blue. To be totally clear, however, the red and blue circles are used to represent species that are chemically identical. Once the symmetry functions have been calculated an adjacency matrix is constructed for only those atoms with a high value for the order parameter (ζ_i or c_i). In this diagram we would thus construct an adjacency matrix for the atoms shown in red in the figure and the i, j element of this matrix would tell us whether or not atoms i and j are within a certain cutoff. When we run the DFS algorithm with this matrix as input we are thus able to identify the three distinct crystalline clusters shown in red in the second panel of this figure. As shown in the right-most panel one possible final step involves calculating how many of the connected components identified by the DFS algorithm contain more than 4 atoms.

performed considering only those atoms that have a large value for the quantity ζ_i defined in equation 7. In other words, the elements of the adjacency matrix are calculated using:

$$A_{ij} = \sigma(r_{ij})S(\zeta_i)S(\zeta_j) \qquad \text{where} \qquad S(\zeta_i) = \begin{cases} 1 & \text{if} \quad \zeta_i > \zeta_0 \\ 0 & \text{otherwise} \end{cases}$$

In this work $\zeta_0 = 0.5$ and $S(\zeta_i)$ was a Heavyside function. There is sufficient flexibility within PLUMED, however, to use a continuous switching function if derivatives are required. The

input that instructs PLUMED to do this calculation is shown below:

```
MOLECULES ...

  MOL1=1,2,1

  MOL2=9,10,9

  ...

  LABEL=m1

...  MOLECULES


smac: SMAC ...

   DATA=m1

   SWITCH={RATIONAL D_0=0.639 R_0=0.1 D_MAX=0.64}

   KERNEL1={TRIANGULAR CENTER=0 SIGMA=0.8}

   KERNEL2={TRIANGULAR CENTER=pi SIGMA=0.7}

   SWITCH_COORD={RATIONAL R_0=0.001}

... smac: SMAC


ff: MFILTER_MORE DATA=smac SWITCH={GAUSSIAN D_0=0.49 R_0=0.5 D_MAX=0.5}

c1: CONTACT_MATRIX ATOMS=ff SWITCH={RATIONAL D_0=0.639 R_0=0.01 D_MAX=0.64}

dfs: DFSCLUSTERING MATRIX=c1

cc1: CLUSTER_PROPERTIES ...

  CLUSTERS=dfs CLUSTER=1

  MORE_THAN={GAUSSIAN D_0=0.49 R_0=0.5 D_MAX=0.5}

... cc1: CLUSTER_PROPERTIES
```

This input was used to analyze a 300 ns well-tempered metadynamics trajectory, which, as discussed in the first paragraph of this section, was taken from a previously published article.[35] In this trajectory urea was observed to transition between the solid and liquid phases multiple times because of the simulation bias. The results of the analysis are shown

in figure 10. In the top panel the total number of molecules that have $\zeta_i > 0.5$, and which are thus solid, is shown. In the bottom panel meanwhile the same quantity is shown calculated for those atoms that form part of the largest connected cluster. These two plots are very similar which suggests that a single solid nucleus forms in these simulations. In other words, all the molecules of solid are connected into one single contiguous mass.
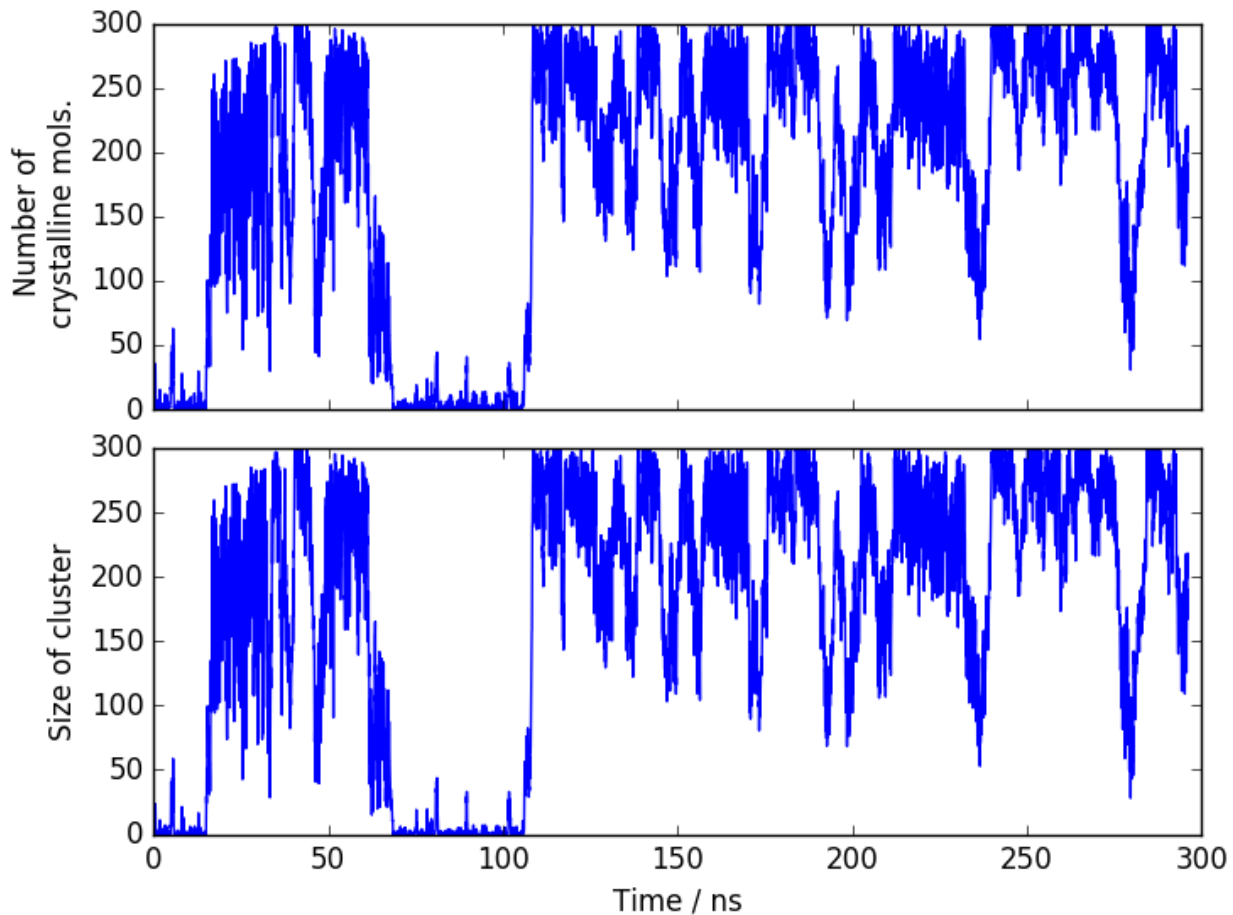
Figure 10: Analysis of a 300 ns crystallisation simulation for urea at 450 K, in which the reversible formation of the solid is observed. In the top panel the total number of molecules that have a $\zeta > 0.5$ and which are thus considered to be solid is shown. In the lower panel the number of molecules with $\zeta > 0.5$, which are are in the largest cluster found in the system is shown. These two curves are almost identical, which suggests that all the crystalline atoms in the system are connected together into a single, contiguous crystalline mass.

5   Crystal Nucleation of GeTe from the supercooled liquid phase

For this final application the formation of a polycrystalline solid within a supercooled liquid phase was examined. In particular, we analysed simulations of the phase change material GeTe whose equilibrium structure at temperatures above 623 K is that of cubic rock salt.[41] Phase change materials such as this one are of great practical relevance,[42,43] as they are currently employed in optical storage devices (such as DVDs and Blu-ray Discs) as well as in electronic nonvolatile memories (Phase Change Memories, PCM).[43] At the heart of these technologies stands the fast (on the ns timescale) and reversible phase transition between the crystalline and amorphous phases of chalcogenide glasses such as GeTe.[44] Devices that use these materials take advantage of the fact that these substances can exist in these two distinct forms when storing binary data. Furthermore, data can be read and thus recovered as the amorphous phase has an optical reflectivity and electrical resistivity that differs by several order of magnitudes from that of the crystalline phase. In PCMs, crystallization from the amorphous phase is achieved by heating the system above its the glass transition temperature. This generates a supercooled liquid which, under certain conditions, can recrystallize in the ordered phase.

The density difference between the supercooled liquid at the melting temperature of 1000 K and the crystalline phase at 0 K is of the order of 5% only.[45] If we are intent on discovering crystalline clusters of GeTe in an MD simulation working with an adjacency matrix that measures whether or not atoms are within a certain cutoff, as described in Sec. 3, is thus clearly not appropriate. Similarly, the method described in Sec. 4 is also not appropriate as the system in this case is composed of atoms, which do not have an orientation *per se*. One can, however, calculate something similar to an orientation for each central atom i by first calculating the set of vectors, $\{\mathbf{r}_{ij}\}$, that connect it to the atoms in its coordination spheres. Once these vectors are calculated the following quantities can be

extracted:

$$q_{lm}^{(i)} = \frac{\sum_{j\neq i}^{N} \sigma(|\mathbf{r}_{ij}|) Y_{lm}(\mathbf{r}_{ij})}{\sum_{j\neq i}^{N} \sigma(|\mathbf{r}_{ij}|)} \qquad \text{where} \qquad \sigma(|\mathbf{r}_{ij}|) = \begin{cases} 1 & \text{if} \quad |\mathbf{r}_{ij}| < d_0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Here the sum runs over the N atoms in the system, $Y_{lm}(\mathbf{r}_{ij})$ is one of the spherical harmonics and in this work $\sigma(|\mathbf{r}_{ij}|)$ is essentially a Heavyside function with d_0 equal to 5.3 Å although there is sufficient flexibility within PLUMED to use a continuous switching function here instead if derivatives are required. It is common within the materials simulation community to combine all the $q_{lm}^{(i)}$ values for a given l value into a single vector quantity. The norm of this vector is then the so-called Steinhardt parameter:[7,8]

$$q_l^{(i)} = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |q_{lm}^{(i)}|^2} \tag{9}$$

The top right panel of fig. 11 shows that this order parameter is not particularly effective at distinguishing those atoms in the crystalline parts of the system from those atoms in the supercooled liquid. Furthermore, we found that the method from Lechner and Dellago[8] that involves taking local averages of these complex vectors was not much more effective at distinguishing between atoms in the solid and in the liquid. A more effective choice is to use *local* Steinhardt parameters:[8,9]

$$Q_6^{(i)} = \frac{\sum_{j\neq i}^{N} \sigma(|\mathbf{r}_{ij}|) \hat{\mathbf{q}}_6^{(i)} \cdot \hat{\mathbf{q}}_6^{(j)}}{\sum_{j\neq i}^{N} \sigma(|\mathbf{r}_{ij}|)} \tag{10}$$

In this expression $\sigma(|\mathbf{r}_{ij}|)$ is a switching function on the distance between atoms i and j with parameter $d_0 = 5.3$ Å. $\hat{\mathbf{q}}_6^{(i)}$ is the versor of the $(2l+1)$-dimensional complex vector $\mathbf{q}_l^{(i)}$ that is constructed by combining all the elements of $q_{lm}^{(i)}$ with $l = 6$. This local order parameter, $Q_6^{(i)}$, thus measures whether or not the atoms in the first and second coordination spheres around atom i have their coordination spheres ordered in a similar fashion. Obviously, for

atoms in the bulk of an ordered solid phase this quantity is large. For atoms in a disordered amorphous or liquid phase, meanwhile, this quantity should be small. It is thus unsurprising to find that, as shown in the lower left panel of figure 11, this approach is better able to distinguish between atoms in the disordered and ordered phases.

The local order parameter $Q_6^{(i)}$ has been used widely in the literature. It is popular because it often gives a far more effective measure of local order than the Steinhardt order parameter $q_l^{(i)}$ (see Eq. 9), which is particularly important when performing enhanced sampling simulations of crystallization.[46] In this work the switching functions in equations 8 and 10 are set large enough so that all the atoms in the first and second coordination spheres of atom i are taken into account when calculating $q_6^{(i)}$ and $Q_6^{(i)}$ because, as shown figure 11, the resulting quantity is better able to distinguish crystalline from amorphous than the function in which only the first coordination spheres are considered.

In figure 12a the atoms are colored according to the value of the $Q_6^{(i)}$ parameter. This figure highlights an additional complexity associated with examining nucleation in this system. In certain temperature regimes (500-600 K) multiple nucleation centers appear on the very short timescale of 10^{2-3} ps over which crystallization is observed and the product that ultimately emerges is thus polycrystalline.[47,48] This product polycrystal contains a number of randomly oriented grains, which can be distinguished by eye in figure 12a. In this figure the regions where the atoms are colored blue and where they are clearly ordered are separated by narrow red regions where the structure is disordered. Ideally the clustering algorithm should find that each of these misaligned domains is disconnected from the others so that each domain is classified as a separate cluster.

The approach outlined in figure 9 shows how the local Steinhardt parameters can be used when finding the crystalline domains. Remember these parameters allow one to distinguish atoms in the ordered parts of the system from those in more disordered parts. The elements
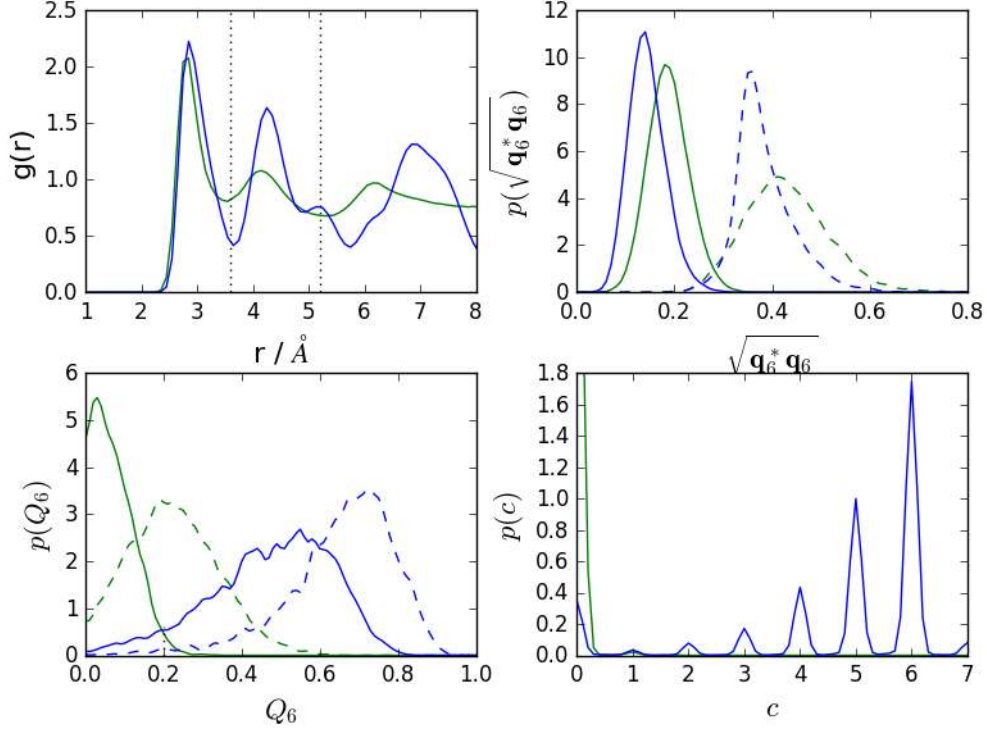
Figure 11: Histograms of symmetry functions for pure GeTe crystal (blue lines) and pure GeTe liquid (green lines). These probabilities were calculated from 2 ns long trajectories of 4096 atoms of solid and liquid, which were run at a temperature of 675 K. Further details can be found in Ref. 47. The top left panel shows the radial distribution functions together with dashed lines which indicate where the first and second coordination spheres are considered to end in this work. The second figure shows the histogram of values we observed for the modulus of the vector $\mathbf{q_6}$ in the solid and liquid systems. The dashed lines in this figure indicate the histogram obtained if only the atoms in the first coordination sphere are considered when calculating this quantity. The solid lines, meanwhile, show what is obtained if both the first and second coordination spheres are considered. Neither of these symmetry functions is particularly good at distinguishing crystalline atoms from those in the supercooled liquid phase. The lower left panel shows the histograms obtained for Q_6. Once again the dashed lines show the distributions that are obtained if the parameter for the switching function in equation 10 is set equal to 3.6 Å(first coordination sphere only), while the solid lines show the distribution that is obtained if this parameter is set equal to 5.2 Å (first and second coordination spheres). For both pairs of histograms the switching function in equation 8 is set equal to 5.2 Å. Once again there is overlap between the distributions of values seen for the solid and liquid and this parameter thus struggles to distinguish the solid from the liquid. The bottom right panel shows the histograms for the quantity, c, which is defined in equation 1. As is clear from the figure this symmetry function is the most effective in distinguishing crystalline atoms from liquid atoms.
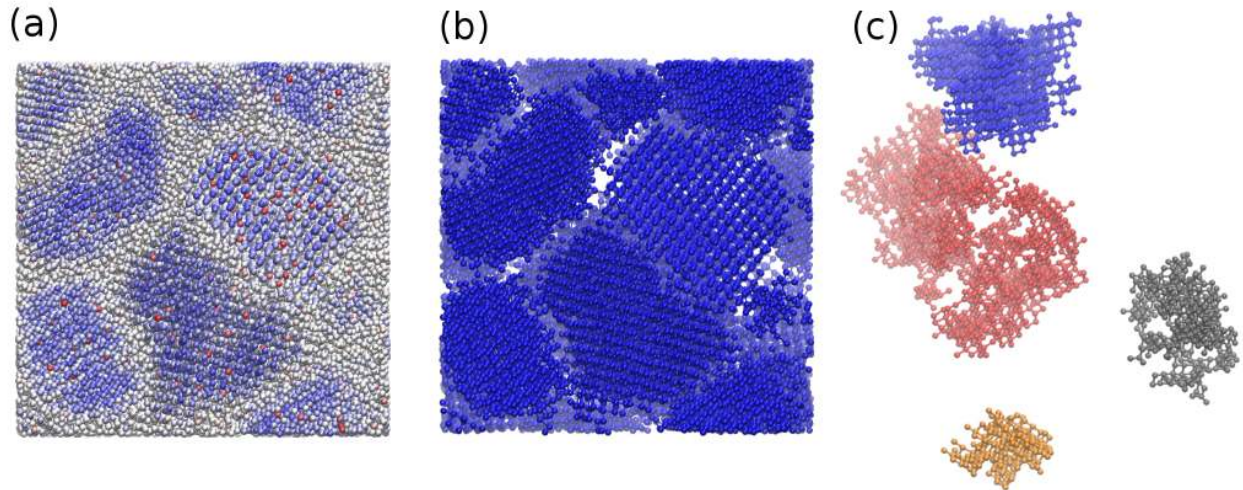
Figure 12: The clusters found in a polycrystalline, 32768-atom configuration of GeTe using various different clustering protocols. Panel (a) shows a snapshot of all the atoms in the system. Those atoms with a high value for Q_6 are colored blue whilst those atoms with a low value are colored white or red. In this figure one can clearly see that the sample is composed of ordered but misaligned domains. Panel (b) shows that the largest cluster contains the majority of the atoms in the system if only those atoms with a low value of Q_6 are excluded from the clustering. A more effective strategy is illustrated in panel (c). The clusters shown in this figure were found by neglecting any atom that did not have a high value for Q_6 and which was not connected to at least six other atoms that also had high values for Q_6 when calculating the adjacency matrix. In this figure the atoms shown in blue are part of the largest domain found, the atoms in red compose the second largest domain, the atoms shown in black form the third largest and the atoms shown in yellow form the fourth largest domain.

of the adjacency matrix are thus calculated using the following expression:

$$A_{ij} = \sigma(|\mathbf{r}_{ij}|) S(Q_6^{(i)}) S(Q_6^{(j)}) \qquad \text{where} \qquad S(Q_6^{(j)}) = \begin{cases} 1 & \text{if} \quad Q_6^{(j)} > q_0 \\ 0 & \text{otherwise} \end{cases}$$

A Heavyside function was used in this work for $S(Q_6^{(j)})$ with $q_0 = 0.2$, while the parameter of the switching function on the distance, $\sigma(|\mathbf{r}_{ij}|)$, was set equal to $d_0 = 3.6$ Å. It is worth noting there is, once again, sufficient flexibility within PLUMED to replace these two discrete switching functions with continuous functions if derivatives are required.

By filtering the atoms using a switching function, $S(Q_6^{(i)})$ on the local q6 parameter we effectively ensure that the atoms in the disordered parts of the system, which have a low value for Q_6, are always disconnected. The overall effect is thus that the DFS clustering is only performed on the set of atoms which have a high value for $Q_6^{(i)}$.

Analyzing the configuration shown in figure 12a using this particular approach gives the result shown in figure 12b. Unfortunately, all of the misaligned crystalline domains are connected into one single contiguous mass. We are thus unable to label the different grains. Looking closely at the cluster found in figure 12c, however, one sees that there is empty space around the various crystalline grains in the polycrystalline sample. This makes a lot of sense as figure 12a shows that each of the regions where the atoms are coloured blue, indicating a high-$Q_6^{(i)}$ value, is surrounded by regions in which the atoms are coloured white and red indicating low values for $Q_6^{(i)}$. The problem is that within these regions, where the $Q_6^{(i)}$ values are for the most part small, there are a few atoms that have anomalously high values. These atoms connect the grains and thus explain why the clustering algorithm finds that all the grains connected into one single domain. To counter balance this effect we thus calculated the following coordination number for each of the atoms:

$$c_i = \sum_{i \neq j} S(Q_6^{(i)}) S(Q_6^{(j)}) \sigma(|r_{ij}|)$$

The parameter of the switching function, $\sigma(|r_{ij}|)$, is set equal to $d_0 = 3.6$ Å in this case so this function measures the number of ordered atoms in the first coordination sphere of atom i. Figure 11 shows the overlap in the histograms of values for this quantity that were found for a trajectory of crystalline structures and a trajectory of liquid structures. The overlap between these two histograms is negligible so this quantity is better than Q_6 at distinguishing atoms in the crystal from atoms in the melt. To incorporate this symmetry function in the cluster analysis the elements of the adjacency matrix were thus calculated using:

$$A_{ij} = f(c_i)f(c_j)\sigma(|r_{ij}|) \qquad \text{where} \qquad f(c_i) = \begin{cases} 1 & \text{if} \quad c_i > c_0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

In this work the parameter in the Heavyside function c_0 was set equal to 6 and once again there is flexibility within PLUMED to make this discontinuous function continuous if derivatives are required. The PLUMED input that allows one to cluster the configuration using this matrix is shown below:

```
q6: Q6 SPECIES=1-32768 SWITCH={GAUSSIAN D_0=5.29 R_0=0.01 D_MAX=5.3} LOWMEM
lq6: LOCAL_Q6 SPECIES=q6 SWITCH={GAUSSIAN D_0=5.29 R_0=0.01 D_MAX=5.3} LOWMEM
flq6: MFILTER_MORE DATA=lq6 SWITCH={GAUSSIAN D_0=0.19 R_0=0.01 D_MAX=0.2}
cc: COORDINATIONNUMBER SPECIES=flq6 SWITCH={GAUSSIAN D_0=3.59 R_0=0.01 D_MAX=3.6}
fcc: MFILTER_MORE DATA=cc SWITCH={GAUSSIAN D_0=5.99 R_0=0.01 D_MAX=6.0}
mat: CONTACT_MATRIX ATOMS=fcc SWITCH={GAUSSIAN D_0=3.59 R_0=0.01 D_MAX=3.6}
dfs: DFSCLUSTERING MATRIX=mat
OUTPUT_CLUSTER CLUSTERS=dfs CLUSTER=1 FILE=cluster1.xyz
```

As in the previous example using this protocol effectively ensures that some of the atoms in the system will always be disconnected from the clusters. In this case, however, rather than just discarding those atoms with a low $Q_6^{(i)}$ parameter we also discard atoms that are bonded to fewer than six atoms that also have a large value for $Q_6^{(i)}$. This effectively removes

25

the atoms in the disordered parts of the system from consideration as, although these atoms have high values for $Q_6^{(i)}$, they are connected to atoms that have low values for $Q_6^{(i)}$. As shown in figure 12d with this setup we find that the various misaligned crystalline domains are disconnected.

Having identified a suitable collective variable we used it to analyze an 800 ps nucleation trajectory that was taken from Ref. 48. This simulation involved 16384 GeTe formula units and was run at 600 K, which corresponds to a supercooling of 400 K. The results from this analysis are shown in figure 13. The top panel shows the sum of $f(c_i)$ calculated for all the atoms in the system as a function of time. In other words, this top panel shows the total number of atoms in crystalline domains. The middle panel of this figure shows how the number of domains containing more than 27 atoms changes as a function of simulation time. This quantity was calculated using:

$$w = \sum_{i=1}^{M} \Omega \left(\sum_{j=1}^{N_i} f(c_j) \right) \qquad \text{where} \qquad \Omega(x) = \begin{cases} 1 & \text{if} \quad x > 27 \\ 0 & \text{otherwise} \end{cases}$$

In this expression the first sum runs over the clusters found using the DFS algorithm. The second sum meanwhile accumulates the sum of the $f(c_i)$ values (see equation 11) for those atoms in each of the clusters found. It is important to note that, because of the difficulties discussed in section 3, this quantity cannot be made continuous by replacing the discontinuous function Ω with a continuous function.

The lower panel in figure 13 shows the number of atoms in each of the six largest domains found in the system as a function of simulation time. This quantity is calculated using the following function:

$$n = \sum_{j=1}^{N} f(c_j)$$

where the sum runs over the atoms in in the cluster of interest and $f(c_j)$ is defined as in equation 11. To calculate the quantities shown in the middle panels of figure 13 the following

PLUMED input must be appended after the input that was shown previously.

```
nclust: CLUSTER_DISTRIBUTION ...

  CLUSTERS=dfs

  TRANSFORM={GAUSSIAN D_0=5.99 R_0=0.01 D_MAX=6.0}

  MORE_THAN={GAUSSIAN D_0=26.99 R_0=0.01 D_MAX=27}

... nclust: CLUSTER_DISTRIBUTION

PRINT ARG=nclust.* FILE=colvar
```
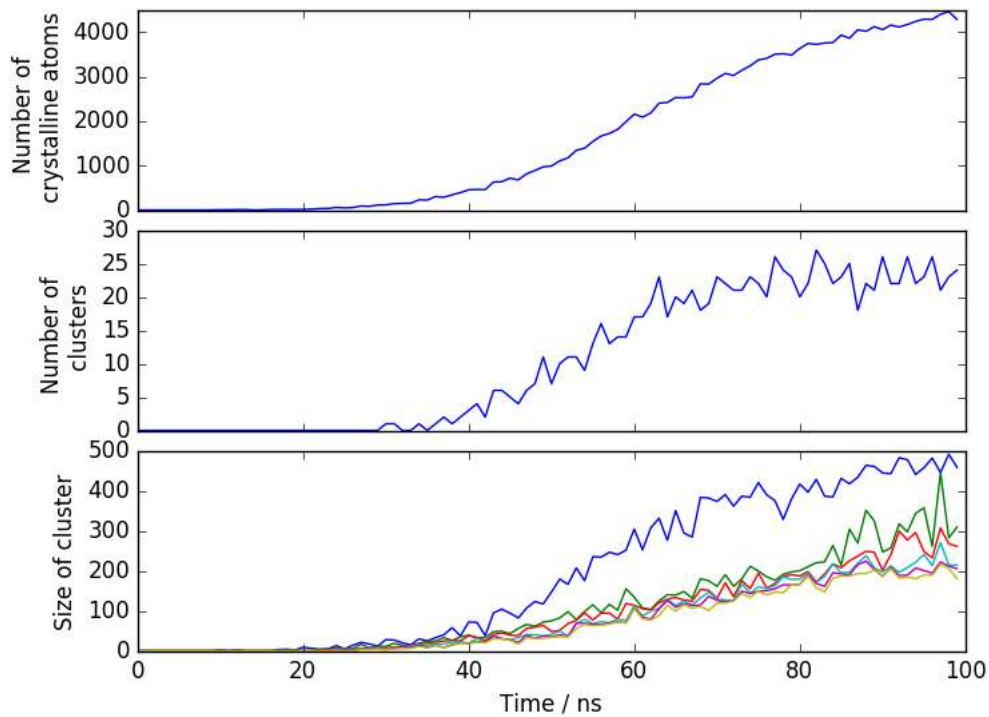

Figure 13: Analysis of a 800 ps crystallisation simulation for GeTe at 600 K, in which the formation of a polycrystal is observed. In the top panel we show how the number of atoms that are identified as crystalline changes as a function of simulation time. This quantity was calculated by counting the number of atoms with a $f(c_i)$ value greater than or equal to 6. The middle panel shows the number of crystalline clusters containing 27 solid atoms or more as a function of time. The lowest panel meanwhile shows how the sizes of the 6 largest crystalline domains found in the sample change as a function of time.

6 Conclusions

In the preceding sections we have discussed the framework that we have developed for examining nucleation and growth phenomena by means of molecular dynamics simulation. This framework is based on a set of collective variables that are inspired by the definition given to the word phase in thermodynamics. A phase of a system is defined as a region of space in which the physical properties of the system are uniform. Hence, when one crosses a phase boundary the physical properties of the system change discontinuously. All the CVs that we have discussed within this article thus work by performing three operations:

1. Local atom-centered, symmetry functions are computed to distinguish between the various different environments in the system.

2. An adjacency matrix is computed, which is then analysed using a clustering algorithm. This procedure returns the set of connected components in the graph.

3. The sum of the symmetry functions in each of the connected components is determined and some linear/non-linear combination of these quantities is used as a final collective variable.

The object oriented structure within PLUMED has allowed us to write an extremely flexible implementation of this method. Firstly, we can use all the various atom-centered symmetry functions that are available in PLUMED in step 1. In step 2 we can construct the adjacency matrix using all the atoms in the system as was explained in section 3 or we can choose to only calculate the adjacency matrix for those atoms that have their symmetry function within a particular range as discussed in sections 4 and 5. Furthermore, the method used to calculate the elements of this adjacency matrix can be changed. We are thus not confined to simply stating that particles i and j are connected if they are within a certain cutoff distance of each other. Lastly, once we have determined a set of connected components using the DFS algorithm, we can calculate a range of different non-linear combinations of the symmetry functions of which they are composed in order to get a final CV.

This new functionality in PLUMED makes it straightforward to create new CVs as complex combinations of variables directly from the input file. In addition, other scientists using PLUMED are not confined to using the functionality that is currently available within it. We are aware that the specific details of any problem under study are important so we have tried to make it as straight forward as possible to implement new methods for calculating symmetry functions, adjacency matrices and even for finding the connected components from the adjacency matrix in PLUMED. It is thus possible to use new methods implemented in this way in any of the work flows described in the earlier sections of this paper.

7   Acknowledgments

References

(1) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961 − 1972.

(2) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Commun.* **2014**, *185*, 604 − 613.

(3) Behler, J.; Parrinello, M. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(4) Angioletti-Uberti, S.; Ceriotti, M.; Lee, P. D.; Finnis, M. W. *Phys. Rev. B* **2010**, *81*, 125416.

(5) Cheng, B.; Tribello, G. A.; Ceriotti, M. *Phys. Rev. B* **2015**, *92*, 180102.

(6) Giberti, F.; Salvalaglio, M.; Mazzotti, M.; Parrinello, M. *Chem. Eng. Sci.* **2015**, *121*, 51–59.

(7) Steinhardt, P. J.; Nelson, D. R.; Ronchetti, M. *Phys. Rev. B* **1983**, *28*, 784–805.

(8) Lechner, W.; Dellago, C. *J. Chem. Phys.* **2008**, *129*, 114707.

(9) Auer, S.; Frenkel, D. *Ann. Rev. Phys. Chem.* **2004**, *55*, 333–361.

(10) Even, S. *Graph Algorithms*, 2nd ed.; Cambridge University Press, 2011; Cambridge Books Online.

(11) Von Luxburg, U. *Stat. and Comput.* **2007**, *17*, 395–416.

(12) MacQueen, J. Proc. Berkeley Symp. Math. Stat. Prob. 1967; p 14.

(13) McQuitty, L. L. *Ed. Psyc. Meas.* **1957**, *17*, 207–229.

(14) Jain, A. K.; Murty, M. N.; Flynn, P. J. *ACM Comp. Surv.* **1999**, *31*, 264–323.

(15) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.

(16) Grubmüller, H.; Heymann, B.; Tavan, P. *Science* **1996**, *271*, 997–999.

(17) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.

(18) Barducci, A.; Bonomi, M.; Parrinello, M. *Wires Comput. Mol. Sci.* **2011**, *1*, 826–843.

(19) Giberti, F.; Salvalaglio, M.; Parrinello, M. *IUCrJ* **2015**, *2*, 256–266.

(20) Chkonia, G.; Wölk, J.; Strey, R.; Wedekind, J.; Reguera, D. *J. Chem. Phys.* **2009**, *130*.

(21) ten Wolde, P. R.; Frenkel, D. *J. Chem. Phys.* **1998**, *109*, 9901–9918.

(22) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(23) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(24) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(25) Lindahl, E.; Hess, B.; Van Der Spoel, D. *Mol. Model* **2001**, *7*, 306–317.

(26) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(27) Pronk, S.; Pll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845.

(28) Bonomi, M.; Barducci, A.; Parrinello, M. *J. Comput. Chem.* **2009**, *30*, 1615–1621.

(29) Wedekind, J.; Reguera, D.; Strey, R. *J. Chem. Phys.* **2006**, *125*, 214505.

(30) Reguera, D.; Bowles, R. K.; Djikaev, Y.; Reiss, H. *J. Chem. Phys.* **2003**, *118*, 340–353.

(31) Grossier, R.; Veesler, S. *Cryst. Growth Des.* **2009**, *9*, 1917–1922.

(32) Salvalaglio, M.; Perego, C.; Giberti, F.; Mazzotti, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, E6–E14.

(33) Theodorou, D. N.; Suter, U. W. *Macromolecules* **1985**, *18*, 1206–1214.

(34) Moroni, D.; ten Wolde, P. R.; Bolhuis, P. G. *Phys. Rev. Lett.* **2005**, *94*, 235703.

(35) Giberti, F.; Salvalaglio, M.; Mazzotti, M.; Parrinello, M. *Chem. Eng. Sci.* **2015**, *121*, 51 – 59, 2013 Danckwerts Special Issue on Molecular Modelling in Chemical Engineering.

(36) Giberti, F.; Tribello, G. A.; Parrinello, M. *J. Chem. Theory Comput.* **2013**, *9*, 2526–2530.

(37) Salvalaglio, M.; Vetter, T.; Giberti, F.; Mazzotti, M.; Parrinello, M. *J. Am. Chem. Soc.* **2012**, *134*, 17221–17233.

(38) Salvalaglio, M.; Vetter, T.; Mazzotti, M.; Parrinello, M. *Angew. Chem., Int. Ed.* **2013**, *52*, 13369–13372.

(39) Salvalaglio, M.; Mazzotti, M.; Parrinello, M. *Faraday Discuss.* **2015**, *179*, 291–307.

(40) Salvalaglio, M.; Giberti, F.; Parrinello, M. *Acta Crystallogr., Sect. C: Struct. Chem.* **2014**, *70*, 132–136.

(41) Raoux, S.; Muñoz, B.; Cheng, H.-Y.; Jordan-Sweet, J. L. *Appl. Phys. Lett.* **2009**, *95*, 143118.

(42) Raoux, S.; Wełnic, W.; Ielmini, D. *Chem. Rev.* **2010**, *110*, 240–267.

(43) Lencer, D.; Salinga, M.; Wuttig, M. *Adv. Mater.* **2011**, *23*, 2030–2058.

(44) Wuttig, M.; Yamada, N. *Nat. Mat.* **2007**, *6*, 824–832.

(45) Sosso, G. C.; Behler, J.; Bernasconi, M. *Phys. Stat. Sol. B* **2012**, *249*, 1880–1885.

(46) Reinhardt, A.; Doye, J. P. K.; Noya, E. G.; Vega, C. *J. Chem. Phys.* **2012**, *137*, 194504.

(47) Sosso, G. C.; Miceli, G.; Caravati, S.; Giberti, F.; Behler, J.; Bernasconi, M. *J. Phys. Chem. Letters* **2013**, *4*, 4241–4246.

(48) Sosso, G. C.; Salvalaglio, M.; Behler, J.; Bernasconi, M.; Parrinello, M. *J. Phys. Chem. C* **2015**, *2015*, 6428–6434.