



Structural Analysis by Statistical Learning Method for Atomic Scale Nanoparticles

M2 Internship Report
FunPhys Master

by

Carlos Rafael SALAZAR LETONA

Lille, July 2023

Supervisor: Dr. Julien LAM
Unité Matériaux Et Transformations
Centre d'Élaboration de Matériaux et d'Etudes Structurales



Contents

1	Introduction	4
2	Theory	6
2.1	Steinhardt Parameters	6
2.1.1	Original Version	6
2.1.2	Average Steinhardt Parameters	7
2.1.3	Voronoi Weighted Steinhardt Parameters	7
2.2	Clustering	8
2.2.1	Gaussian Mixture Model	9
2.2.2	Maximum Likelihood Classifier	10
2.2.3	Expectation-Maximization Algorithm	11
3	Methodology	12
3.1	MD Simulation Protocol	12
3.2	Database	12
3.3	Classification	14
3.4	Seeding Technique	14
3.4.1	Seed Creation	17
3.4.2	Cluster Analysis	17
3.4.3	Relaxation	17
3.4.4	Growth/Melting Simulation	18
4	Results and Discussion	20
4.1	Gaussian Mixture Model on Database	20
4.2	Melting Analysis	23
4.3	Relaxation	27
4.4	Growth and Critical Temperature	27
4.4.1	BCT Cluster Growth	29
4.4.2	WRZ Cluster Growth	32
5	Conclusion and Outlook	34
6	References	37
	Appendix	45

Abstract

Observing nucleation in simulations remains challenging due to the long time needed to cross the free energy barrier required to form a nucleus. Furthermore, the structural analysis necessary to understand the process involved in nucleation can be complicated to perform in materials with multiple available structures. In this work, we present a new classification methodology for structural analysis and apply it to the study of melting and crystallization in Zinc Oxide (ZnO), using also the seeding technique to estimate the critical temperature of body-centered tetragonal (BCT) and wurtzite (WRZ) crystalline clusters in nanoparticles. In our classification methodology, we make use of the average Voronoi-weighted Steinhardt parameters as descriptors and the Gaussian Mixture Model (GMM) with the Maximum Likelihood Classifier (MLC) as a machine learning classification method. We show that our classification method makes physically meaningful structural predictions by performing multiple tests on ZnO structures. In our study of the melting of BCT and WRZ bulk crystals, the results support the hypothesis of a two-step transition between the WRZ and liquid phases. For the study of nucleation, we perform the seeding technique by inserting a crystal seed into a liquid droplet and analyzing changes in size at different temperatures. We found that the critical temperature for BCT and WRZ clusters is similar for approximate sizes of 180 and 360 atoms. For smaller clusters, around 70 atoms, we found a significantly higher critical temperature for the BCT cluster. By performing structural analysis in the crystalline clusters we observe a transition of the BCT cluster to a WRZ cluster. On the other hand, the WRZ cluster is commonly surrounded by BCT and HBN atoms. Our results support the hypothesis of the two-step transition between the WRZ and the liquid phases in the melting and crystallization processes. Furthermore, we shed light on the crystallization of ZnO nanoparticles and put into question the stability of BCT crystalline clusters.

1 Introduction

Crystallization is a process that is present in both natural and synthetic environments. Studying and being able to control this process has allowed many technological advances. Multiple works have been dedicated to studying the emergence of the nucleation core [1–5] and the role that the nucleation core has on the final crystal structure [6–12]. However, nucleation is a stochastic process associated with crossing a free energy barrier. Because of this, its observation requires long simulations and nucleation is considered to be a rare event [13]. Several methods have been devised for the sampling of rare events, such as biased simulations [14, 15], trajectory sampling [16], and the seeding technique [17, 18].

From the fundamental viewpoint, Classical Nucleation Theory (CNT) [19–21] has been utilized to describe nucleation processes. In CNT, the critical cluster size required for crystallization is achieved by the fluctuations in number density that are inherent in a system at finite temperatures. CNT also stipulates that there is a single energy barrier that needs to be overcome in order to form a nucleus, which is the energy required to form a surface of the critical cluster size. This approach was successful in the qualitative explanation of numerous observations and was hence the main nucleation theory for over 100 years. However, over the past 20 years, advances in imaging and spectroscopy tools allowed for observations that reveal hierarchical nucleation pathways. These observations challenge the fundaments of CNT and in recent years it has been shown that for many nucleation processes the Ostwald step mechanism [22–24] is the preferred nucleation pathway. In this mechanism, the structures that are closest in free energy to the initial phase are the first ones to form, while the more thermodynamically stable phases form after [25]. This mechanism differs from CNT as it requires the presence of at least two energy barriers, one to go from the liquid state to the kinetically stable crystal, and a second barrier to go from the kinetically stable crystal to the thermodynamically stable crystal.

Due to the complexity, size, and length of nucleation simulations, it is hard to understand the mechanisms involved. Unknown structures may be encountered at the local or global scale. It has also been found that well-known materials can present a hidden structural order in the middle range of a transition [26–28]. Several methods have been traditionally used to analyze the structure of materials, such as Radial Distribution Functions (RDFs) [29–31], Common Neighbor Analysis (CNA), [32] and Centrosymmetry Parameter (CSP) [33].

In this context, the study of nucleation presents two challenging tasks. On the one hand, because it is a rare event, one requires original simulation strategies to observe nucleation in molecular dynamics. On the other hand, because of the structural complexity that is involved, one requires innovative analysis tools to determine the structure of the emerging crystal.

In this project, we focus on the Zinc-Oxide (ZnO) system. This system has raised interest in the pharmaceutical, cosmetic, food, rubber, commodity chemical, paint-

ing, ceramic, and glass industries [34]. It has been found that the nucleation process in bulk ZnO systems is also a two-step process [35], illustrated in Fig 1. However, it was not possible to confirm the Ostwald step mechanism in ZnO due to the long simulations needed to compute free energies. Here, we expand this work by studying the nucleation of ZnO using the seeding technique first introduced by Bai and Li [17, 18]. In this way, nucleation can be studied with shorter simulations. However, its application is dependent on the classification of liquid and crystalline atoms. The classification task in ZnO systems can be problematic since it presents many crystalline phases [36]. To perform structural analysis we present a new methodology using average Voronoi-weighted Steinhardt parameters as descriptors and the Gaussian Mixture Model (GMM) as a classification method.

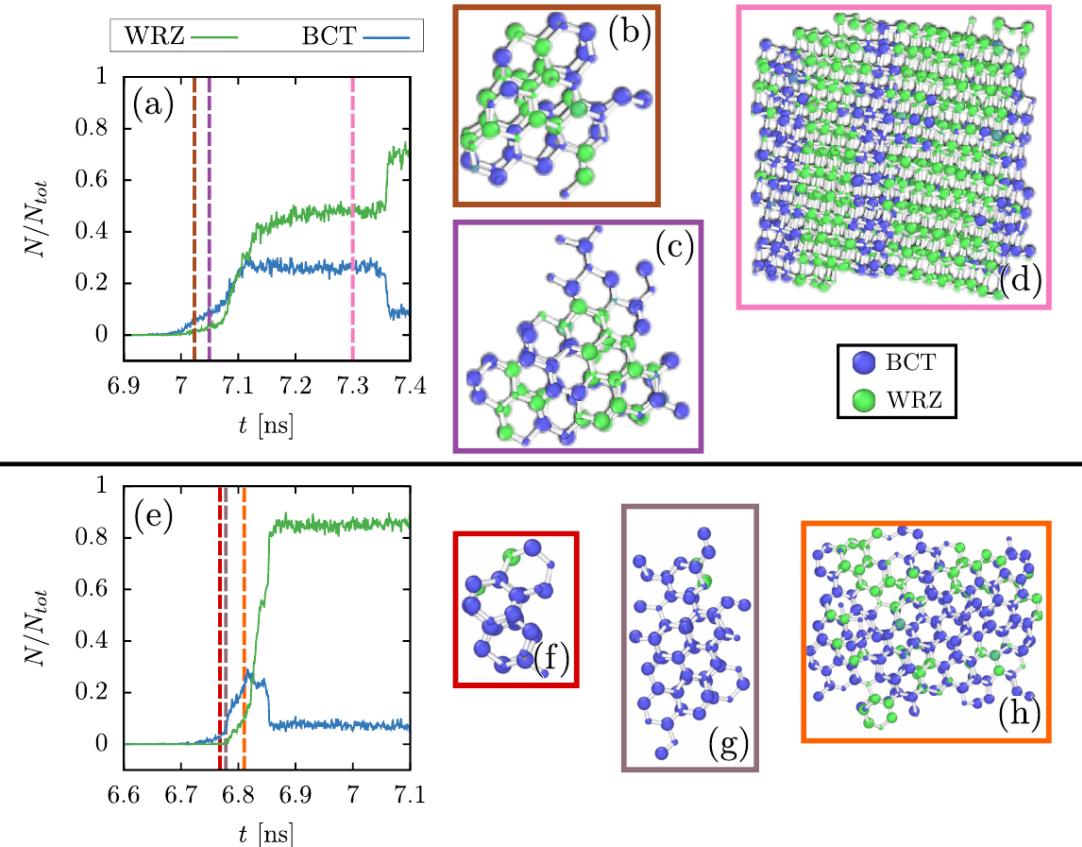


Figure 1: Number of atoms detected in WRZ or BCT environments along the cooling path in two different scenarios. The top figure represents a scenario where nucleation starts with a BCT core and few WRZ atoms. The bottom figure represents a scenario where the nucleation starts only with a BCT core. Colored dash lines point out where the structures shown on the right are picked from along the simulations. Figure obtained from Ref [35].

2 Theory

Recently, several methods for local structure detection have been proposed [37–43]. A common feature is that the identification of atomic structure is done in two main steps. First, the local structure environment is numerically characterized using a high-dimensional descriptor. Then, a comparison is made between the descriptors to classify the atomic structures. This is done using supervised or unsupervised Machine Learning (ML) methods. Dimensionality reduction has also been used as an extra step before classification [42–47].

2.1 Steinhardt Parameters

2.1.1 Original Version

In 1983, Steinhardt et. al. proposed the bond-orientational order parameters [48], also known as Steinhardt Parameters. These local descriptors are based on spherical harmonics to quantify the rotational symmetry of an atomic environment. They are used extensively in the identification of different crystalline phases [49–55], identification of solid and liquid phases in a system [56–58], and the study of interfaces [59].

The first step of computing the Steinhardt parameters is to compute the complex vector $q_{lm}(i)$ for a particle i defined as

$$q_{lm}(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} Y_{lm}(\mathbf{r}_{ij}), \quad (1)$$

where the functions $Y_{lm}(\mathbf{r}_{ij})$ are the spherical harmonics evaluated for the vector \mathbf{r}_{ij} going from particle i to its neighboring particle j . The index l is a free integer parameter that will define the degree of the Steinhardt parameter, while the index m is an integer in the range $[-l, +l]$. The spherical harmonics functions are averaged over the set $N_b(i)$ of nearest neighbors of particle i . Therefore, the value of the Steinhardt parameters is dependent on the neighbor definition.

The Steinhardt parameters are then computed by averaging the previous complex vectors over the index m as

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

Depending on the choice of the degree l the parameters will be sensitive to different crystal symmetries. Commonly, the parameter q_6 is used for the distinction between liquid and solid, while the combination of q_4 and q_6 is used for the identification of crystal structures [37, 60].

2.1.2 Average Steinhardt Parameters

Lechner and Dellago found that the previous definition of the Steinhardt parameters could be improved by taking the average of the atom and its first-shell neighbors [37]. They showed that it was easier to distinguish different crystal structures in a scatter plot of the average Steinhardt parameters than in a scatter plot of the simple Steinhardt parameters. The improvement observed by Lechner and Dellago is shown in Fig 2 for the Lennard-Jones system in three different crystalline structures and in the liquid phase.

The averaged Steinhardt parameters are computed using the average of the complex vector as

$$\bar{q}_{lm}(i) = \frac{1}{\tilde{N}_b(i)} \sum_{k=0}^{\tilde{N}_b(i)} q_{lm}(k), \quad (3)$$

where the sum from $k = 0$ to $\tilde{N}_b(i)$ iterates over particle i and all of its neighbors. With these newly defined vectors, the Steinhardt parameter is computed as before using Eq 2.

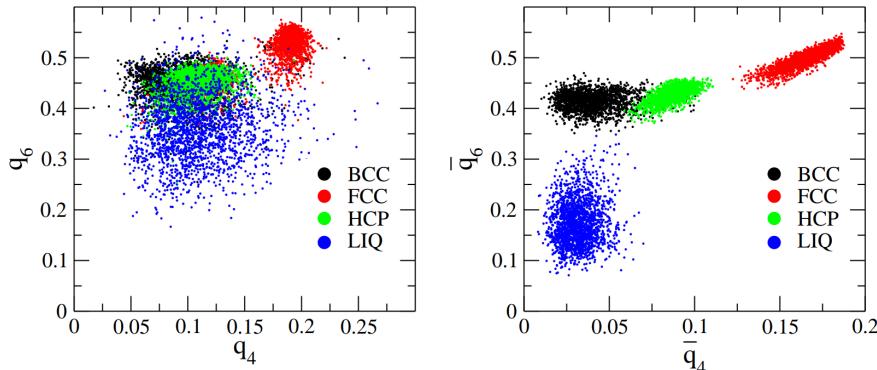


Figure 2: Comparison between simple Steinhardt parameters (left) and averaged Steinhardt parameters (right). Figure obtained from Ref [37].

2.1.3 Voronoi Weighted Steinhardt Parameters

The Steinhardt parameters are computed using the neighbor network for a given atom. This makes the neighbor definition a crucial step to obtaining meaningful descriptions of local structures. It has been found that non-crystalline systems can locally have the same q_6 values as HCP or FCC environments [61]. Mickel et al. [60] have demonstrated that the use of q_6 as a structure metric in disordered systems shows many discrepancies. This is in accordance with difficulties found in the application of q_6 in ordered and disordered systems [62–64].

Mickel et al. [60] also demonstrate that the Steinhardt parameters show a significant dependence on the neighbor network. They argue that changes in the neighbor definition have an effect not only on the absolute values of the Steinhardt parameters but also on the functional trends. To tackle this issue Mickel et al. proposed the Voronoi weighted Steinhardt parameters [60]. In this method, the neighbors for a given particle are selected using Voronoi tessellation [65]. The certainty of different bonds can then be quantified by the surface area of the Voronoi cell $A(f)$ for a given facet f separating two neighboring atoms. Using this, the complex vector presented in Eq 1 is modified as

$$q_{lm}(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} \frac{A_{ij}}{A} Y_{lm}(\mathbf{r}_{ij}), \quad (4)$$

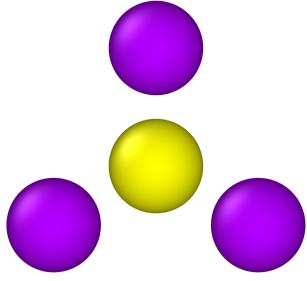
where A_{ij} is the area of the facet connecting atoms i and j , and A is the total area of the Voronoi cell. In this way, the contribution of each neighboring atom to the Steinhardt parameter will be weighted by a relative area factor.

Figs 3 and 4 show examples of Steinhardt parameter values for different symmetric systems computed using different techniques. In Fig 3a the values of the non-averaged Steinhardt parameters are the highest for q_3 and q_6 , implying that the system has a 3-fold and 6-fold rotational symmetry. However, this is not the case for the averaged parameters, since their value depends also on the values for the neighbors. From the point of view of the neighbors, the symmetry of the system is not the same. Similarly, Fig 3b shows a system with 4-fold and 8-fold symmetry. Fig 4 exemplifies one of the shortcomings of using simple Steinhardt parameters with a cutoff radius. For both systems in Figs 4a and 4b the Steinhardt parameters are the same when computed with only a cutoff, even though both systems are visually different. This type of problem is addressed by the averaged and the Voronoi-weighted Steinhardt parameters. Furthermore, the power of Voronoi tessellation to select the neighbors of the central atom is exemplified by the agreement of non-averaged Voronoi weighted Steinhardt parameters between Fig 3a and Fig 4a, and Fig 3b and Fig 4b.

2.2 Clustering

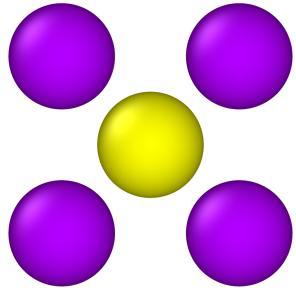
Clustering is a general technique in data science through which objects are assigned to a group of similar objects (clusters). Objects are usually described as vectors of features. They can be numerical or categorical. The assignment of an object to a cluster can be hard or fuzzy. In a hard clustering approach, the object belongs to only one cluster. On the other hand, fuzzy clustering means that an object can belong to multiple clusters, having different probabilities for each one [66]. In the context of clustering, the definition of a distance measure is also important, as it quantifies the similarity between two objects.

The different clusters in a model can, for example, represent different classes of



(a)

	Cutoff	Cutoff Average	Voronoi	Voronoi Average
q_3	0.7906	0.0000	0.7529	0.3506
q_4	0.3750	0.3750	0.4048	0.4805
q_6	0.7408	0.3125	0.6868	0.2846
q_8	0.5331	0.2734	0.5337	0.4229



(b)

	Cutoff	Cutoff Average	Voronoi	Voronoi Average
q_3	0.0000	0.0000	0.0000	0.0000
q_4	0.8292	0.4031	0.8132	0.4591
q_6	0.5863	0.3279	0.5373	0.1359
q_8	0.7979	0.6887	0.7786	0.6572

Figure 3: Values of q_3 , q_4 , q_6 , and q_8 parameters for single-shell systems computed on the central atom (yellow) using different techniques.

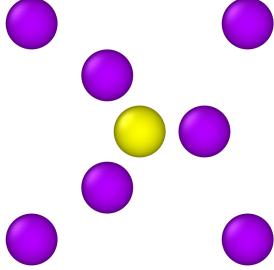
molecules or configurations of a given system [67]. Multiple unsupervised clustering techniques have been applied to the study of atomic structure such as Random Forest Classifier [45], Gaussian Mixture Model [42, 43, 68], and Kernel methods [69].

2.2.1 Gaussian Mixture Model

The Gaussian Mixture Model (GMM) is an unsupervised clustering technique that assumes that the data is composed of a set of Gaussian distributions [70]. The classes are defined by their parameters \mathbf{m}_i and \mathbf{C}_i , representing the mean vector and covariance matrix of the corresponding Gaussian distribution ω_i . The assumption of the GMM is expressed mathematically as a linear mixture of K Gaussian distributions of the form

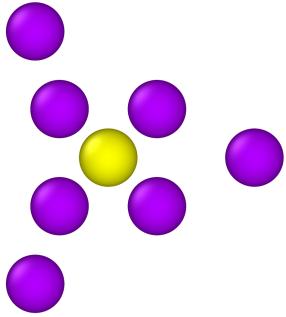
$$p(\mathbf{x}) = \sum_{k=1}^K \alpha_k \mathcal{N}(\mathbf{x} | \mathbf{m}_k, \mathbf{C}_k), \quad (5)$$

where α_k are mixture proportions also denoted as $p(\omega_k)$ and \mathbf{m}_k and \mathbf{C}_k are the mean vector and the covariance matrix of each Gaussian component ω_k . The mixture



	Cutoff	Cutoff Average	Voronoi	Voronoi Average
q_3	0.3388	0.0518	0.7529	0.2699
q_4	0.5650	0.4712	0.4048	0.4149
q_6	0.5108	0.3824	0.6868	0.3761
q_8	0.5447	0.3258	0.5337	0.3861

(a)



	Cutoff	Cutoff Average	Voronoi	Voronoi Average
q_3	0.3388	0.0690	0.0000	0.0979
q_4	0.5650	0.3774	0.8132	0.4814
q_6	0.5108	0.3219	0.5373	0.3227
q_8	0.5447	0.3234	0.7786	0.4551

(b)

Figure 4: Values of q_3 , q_4 , q_6 , and q_8 parameters for double-shell systems computed on the central atom (yellow) using different techniques.

proportions satisfy the conditions $0 \leq \alpha_k \leq 1$ and $\sum_{k=1}^K \alpha_k = 1$. In this way, Eq 5 represents the probability of finding an object \mathbf{x} from the mixture.

In most applications, the number of Gaussian components is selected automatically. Some of the criteria to choose the number of Gaussian components are the Bayesian Information Criterion (BIC) [71] and the Integrated Completed Likelihood (ICL) [72]. The number of Gaussian components is chosen such that the BIC and ICL functions are minimized. In this project, under the assumption that our database follows closely Gaussian distributions, we manually selected the number of Gaussian components, setting it equal to the number of structure types in our database.

2.2.2 Maximum Likelihood Classifier

In order to perform classification we would like to know the conditional probability $p(\omega_i|\mathbf{x})$ that object \mathbf{x}_i belongs to cluster ω_i . The Maximum Likelihood Classifier (MLC) makes use of Bayes' theorem [70] to compute the probability as

$$p(\omega_k|\mathbf{x}_i) = r_{ik} = \frac{p(\mathbf{x}_i|\omega_k)p(\omega_k)}{p(\mathbf{x}_i)} = \frac{\alpha_k \mathcal{N}(\mathbf{x}_i|\mathbf{m}_k, \mathbf{C}_k)}{\sum_{j=1}^K \alpha_j \mathcal{N}(\mathbf{x}_i|\mathbf{m}_j, \mathbf{C}_j)}, \quad (6)$$

also known as the responsibility r_{ik} that cluster ω_k has on object \mathbf{x} . Where the denominator of the fraction comes from Eq 5 and the numerator is the probability for just one Gaussian component. In the MLC, the responsibilities are then compared and an object is said to belong to the cluster whose responsibility is highest over the object. In this work, we slightly modify the classification rule and consider an atom to belong to a cluster only when its responsibility over the object is higher than 50%.

2.2.3 Expectation-Maximization Algorithm

In order to estimate the unknown parameters, the GMM uses the Expectation-Maximization (EM) algorithm [73]. The EM algorithm is an iterative technique to optimize the estimated parameters of a GMM. First, the parameters of the GMM are randomly initialized. Then, the expectation step is performed, in which the probability that each object belongs to a Gaussian component is computed with the current parameters. This is done using Eq 6. Then, in the maximization step, the parameters are updated based on the probabilities for each object. These steps are iteratively repeated until convergence [70].

3 Methodology

3.1 MD Simulation Protocol

The Molecular Dynamics (MD) simulations in this work were performed using the Large-scale Atomic/Molecular Massively Parallel Simulator (*LAMMPS*) software [74]. Two different MLIPs were utilized when indicated, the Physical LassoLars Interaction Potential (PLIP) without electrostatic interactions [35] and a more accurate PLIP potential with electrostatic interactions. For all the simulations the *metal* units were used, as defined by *LAMMPS*. The default integration timestep of 0.001 ps for *metal* units was used. Unless indicated, all simulations were performed using NVT time-integration with the Nose-Hoover thermostat [75]. The NVE time-integration method was also used together with the Langevin thermostat [76] as a tool to introduce randomness to a system during a simulation instead of resetting the velocities at the start. The simulations were performed on bulk and nanoparticle ZnO systems. The nanoparticle systems consisted of 2000 atoms in a cubic simulation box with a length of 60 Å. The nanoparticles had an approximate diameter of 30 Å and were therefore surrounded by a vacuum. The bulk systems consisted of around 2800 atoms in a box with an approximate length of 30 Å, which varied lightly among the crystalline systems used in this work.

3.2 Database

The database used for this project consisted of snapshots of NVT simulations of different crystalline systems and a liquid system. The ZnO crystals included are wurtzite (WRZ), zinc blend (ZBL), body-centered tetragonal (BCT), sodalite (SOD), h-BN (HBN), cubane (CUB), and rock salt (RCK). Since we are using a data-driven ML model, it is important that we not only include the perfect crystals but also physically feasible structures where the atoms vibrate around their crystalline position. For this purpose, NVT simulations were performed with the *LAMMPS* [74] software for all the systems using the PLIP potential [35]. The atomic structures of the perfect crystals are shown in Fig 5, for which the temperature was varied from 200K to 1500K during the NVT simulation. The maximum temperature was chosen in order to avoid melting the crystals. In the case of the liquid system, the NVT simulation was performed at a constant 2500K, for which a representative structure is illustrated in Fig 5h. 21 snapshots were stored for each of the NVT simulations with the purpose of representing the different structures found for each crystal at different temperatures.

For each of the snapshots in the database, we computed the averaged Voronoi weighted Steinhardt parameters using the *Pyscal* [77] library in *Python*. The Steinhardt parameters computed were q_2 to q_8 . Then, q_{mono} parameters were also computed by creating secondary systems consisting of only one atom type (zinc or oxygen) and computing the Steinhardt parameters as before. In this way, 14 Steinhardt parame-

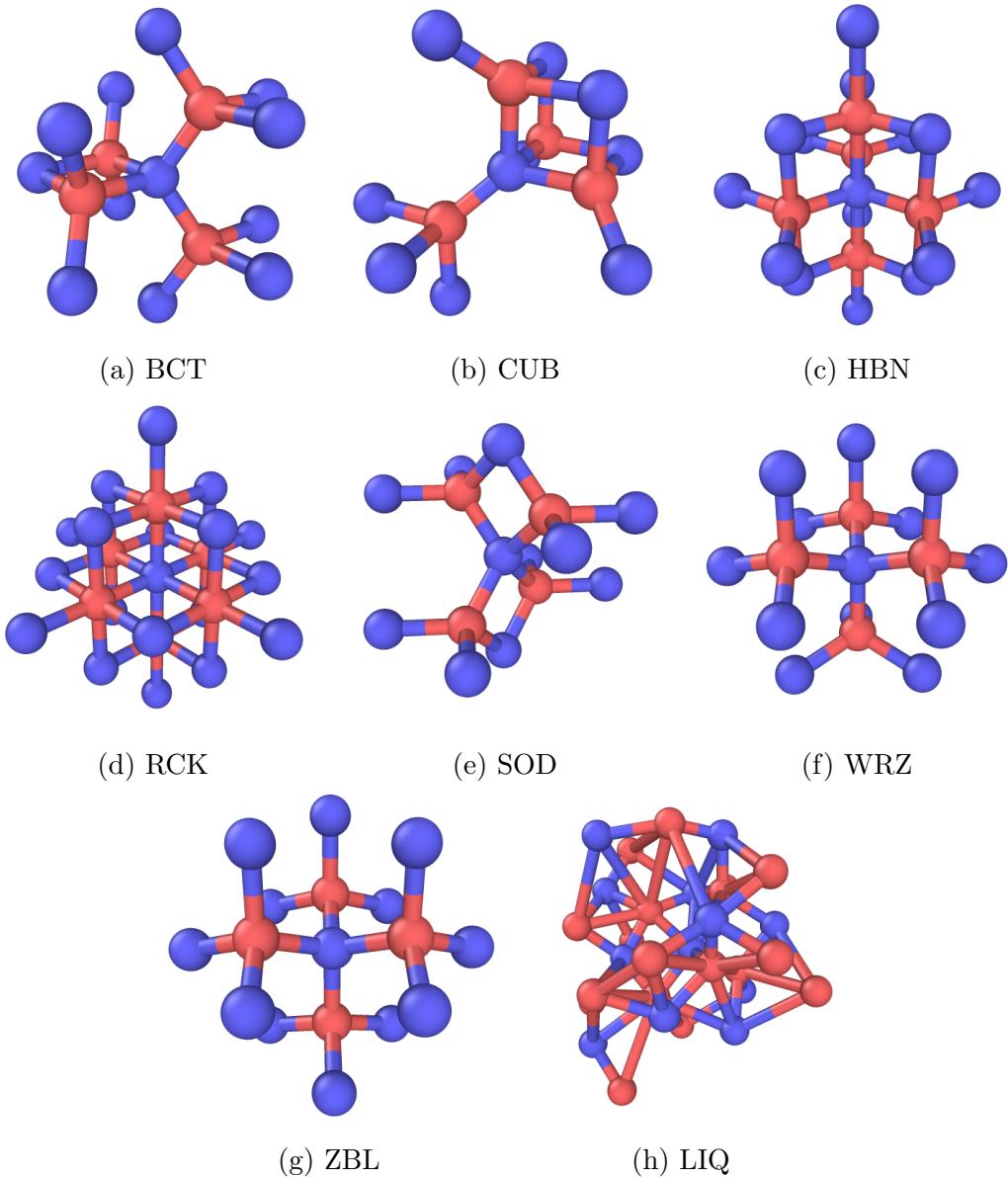


Figure 5: Atomic environments in the database including the second shell of neighbors selected using the Voronoi method. Zinc and oxygen atoms are colored red and blue respectively.

ters were computed for each atom in a snapshot. All Steinhardt parameters were later included in the clustering procedure, although we found by plotting different pairs of parameters that some are more useful than others. Fig 6c shows the distribution of q_4 and q_8 for the different systems in the database. It can be seen that q_4 and q_8 are already good candidates for the classification of the different crystal structures. However, the other values are still used to address the overlap between the clusters

in Fig 6c.

Fig 5 shows the structure of the different systems in the database as seen by our computation method. A central atom is chosen and the neighbors are selected using Voronoi tessellation. In order to include all the atoms that influence the Steinhardt parameter of each atom the second shell is also drawn, since it will be included when computing the average of Steinhardt parameters. In this way, the local neighborhood of atoms in each system is shown independently of the unit cell and as it will be seen by the clustering method. A visual comparison can be done between the neighborhood of the perfect crystal and atoms in other systems to confirm the predictions.

3.3 Classification

The clustering of the database was performed using the GMM as implemented in the *scikit-learn* [78] library in *Python*. The unknown parameters of the GMM were estimated using the Expectation-Maximization algorithm [73] iteratively. The GMM was trained using full covariance matrices and 100 k-means initializations. As opposed to other uses of the GMM in structural analysis, we do not select the number of Gaussian components automatically. We chose the number of Gaussian components equal to the number of structure types in our database. In this way, we give priority to the physical meaning of our database in which an NVT simulation is performed at temperatures at which the system should be crystalline, or transitioning towards liquid. Therefore, we trained our model using 8 Gaussian components, corresponding to the 7 crystals in our database and the liquid system. A non-automatic number of Gaussian components is justified by the Gaussian shapes in the database as seen in Fig 6c. For classification, the Maximum Likelihood Classifier is utilized. The responsibilities shown in Eq 6 are obtained when evaluating the model on a structure. These values can then be interpreted as the probability of belonging to a cluster in the model (the probability of an atom being in one of the 8 structure types in the database).

After the model is trained, it is tested by predicting the cluster of the perfect crystals and one liquid structure. In this way, the labels are obtained for the different Gaussian components and the training performance can be assessed. This is another way in which our method differs from previous uses of the GMM. We first train a model with a soft-labeled database. Since the clusters in our database are approximately Gaussian, it is expected that the EM algorithm correctly assigns a Gaussian component to each cluster. This model can then be used to analyze systems different than the ones encountered in the database.

3.4 Seeding Technique

The seeding technique is a method to sample rare events, more specifically nucleation in supercooled fluids. Traditionally, the process of crystallization is divided into two

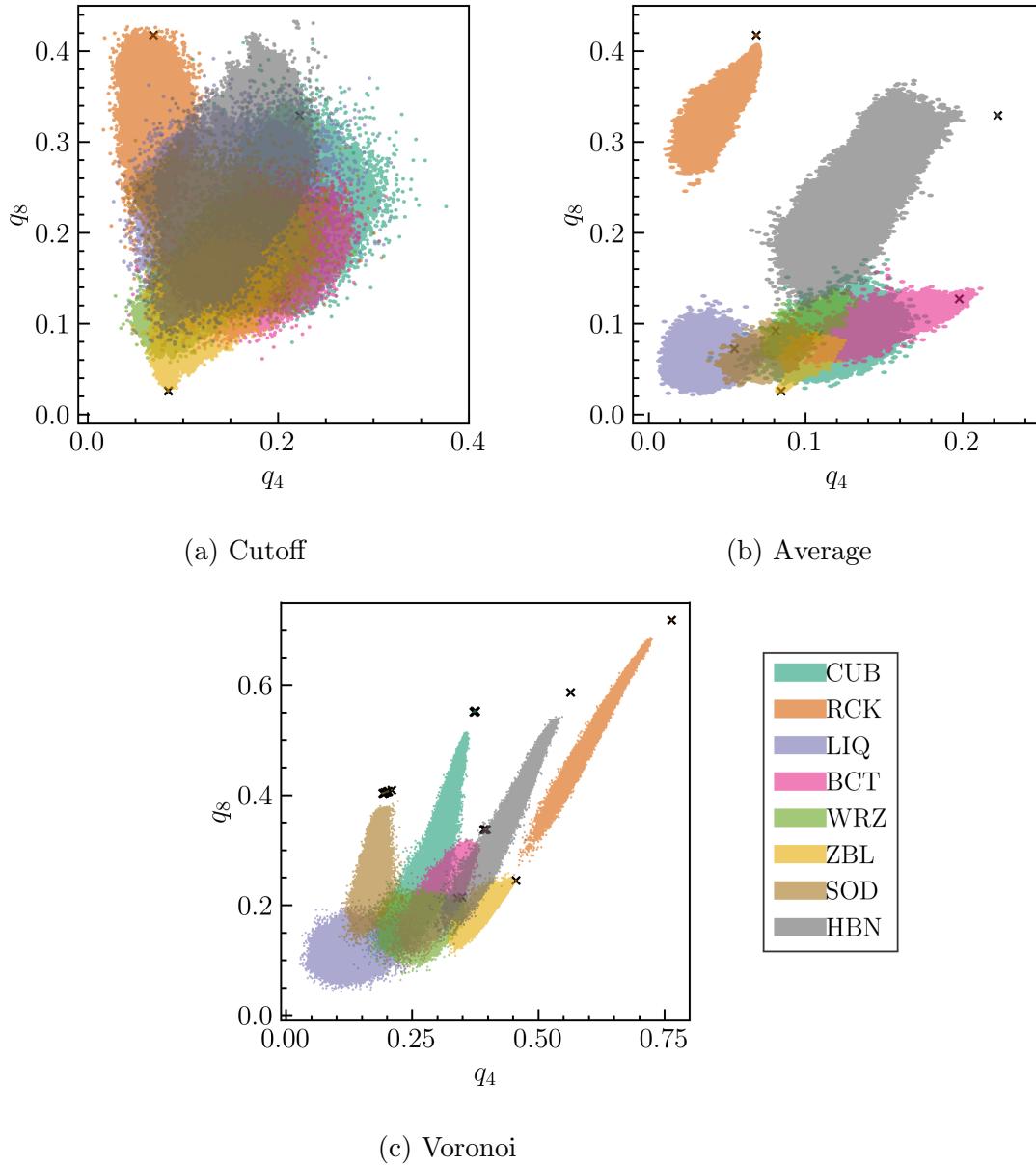


Figure 6: Scatter plot of the q_4 and q_8 values for different structures in the database. The dark crosses represent the q_4 and q_8 values for the perfect crystals. Using the simple Steinhardt parameters with a cutoff radius of 4 Å does not lead to easy discrimination of the different clusters, as shown in Fig (a). The discrimination is significantly improved by the average Steinhardt parameters, used in Fig (b). Meanwhile, the best discrimination for this database is achieved with average Voronoi-weighted Steinhardt parameters, shown in Fig (c).

phases, nucleation, and growth. From the two, nucleation is a rare event, since it involves overcoming a free energy barrier. Then, crystallization proceeds downhill in free energy once the nucleation barrier has been overcome [20]. The seeding technique focuses on the first phase of crystallization. It consists in manually inserting a crystal seed in the fluid and determining the temperature at which such clusters are critical [79]. By inserting a crystal seed, the energy barrier of nucleation is artificially overcome, allowing for the study of this rare event in shorter simulations. The seeding technique has been applied to the study of nucleation in conditions of shallow metastability [18, 79–84].

To study the nucleation of ZnO nanoparticles we used the seeding technique, based on the procedure described by Espinosa et al [79, 85], illustrated in Fig 7. This technique consists of inserting a crystalline cluster within a supercooled liquid and performing simulations at different temperatures while monitoring the size of the cluster. With this method, the critical temperature can be found, which is the temperature at which a crystalline cluster remains the same size on average. In this project, we used the seeding technique to find the critical temperature of WRZ and BCT crystalline clusters. This is done in order to address the competition that was previously observed in the bulk simulation illustrated in Fig 1. Furthermore, by applying the seeding technique on nanoparticles we are able to assess their preferred crystallization pathway.

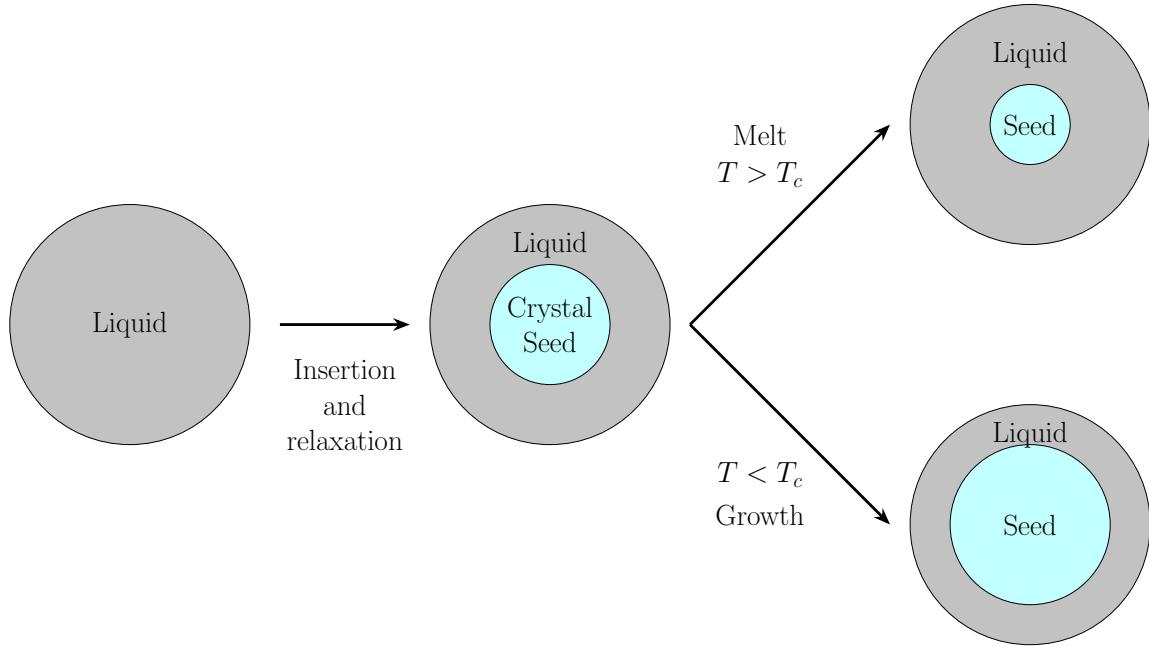


Figure 7: Schematic of the main steps in the seeding technique. First, a crystal seed is inserted into a liquid system. Then, the system is relaxed to allow the liquid to form an interface around the crystal. Lastly, simulations are performed at different temperatures, for which the crystalline cluster can increase or decrease in size.

3.4.1 Seed Creation

The first step of the seeding technique is to insert a crystalline cluster in a liquid droplet. To prepare a liquid droplet with an embedded crystal we used the *Ovito* [86] library in *Python*. Since a version of the PLIP potential with Coulomb interaction was used, one of the requirements of the system was that the number of zinc and oxygen atoms remained the same. A bulk crystal system and a liquid droplet system were used as starting points. In the droplet system, the N atoms closest to the zero coordinate were deleted, creating a cavity in the droplet. In the bulk system, all atoms were deleted except the N atoms closest to the zero coordinate, creating a spherical seed. These two systems could then be merged, creating a liquid droplet with a crystalline cluster inside. In order to avoid atoms being too close together, the liquid droplet was scaled by a factor ensuring that the minimum distance between cluster and droplet atoms was 1 Å.

3.4.2 Cluster Analysis

To count the number of atoms in the crystalline cluster we used again the *Ovito* [86] library in *Python*. After performing our classification method, we obtained probabilities for each atom to be liquid or crystalline. Using these values we select all the atoms that have a probability of being a liquid lower than 50%. From the atoms that are not liquid, the largest cluster is selected. This is done to avoid counting single crystalline atoms within the liquid as part of the crystalline cluster.

3.4.3 Relaxation

Since the crystalline cluster was artificially inserted in the droplet, the next step is to relax the system. The relaxation was performed in two steps. First, dynamics are performed on the liquid droplet, while keeping the crystalline cluster static. This was performed in *LAMMPS* by using the Nose-Hoover [75] thermostat only on the liquid droplet for 8 ps. The relaxation was performed at a temperature of 1350K, since after testing it was estimated to be lower than the critical temperature. By allowing only the liquid droplet to move, some liquid atoms will become part of the crystal. This increase is shown in step 1 of Fig 8. Furthermore, the gap created in the previous step between the crystal and the liquid will be filled during this first relaxation step.

During the second step of the relaxation, the atoms in the crystalline cluster are allowed to move and their temperature is increased. Two separate Nose-Hoover thermostats are used for 8 ps. One is applied to the liquid droplet and maintains a constant temperature of 1350K. The second one is applied to the crystalline cluster and does a temperature ramp from 100K to 1350K. During this step of the relaxation, we attempt to keep the size of the cluster constant. The purpose of this step is to slowly increase the temperature of the crystalline cluster until the desired temperature of 1350K is achieved for the whole nanoparticle.

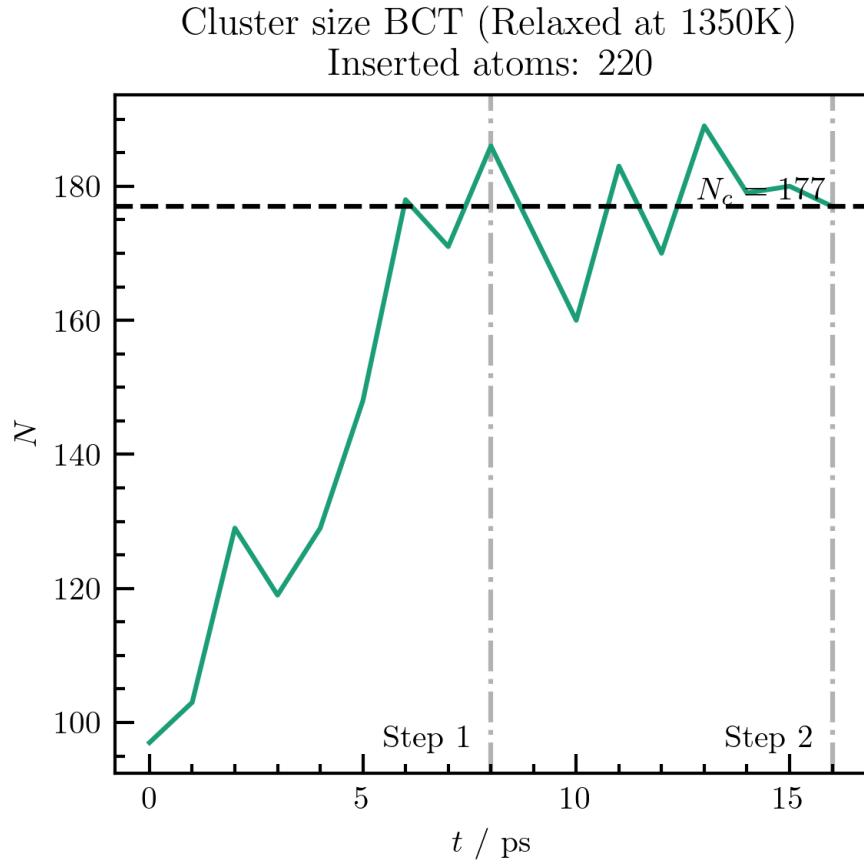


Figure 8: Crystalline cluster size as a function of time during the relaxation of the system.

3.4.4 Growth/Melting Simulation

The next step in the seeding technique is to perform simulations of the relaxed system at different temperatures and monitor the size of the cluster. In order to preserve the stability of the system it is important to preserve the atomic velocities from the relaxation step. Because of this, the Langevin thermostat [76] was used during 2 ps to add randomness to the system without resetting the atomic velocities. During the 2 ps, a temperature ramp is applied to the system to change the temperature from 1350K to the temperature to be studied. After the 2 ps temperature ramp, the traditional Nose-Hoover thermostat is used for 32 ps at a constant temperature. Thanks to the randomness introduced by the Langevin thermostat it is possible to run multiple simulations for the same temperature using a different random seed and compute the average and standard deviation. This is done to account for the stochasticity of the nucleation process and verify the validity of our results.

To study the critical temperature of clusters of different sizes we perform growth/melting

simulations at different temperatures. For each simulation, we monitor the size of the cluster. The critical temperature is then found between the lowest temperature at which the cluster shrinks and the highest temperature at which it grows. Ideally, the critical temperature is such that the slope of the cluster size as a function of time is zero. With this idea in mind we propose using the bisection algorithm [87] to sample different temperatures and find the zero of the slope as a function of temperature. To do this, a temperature range is selected based on information from preliminary simulations. Then, growth simulations are performed at the upper and lower limits of the range, as well as in the middle. Afterward, the cluster sizes are plotted and the temperature for the next simulation is chosen as the value in between the temperatures with opposite cluster size slope. With this method, we ensure that with each iteration we get closer to the critical temperature of the cluster. An example of this is shown in Fig 9 for the BCT critical cluster illustrated in Fig 8. First, the temperatures of 1375K, 1500K, and 1437.5K were sampled. Since at a temperature of 1437.5K, the cluster size remains more or less constant, we decided to sample both temperature ranges. In this way, we found that the slope at 1406.25K was closer to the slope at 1437.5K, and therefore chose it as the next sampling region. In this case, after sampling the temperature range between 1406.25K and 1437.5K, no significant difference could be observed in the slope, for this reason, the critical temperature was defined as $1421.875K \pm 15.625K$.

All nanoparticle simulations were performed using a modified version of the PLIP potential introduced in Ref [35]. This potential includes also electrostatic Coulomb interactions between the positively charged zinc atoms with the negatively charged oxygen atoms.

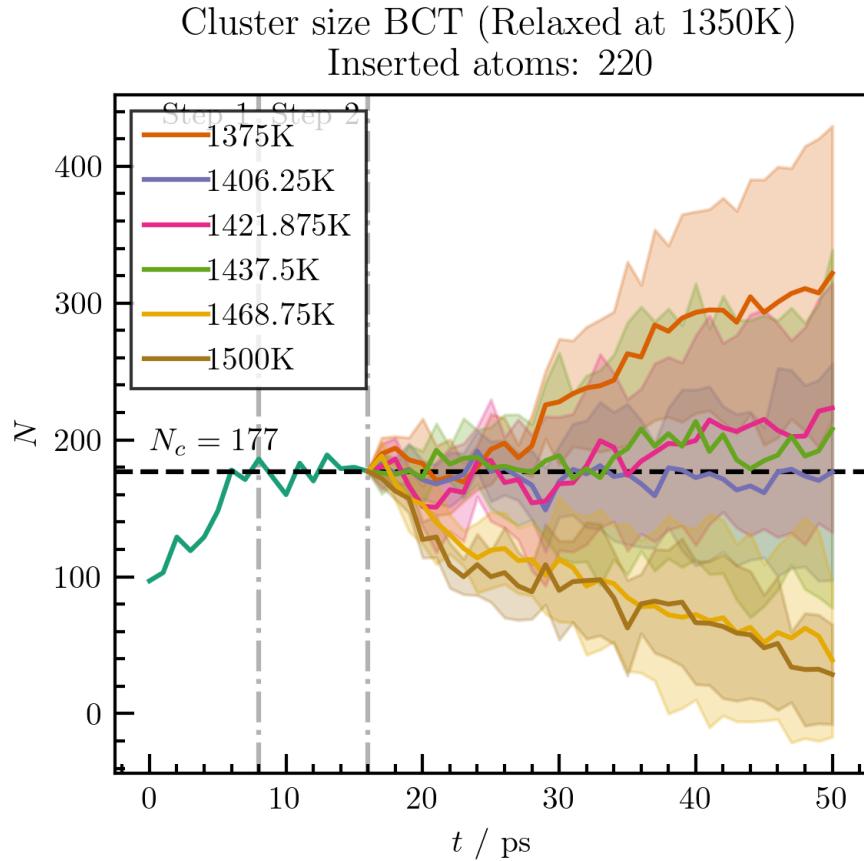


Figure 9: Crystalline cluster size as a function of time during the relaxation and growth/melting simulations.

4 Results and Discussion

4.1 Gaussian Mixture Model on Database

In order to validate the trained GMM, we performed an analysis on the same structures used during training. Since our database consisted of NVT simulations at low temperatures, we expect that throughout the simulation most crystals should still be considered to be crystals by our model. The results are shown for the first and last snapshots of the NVT simulations in Fig 10, where the atoms in the system are colored according to their probability to be a certain crystal. In most cases, the majority of the atoms are considered to be crystalline in the perfect crystal as well as at 1500K. This is expected since 1500K is a temperature lower than the melting temperature of ZnO. Similarly in the case of the liquid, after melting, no crystal structure is expected at the temperature of 2500K. In the case of cubane, there are many atoms that are

not predicted to be in the crystal phase and are predicted to be liquid. This may be an indication that our model is not good at recognizing lowly-ordered cubane, or that cubane melts at lower temperatures than the other crystal phases.

Next, the model was tested on a melting simulation of the WRZ bulk crystal. The melting was performed with an NVT simulation using the PLIP potential [35] without electrostatic interactions. The temperature in the simulation was varied from 1500K to 3000K, such that the crystal would completely melt. In a process like this, it would be expected for our model to correctly predict the beginning and ending states of the simulation. The probabilities for atoms to be in WRZ and liquid structure are shown in Fig 11 for multiple snapshots during the simulation. It can be seen that our model correctly predicts the starting WRZ crystal and the transition toward liquid. In the end, the system is almost completely predicted to be in the liquid state, as expected at a temperature of 3000K.

Lastly, to justify the usage of our classification method in the seeding technique we analyzed the structure of seeded nanoparticles before relaxation, shown in Figs 12 and 13 for BCT and WRZ crystal seeds. It was found that in both cases our model correctly identifies the crystalline cluster surrounded by atoms in the liquid phase. In the case of the BCT crystal seed, shown in Fig 12, the number of atoms predicted to be in the BCT phase is lower than the number of inserted atoms. This is expected from our model since at the interface between crystal and liquid, the atomic environments differ significantly from the environment illustrated in Fig 5a. Even though the atoms at the interface still present some order in their structure, according to the Maximum Likelihood Classifier described previously, they are considered to be closer to the liquid structure. Similarly, in the case of the WRZ crystal seed, shown in Fig 13, the number of atoms predicted to be in the crystalline phase is lower than the number of inserted atoms. This time, however, not all atoms in the crystalline cluster are predicted to be in the WRZ phase. At the surface of the crystalline cluster, some atoms are predicted to be in the HBN phase instead. This is not expected from our model, since all the inserted atoms were obtained from a WRZ bulk crystal. Due to this, we argue that atoms at the interface between liquid and WRZ phases may be mislabeled as being in an HBN structure. This is possibly caused by the similarity in WRZ and HBN environments observed in Figs 5f and 5c.

With these tests, we have shown the physical meaning of the predictions performed using the GMM. The different crystals were still predicted to be the same even after increasing the system's temperature. On the other hand, when increasing the temperature of a WRZ crystal to a temperature much higher than the melting temperature, the model correctly predicted the change in structure toward liquid. Lastly, crystalline atoms were correctly identified when surrounded by a liquid, with the exception of atoms at the interface between liquid and WRZ phases. These results confirm that the training parameters of the model were appropriately chosen, and the physical meaning of the database was retained.

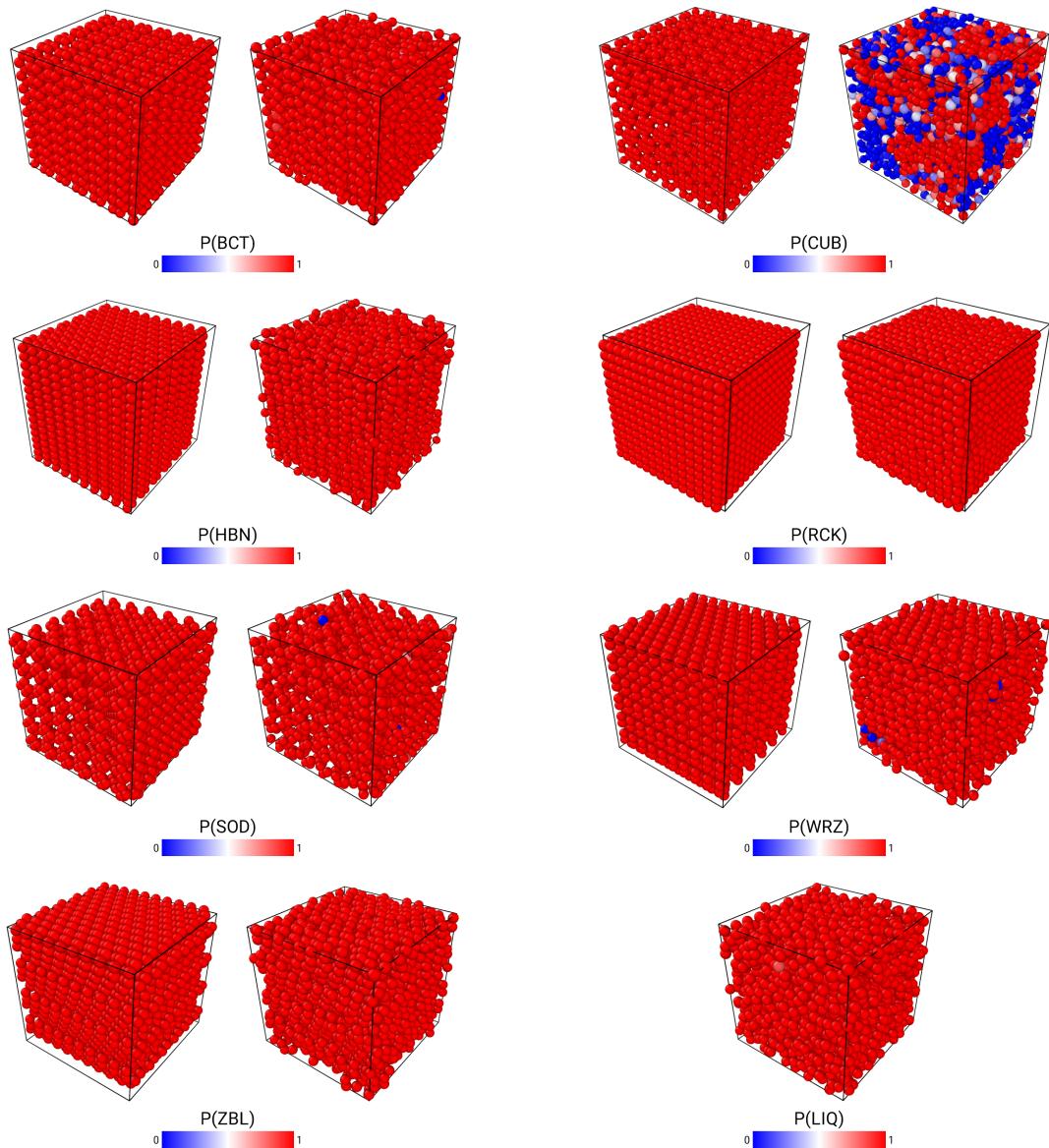


Figure 10: Atom probabilities to belong to different clusters during NVT simulations of the corresponding crystal (200K - 1500K). In the case of liquid, the simulation was performed at a constant 2500K.

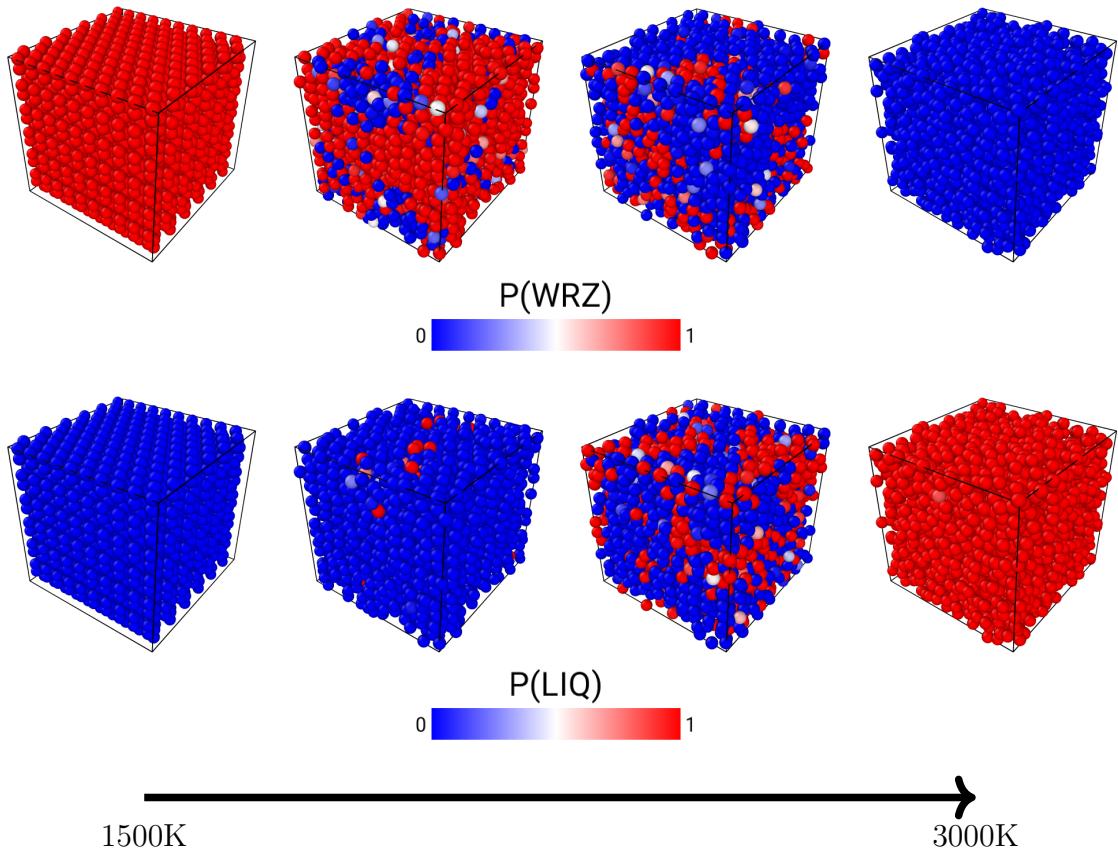


Figure 11: Atom probabilities to be in WRZ or LIQ structures during an NVT melting simulation (1500K - 3000K).

4.2 Melting Analysis

We then studied the melting of ZnO more in-depth by performing NVT simulations of the WRZ and BCT bulk crystals. The temperature was varied from 1000K to 3000K during three different durations of 10 ps, 50 ps, and 100 ps. We then used our classification method to count the number of atoms in LIQ, BCT, WRZ, and HBN configurations throughout the simulations. The results are shown in Figs 14 and 15.

First, we studied the melting of a WRZ bulk crystal. The number of atoms in the different configurations throughout the simulation is shown in Fig 14. At the beginning of the simulation, all atoms are predicted to be in the WRZ phase. As the temperature is increased, some atoms in the WRZ phase start changing into the BCT phase. It is important to note that this transition starts happening even before any atoms are predicted to be in the liquid phase. Then, the number of atoms predicted to be in the liquid phase starts increasing. After the system reaches a point where it is composed of one-third of WRZ atoms, one-third of BCT atoms, and one-third of liquid atoms, the number of atoms predicted to be in the BCT phase starts decreasing,

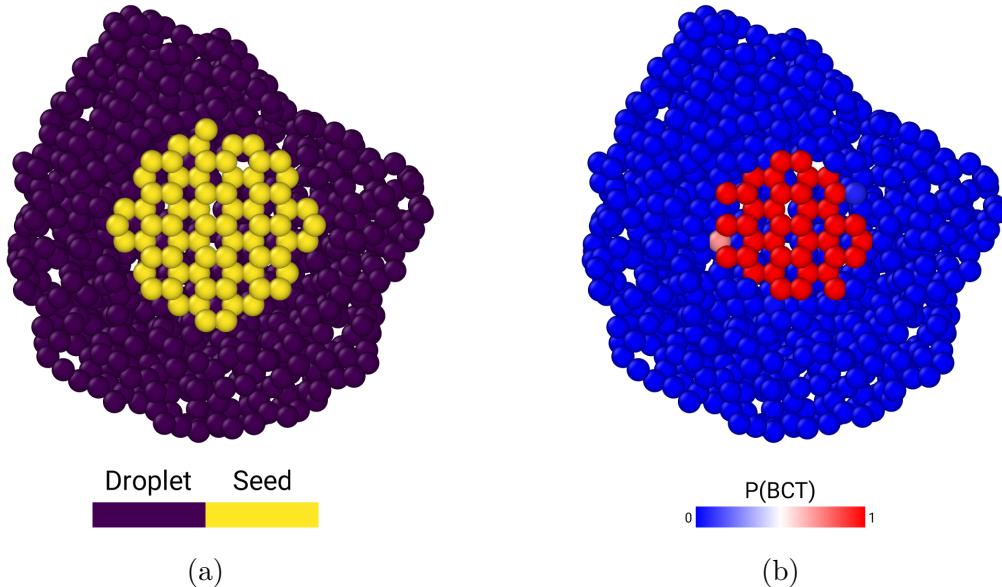
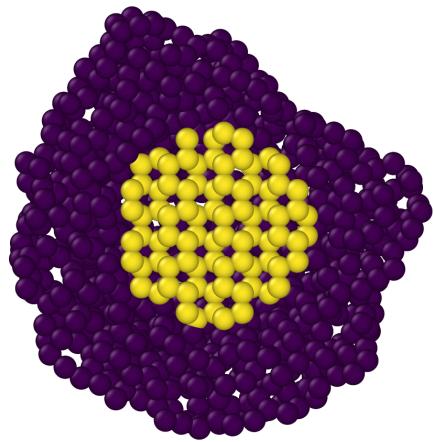


Figure 12: Liquid droplet system seeded with a BCT crystal. Fig (a) shows the atoms from the original droplet in purple, and the inserted crystalline atoms in yellow. The same system is shown in Fig (b) where the atoms are colored according to our model’s predicted probability to be in the BCT phase.

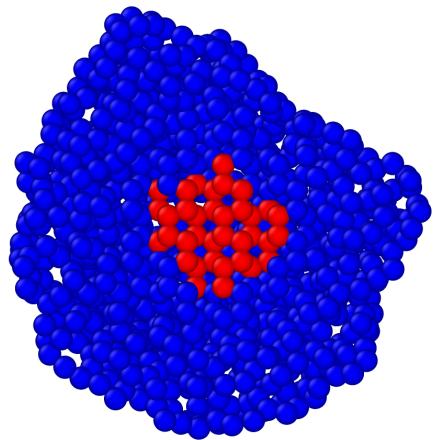
and all atoms in the system quickly transition to the liquid phase. The simulation was performed in different timescales while keeping all other parameters equal in order to ensure that the observed behavior was not an anomaly of fast melting. In all three simulations the behavior observed was the same, where, after reaching one-third of BCT, WRZ, and liquid atoms, all atoms quickly transition to the liquid phase. This result supports the nucleation results obtained in Ref [35], where the same two-step process was observed in nucleation. This supports the idea that there are two free energy barriers that need to be overcome when transitioning between WRZ crystal and liquid phases.

Then, we performed the same analysis on the melting of a BCT bulk crystal. As before, at the beginning of the simulation, all the atoms are predicted to be in the BCT phase. In this case, however, as temperature increases and the number of atoms in the BCT phase start decreasing, the number of atoms in the liquid phase starts increasing right away. This result was observed in the three timescales studied. However, in the longer simulations of 50 ps and 100 ps, there is a small number of atoms predicted to be in the WRZ phase. Since the number of atoms in the WRZ phase starts increasing just after some atoms in the liquid phase have already appeared it cannot be said that WRZ is involved in the transition between BCT and liquid. The appearance of WRZ can be attributed to the stochasticity of the process, in which some atoms cross the energy barrier in the opposite direction and transition



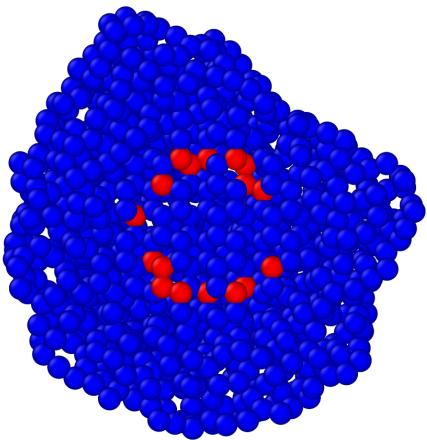
Droplet Seed

(a)



$P(\text{WRZ})$

(b)



$P(\text{HBN})$

(c)

Figure 13: Liquid droplet system seeded with a WRZ crystal. Fig (a) shows the atoms from the original droplet in purple, and the inserted crystalline atoms in yellow. The same system is shown in Figs (b) and (c) colored according to the atom's probability to be in the WRZ and HBN phases respectively, as predicted by our classification model.

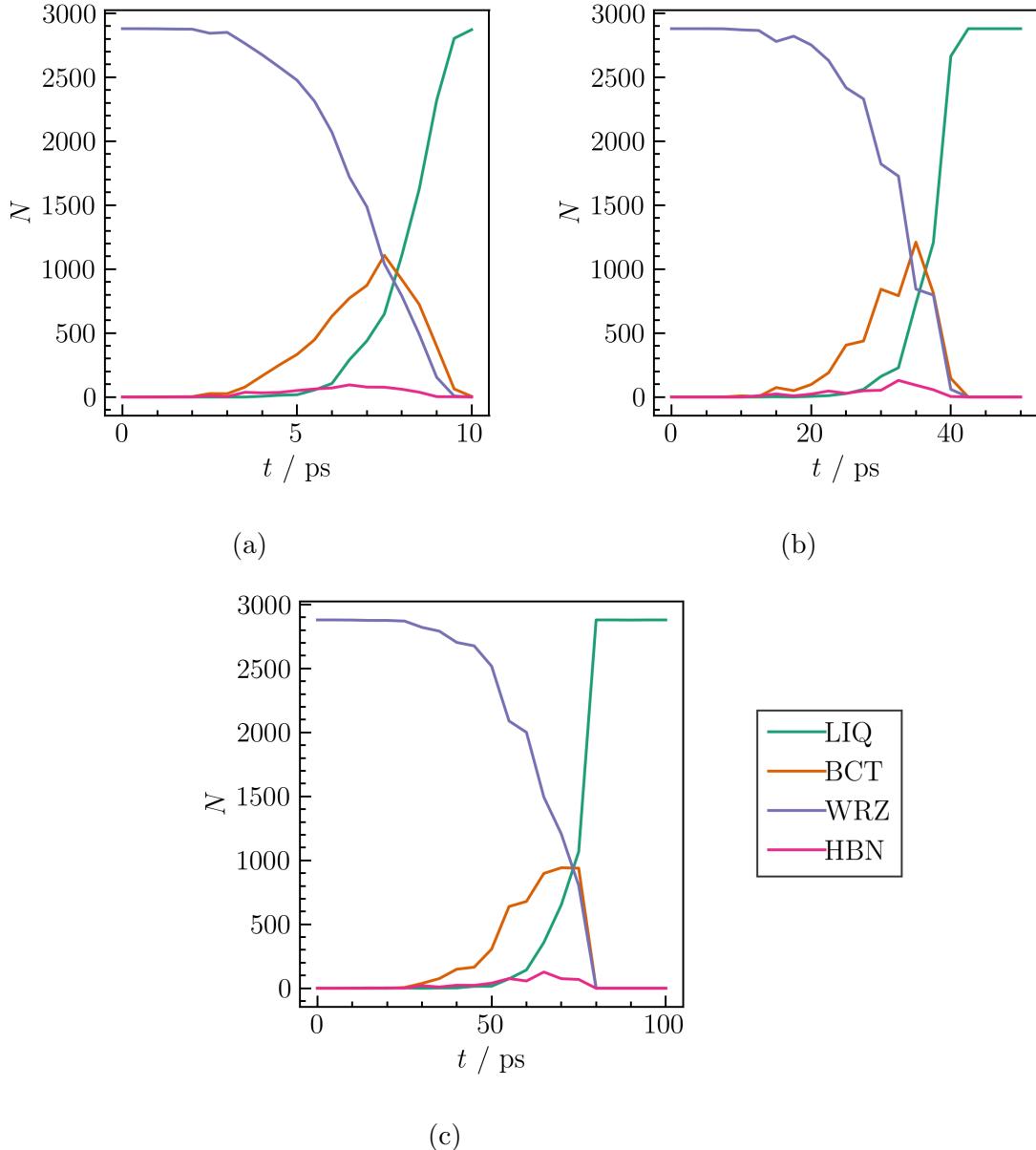


Figure 14: Count of atoms in LIQ, BCT, WRZ, and HBN configurations throughout the melting simulation of a WRZ bulk crystal with 2880 atoms. Results for simulations with a duration of 10 ps, 50 ps, and 100 ps are shown in Figs (a), (b), and (c) respectively.

to WRZ instead of liquid.

4.3 Relaxation

During the first step of the seeding technique, crystalline atoms are inserted in a liquid droplet, and then, the system is relaxed. The crystalline cluster at the end of the relaxation is the one for which the critical temperature will later be studied. The structures obtained at the end of the relaxation depend drastically on the initial velocities, and by artificially inserting a crystal seed we have control only over its initial size, but not over the final outcomes. Furthermore, the behavior throughout the relaxation was not consistent, and multiple relaxation simulations had to be discarded due to inconstant behavior in the second step of the relaxation. Ideally, we want to obtain relaxations like the one shown in Fig 8, where the size of the crystalline cluster increases in the first step, but remains more or less constant in the second step. Then, relaxations were chosen such that the crystalline cluster size was as similar as possible between BCT and WRZ, trying not to have differences higher than 5%. This would allow us to study the difference in critical temperature for clusters of similar size but different structures. The critical cluster size of the selected relaxation simulations is shown in Tab 1, and the corresponding relaxation curves are shown in the appendix in Fig A.1.

	WRZ			BCT		
	N_{inserted} 100	N_{inserted} 210	N_{inserted} 390	N_{inserted} 140	N_{inserted} 220	N_{inserted} 380
N_c	71	186	366	72	177	364

Table 1: Number of crystalline atoms inserted in the droplet and the size of the cluster after relaxation.

4.4 Growth and Critical Temperature

With the purpose of shedding light on ZnO nucleation, we then studied the critical temperature of BCT and WRZ crystal clusters. To do this, NVT simulations were performed at different temperatures using the Nose-Hoover thermostat. These simulations were initialized with a short NVE simulation using the Langevin thermostat in order to add randomness to the results. In this way, the average and standard deviation could be computed and we could assess whether for a given temperature the cluster increased or decreased in size or remained the same. The critical temperature was then found in the range between the lowest temperature that makes the cluster melt and the highest temperature that makes it grow. The average between these two temperatures was taken as the critical temperature and the bounds of the range were taken as an error range. The growth/melting curves from which

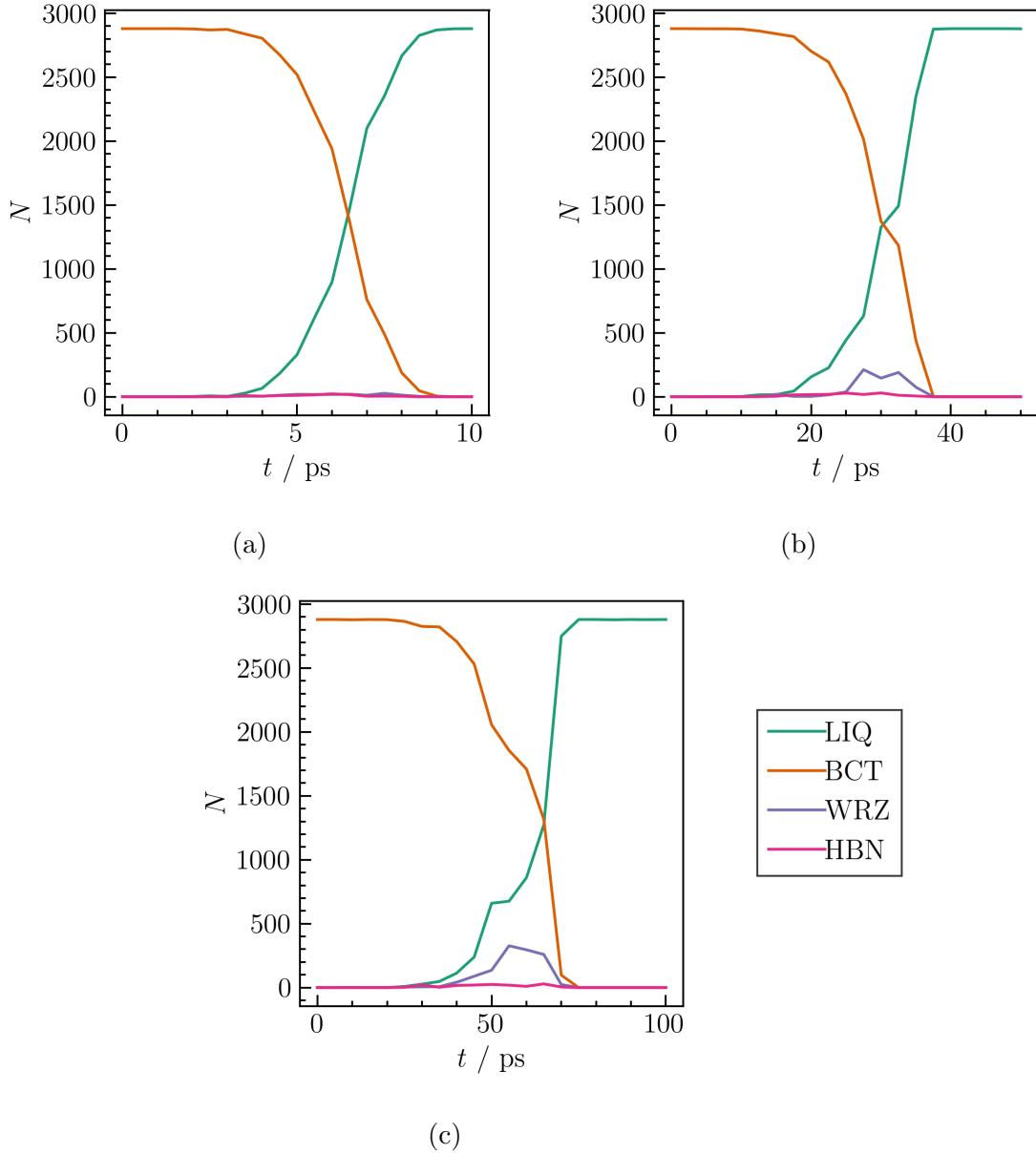


Figure 15: Count of atoms in LIQ, BCT, WRZ, and HBN configurations throughout the melting simulation of a BCT bulk crystal with 2880 atoms. Results for simulations with a duration of 10 ps, 50 ps, and 100 ps are shown in Figs (a), (b), and (c) respectively.

the critical temperatures were estimated are shown in the appendix in Figs A.2, A.3, and A.4. The critical temperatures found for different crystalline cluster sizes are shown in Fig 16. It can be seen that for smaller clusters, the difference in critical temperature between BCT and WRZ is significative, being higher for BCT. As the size of the cluster increases, it becomes harder to discern a difference in temperature between the two crystals. This means that for small seeds (around 70 atoms), there is a temperature range at which it is easier to grow a BCT crystal seed than a WRZ crystal seed, while for bigger seeds they are both similarly likely to grow.

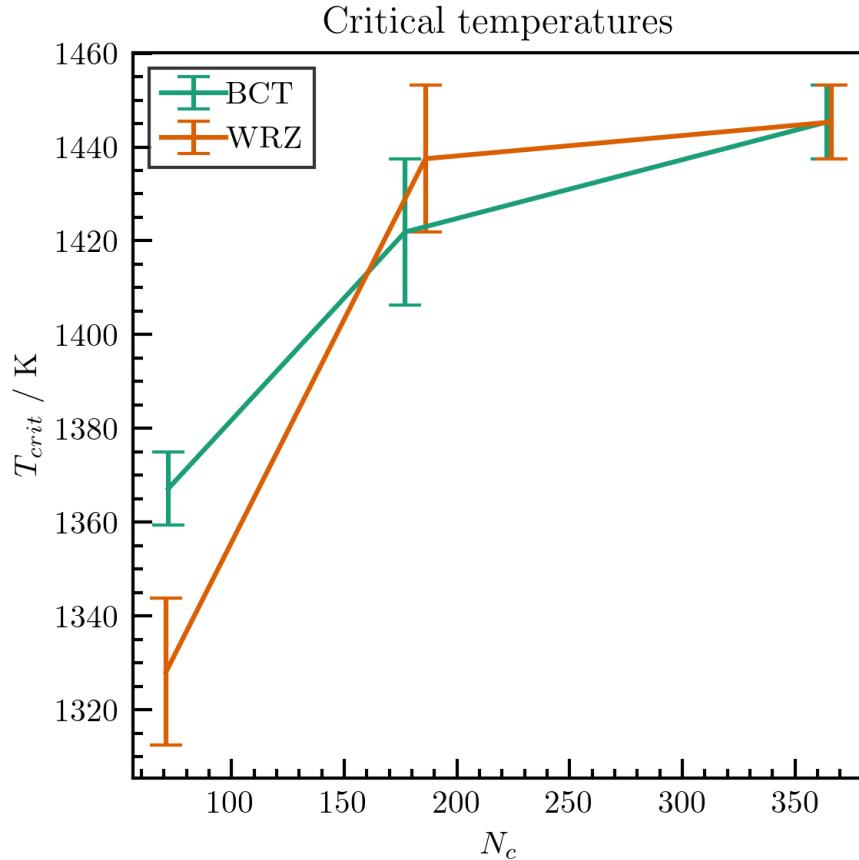


Figure 16: Critical temperature as a function of critical cluster size for BCT and WRZ seeds.

4.4.1 BCT Cluster Growth

During the growth simulations performed to find the critical temperature of the BCT critical clusters, it was found that a transition to WRZ occurred in most cases. An example of this is shown in Fig 17 for the BCT cluster of 364 atoms at 1375K, which

grows as shown in Fig A.4a. It can be seen that the composition of the crystalline cluster is completely different at the beginning and at the end of the simulation. At 50 ps, the center of the crystalline cluster is composed mostly of atoms in the WRZ structure, while the outer layers are composed of atoms in HBN and BCT structures. This observation suggests that the preferred crystallized structure for ZnO in these conditions is WRZ. With the current scope of this work, we are not able to comment on the physical meaning of the appearance of the HBN structure in the crystalline cluster. As it can be observed in Figs 5c and 5f, the atomic environment as seen by our method is similar for HBN and WRZ atoms. Due to this, it is possible that atoms in the WRZ structure are sometimes mislabeled as being in the HBN structure. This was also observed during the relaxation of the WRZ clusters, as atoms in the interface between WRZ and liquid are many times predicted to be in the HBN structure, as shown in Fig 13c.

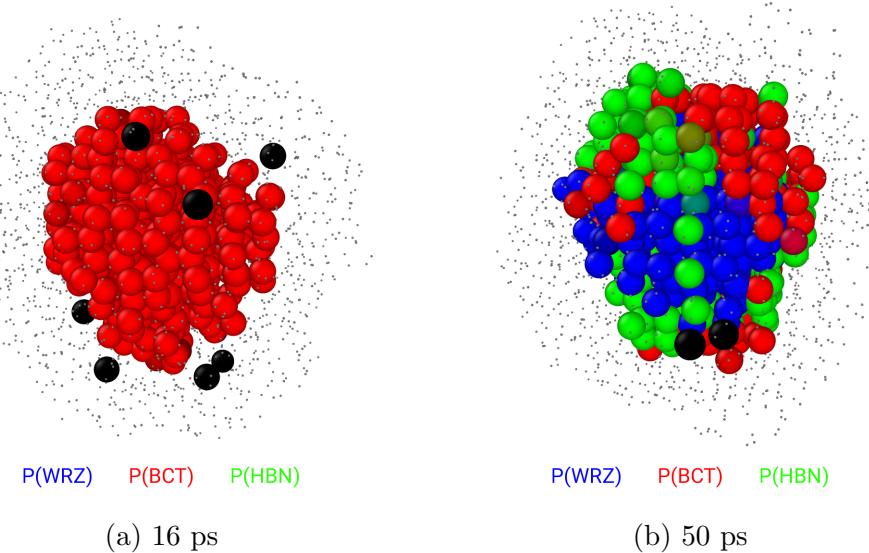


Figure 17: Atoms in the BCT cluster during a growth simulation at 1375K (initial size of 364 atoms), colored according to their probability to be in BCT (red), HBN (green), or WRZ (blue) phases. The atoms that do not form part of the cluster have been reduced in size and colored gray for clarity. The few atoms that are colored black form part of the cluster, but are predicted to be in neither WRZ, BCT, or HBN phases. In this case, these atoms are predicted to be in SOD structure.

To do a more in-depth analysis of the transition between BCT and WRZ we studied the atom counts for each structure as a function of time during the growth simulation, shown in Fig 18 for three different cluster sizes. In all three cases, it can be seen that the total number of atoms in the crystalline cluster increases as a function of time, as expected for temperatures lower than the critical temperature. In the beginning, almost all the atoms are predicted to be in the BCT phase. Afterward,

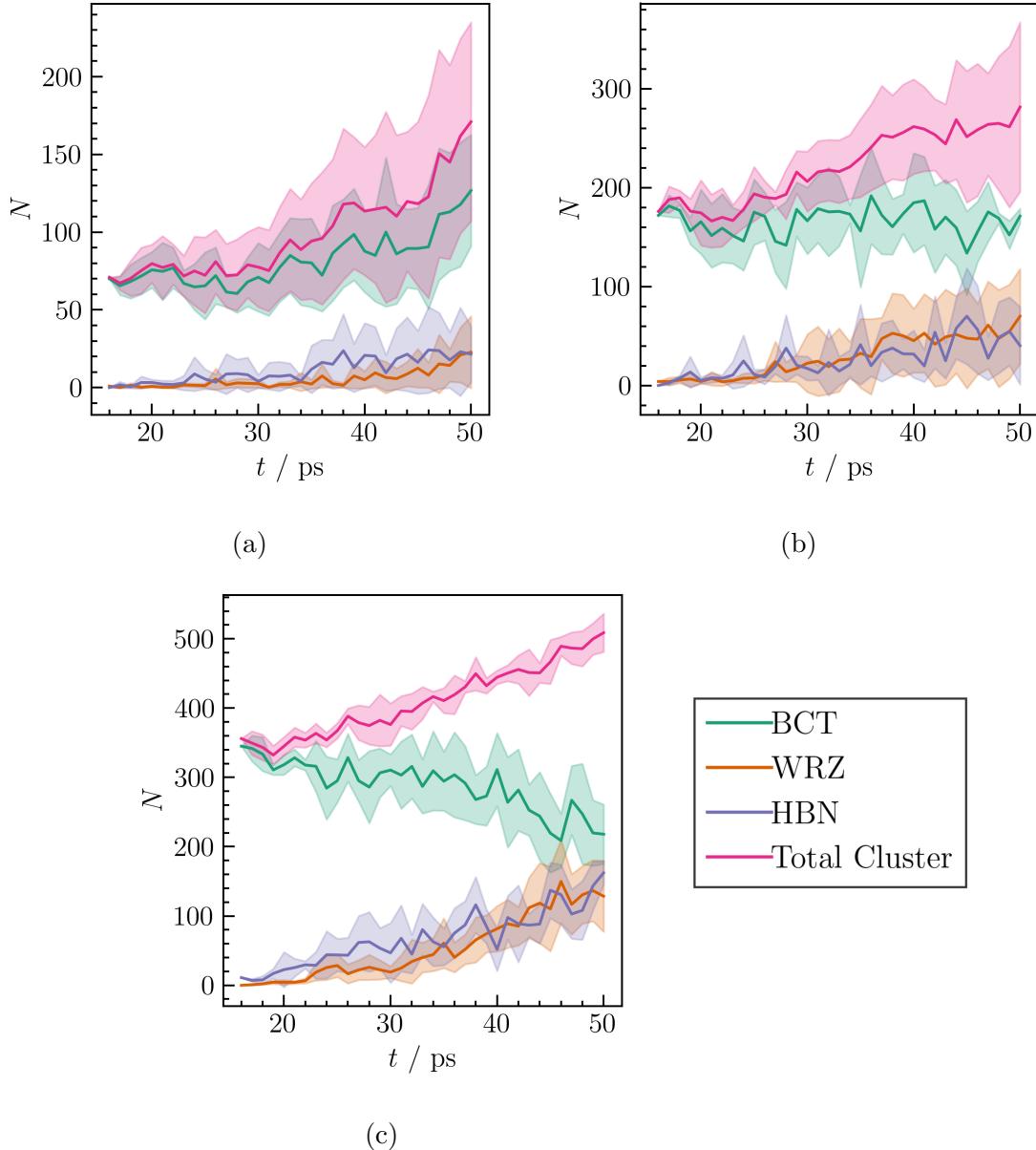


Figure 18: Count of atoms in BCT, WRZ, and HBN configurations, as well as the total number of atoms in the cluster throughout the growth simulation of BCT clusters. Figs (a), (b), and (c) show the atom counts for the clusters of 72, 177, and 364 atoms respectively. For the cluster of 72 atoms, the results shown correspond to the growth simulation at 1250K, while for the other two sizes, the results correspond to the growth simulations at 1375K.

the growing behavior differs for each of the three crystalline cluster sizes, most notably when observing the change in the number of atoms in the BCT phase. For the smallest cluster, the number of atoms in the BCT phase increase with time, suggesting that the crystallization of BCT is the main mechanism for cluster growth. In the medium-sized cluster, the number of atoms in the BCT phase remains more or less constant, while for the biggest cluster, the number of atoms in the BCT phase decreases since the beginning of the simulation. This suggests that for larger clusters, the main mechanism for growth is the crystallization of WRZ and HBN, while BCT becomes unstable.

4.4.2 WRZ Cluster Growth

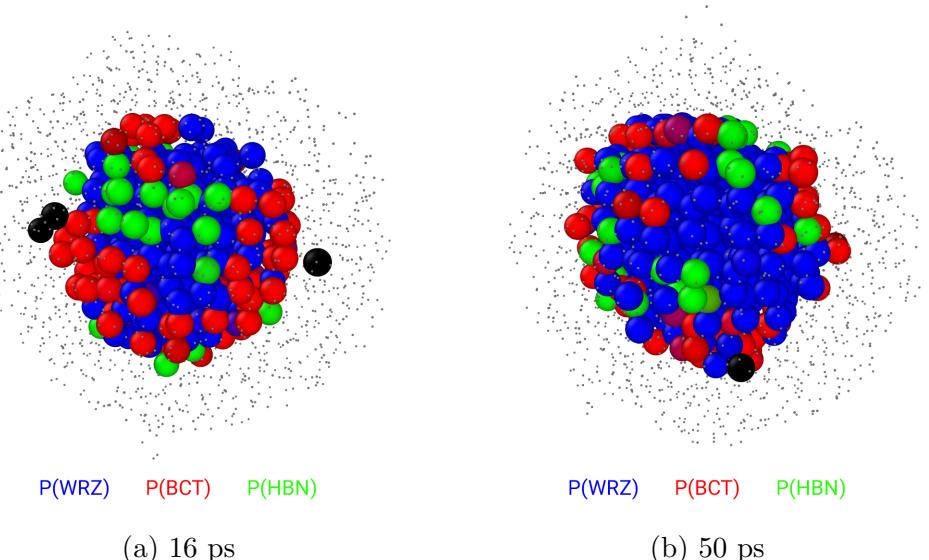


Figure 19: Atoms in the WRZ cluster during a growth simulation at 1375K (initial size of 366 atoms), colored according to their probability to be in BCT (red), HBN (green), or WRZ (blue) phases. The atoms that do not form part of the cluster have been reduced in size and colored gray for clarity. The few atoms that are colored black form part of the cluster, but are predicted to be in neither WRZ, BCT, or HBN phases.

The growth of WRZ clusters was also studied more in-depth. During the growth simulations, we could remark that the cluster composition was consistent for the different sizes. In all cases, the cluster core was composed of atoms in the WRZ phase. On the other hand, the outer layers of the cluster are many times composed of atoms in BCT and HBN phases, as shown in Figs 19 and 20, corresponding to the growth simulations at 1375K and 1250K in Figs A.4b and A.2b respectively. As pointed out before, the presence of atoms predicted to be in the HBN phase at

the surface of the crystalline cluster might be misleading due to mislabeling at the interface between WRZ and liquid phases. On the other hand, the presence of atoms in the BCT phase is more justified since it was shown in the previous results that BCT is a middle step in the transition from WRZ to liquid phases. Furthermore, the crystallization of supercooled fluid was shown to be a two-step process involving BCT and WRZ phases [35]. The fact that the atoms in the BCT phase are mostly located on the surface of the crystalline cluster suggests that the growth of the WRZ cluster is preceded by the formation of an unstable BCT crystal at the interface with the liquid. Similar behavior can be observed also for smaller clusters, as shown in Fig 20, but with a less regular shape. Further simulations can be dedicated to confirming the growth of the WRZ cluster through an unstable BCT crystal at the interface with the liquid phase.

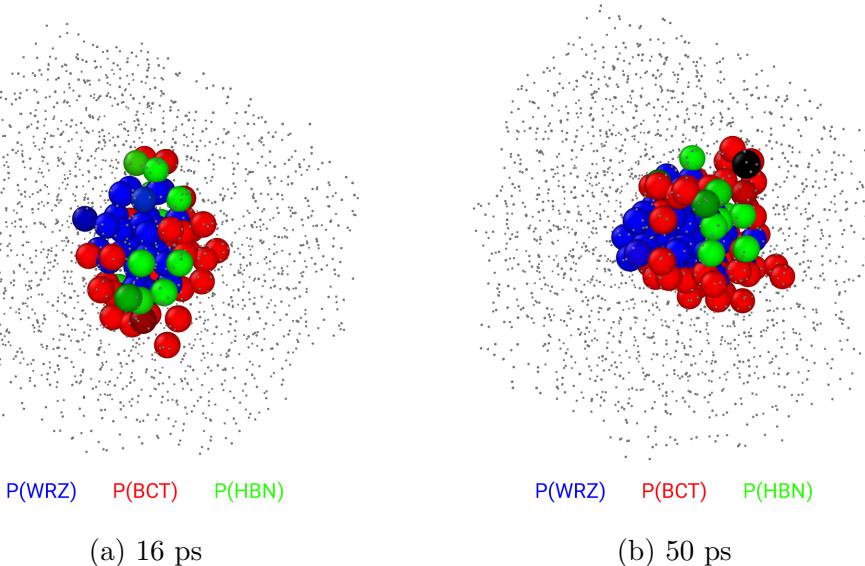


Figure 20: Atoms in the WRZ cluster during a growth simulation at 1250K (initial size of 71 atoms), colored according to their probability to be in BCT (red), HBN (green), or WRZ (blue) phases. The atoms that do not form part of the cluster have been reduced in size and colored gray for clarity. The few atoms that are colored black form part of the cluster, but are predicted to be in neither WRZ, BCT, or HBN phases.

5 Conclusion and Outlook

The study of the melting and crystallization of ZnO in nanoparticles is a multi-faceted problem. Firstly, to study transitions between different phases it is necessary to be able to distinguish and classify the phases as they appear in a system. This is the task of structural analysis, which in ZnO systems becomes a complex task due to the multiple crystal structures that are available. Secondly, in the case of crystallization, a free energy barrier needs to be overcome for nucleation to occur. Because of this, nucleation is a rare event that requires long simulations to be observed. Thirdly, ZnO, as many other systems studied in recent years, presents multi-step crystallization and nucleation processes. This complicates the analysis since classical study techniques cannot be applied to this system, such as Classical Nucleation Theory.

In this work, we have presented a new methodology for structural analysis and implemented it in the study of the melting and crystallization of ZnO. In our method, we used the Steinhardt parameters as descriptors due to their popular application in the study of ordered matter. In specific, we used the average Voronoi weighted variant of the Steinhardt parameters as it has been shown to present improvements in the detection of different ordered phases of matter and allow for the automatic detection of neighboring atoms. For classification, we used the Gaussian Mixture Model to fit our database with Gaussian clusters and the Maximum Likelihood Classifier to assign clusters to the studied atoms. Our database consisted of NVT simulations of 7 different crystalline phases and the liquid phase. After testing, we showed that our structural analysis method was able to make physically meaningful predictions of structures in ZnO systems.

With our structural analysis method, we set off to study the melting of both BCT and WRZ bulk crystals. For this purpose, we performed NVT simulations in a temperature range from 1000K to 3000K. In the WRZ melting simulation, we observed a transition of WRZ crystalline atoms to BCT crystalline atoms before the formation of the liquid phase started. This result was validated for three different timescales, confirming that it is not an artifact of slow melting. This result is in agreement with previous work on the nucleation of ZnO in which it was shown that the crystallization of ZnO in the WRZ phase is a two-step process [35]. The crystallization and melting results support the presence of an Ostwald step mechanism in the transition between WRZ and liquid phases, where the intermediary free energy barrier represents the transition to the BCT phase. Furthermore, the existence of a critical point was observed. After the count of atoms in the WRZ phase decreases and meets the increasing count of atoms in the BCT and liquid phases, the number of atoms in the liquid phase sharply increases, and the number of atoms in either the BCT or WRZ phases declines. The melting of the BCT bulk crystal was also studied. It was found that the transition between BCT and liquid phases occurs directly without an intermediary phase. Similarly, this behavior was reproduced at three different timescales. In the longer melting simulations, a small portion of atoms is predicted

to be in the WRZ phase, around 10% of the total number of atoms. This can be attributed to the stochasticity of the system, due to which some atoms might cross the free energy barrier in the opposite direction, transitioning from the BCT phase to the WRZ phase.

To measure the critical temperature of ZnO we implemented the seeding technique to study the critical temperature of BCT and WRZ crystalline clusters. The BCT and WRZ clusters were chosen for this study since they were shown to be involved in the crystallization of ZnO from a supercooled liquid. This was performed by first substituting atoms in a liquid droplet with crystalline atoms in either BCT or WRZ structure. The resulting structure is in many cases not stable, so a relaxation needs to be performed. During this relaxation, the crystalline atoms are kept static, allowing the liquid atoms to move and adjust around the crystalline cluster. The relaxation is then continued by allowing all the atoms to move, but slowly increasing the temperature of the crystalline cluster to avoid disrupting the structure. Growth/melting simulations are then performed on the crystalline cluster to estimate its critical temperature. In this way, we are able to study the properties of nucleation in the system without having to wait for a free energy barrier to be crossed.

Then, we set to study the BCT and WRZ crystallizations in a liquid droplet of 2000 atoms. For this, the seeding technique was used to create crystalline clusters and estimate their critical temperatures. It was found that for a cluster size of approximately 70 atoms, the difference in critical temperature between BCT and WRZ is significative. At this size, the BCT cluster has a critical temperature of $1367.1875K \pm 7.8125K$, while the WRZ cluster has a critical temperature of $1328.125K \pm 15.625K$. Meanwhile, for bigger critical clusters a significant difference was not observed. This means that for clusters of around 70 atoms in size, there is a temperature range at which a BCT crystal would continue to grow and the WRZ cluster would melt and disappear. We then performed further analysis to understand the mechanisms through which the different crystalline clusters grow. For the BCT cluster, it was observed that it grows mainly as BCT only in the smallest cluster studied. For the medium and big clusters, the number of atoms predicted to be in the BCT phase decreased or remained more or less the same. In these cases, the growth of the clusters was driven by an increase in the number of atoms in the WRZ and HBN phases, indicating as well a transition of the cluster from the BCT phase to the WRZ phase. On the other hand, during the growth simulation of the WRZ critical cluster, the number of atoms in the WRZ phase always increased as the cluster grew. However, when doing a visual analysis of the WRZ cluster, it was observed that commonly the WRZ cluster contains also atoms in the BCT and HBN phases at its surface. This indicates again the two-step nature of the WRZ crystallization process where liquid atoms transition to the BCT phase before transitioning once more to the WRZ phase.

The results obtained in this work shed light on the melting and nucleation processes of ZnO. In the case of melting, the results support previous findings in the bulk crystallization [35]. The utilization of the seeding technique on the formation

of nanoparticles helped us estimate critical temperatures for BCT and WRZ clusters of different sizes. By performing an analysis of the composition of the crystalline clusters we were able to speculate on the growth mechanism of the different clusters. Future work will be dedicated to studying in-depth the growth mechanism of the crystalline clusters. A few questions have been raised from the results of this work. Firstly, the stability of the BCT cluster in the nanoparticle was put into question. Longer growth simulations can be performed to study the composition of the cluster at different sizes. Furthermore, the energy of the completely crystallized nanoparticle can be computed and compared for BCT and WRZ crystal structures, allowing us to assess the stability of each structure. Secondly, the stability of the WRZ cluster for a small size was observed to be lower than that of BCT, due to their different critical temperatures. This result can be confirmed by performing the same analysis on smaller clusters and studying also the composition of the clusters throughout the simulations.

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. Julien Lam, for his invaluable guidance and support throughout my internship. His expertise and encouragement helped me to complete this research and write this thesis. I would also like to thank Dr. Akshay Krishna for his support and meaningful discussions during this project. They helped me give shape and orientation to the work dedicated to this project. I would also like to thank all my professors. Without them, I would not be able to understand the topics I can understand today, and I would not have the passion for science required to enjoy it. I would also like to thank my family, who have always been there providing moral support. I would like to thank my dad for the opportunity he gifted me of undertaking my higher education abroad, as I have always dreamt about when I was a kid. I want to thank my mother, who is now resting in peace, for her titanic effort in making sure that I had the best education possible. I want to thank my sisters for always being there for me and being my closest friends. I want to thank my dogs for giving color to my life and always making me smile. I want to thank my friends who make my life better and motivate me to keep moving forward. Lastly, I would like to thank my girlfriend for all her love and support during the past two years.

6 References

- [1] Pieter Rein ten Wolde, Maria J. Ruiz-Montero, and Daan Frenkel. “Numerical Evidence for bcc Ordering at the Surface of a Critical fcc Nucleus”. In: *Phys. Rev. Lett.* 75 (14 Oct. 1995), pp. 2714–2717. DOI: [10.1103/PhysRevLett.75.2714](https://doi.org/10.1103/PhysRevLett.75.2714).
- [2] Pieter Rein ten Wolde and Daan Frenkel. “Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations”. In: *Science* 277.5334 (1997), pp. 1975–1978. DOI: [10.1126/science.277.5334.1975](https://doi.org/10.1126/science.277.5334.1975).
- [3] Stefan Auer and Daan Frenkel. “Prediction of absolute crystal-nucleation rate in hard-sphere colloids”. In: *Nature* 409 (Mar. 2001), pp. 1020–3. DOI: [10.1038/35059035](https://doi.org/10.1038/35059035).
- [4] James F. Lutsko and Julien Lam. “Long-wavelength density fluctuations as nucleation precursors”. In: *Phys. Rev. E* 101 (5 May 2020), p. 052122. DOI: [10.1103/PhysRevE.101.052122](https://doi.org/10.1103/PhysRevE.101.052122).
- [5] James F. Lutsko. “How crystals form: A theory of nucleation pathways”. In: *Science Advances* 5.4 (2019), eaav7399. DOI: [10.1126/sciadv.aav7399](https://doi.org/10.1126/sciadv.aav7399).
- [6] Mike Sleutel and Alexander E. S. Van Driessche. “Role of clusters in nonclassical nucleation and growth of protein crystals”. In: *Proceedings of the National Academy of Sciences* 111.5 (2014), E546–E553. DOI: [10.1073/pnas.1309320111](https://doi.org/10.1073/pnas.1309320111).
- [7] Alexander E S Van Driessche, Nani Van Gerven, Paul H H Bomans, Rick R M Joosten, Heiner Friedrich, David Gil-Carton, Nico A J M Sommerdijk, and Mike Sleutel. “Molecular nucleation mechanisms and control strategies for crystal polymorph selection”. In: *Nature* 556.7699 (Apr. 2018), pp. 89–94. ISSN: 0028-0836. DOI: [10.1038/nature25971](https://doi.org/10.1038/nature25971).
- [8] Sarath Menon, Grisell Díaz Leines, Ralf Drautz, and Jutta Rogal. “Role of pre-ordered liquid in the selection mechanism of crystal polymorphs during nucleation”. In: *The Journal of Chemical Physics* 153.10 (Sept. 2020). 104508. ISSN: 0021-9606. DOI: [10.1063/5.0017575](https://doi.org/10.1063/5.0017575).
- [9] Julien Lam and James F. Lutsko. “Lattice induced crystallization of nanodroplets: the role of finite-size effects and substrate properties in controlling polymorphism”. In: *Nanoscale* 10 (10 2018), pp. 4921–4926. DOI: [10.1039/C7NR08705E](https://doi.org/10.1039/C7NR08705E).
- [10] Jonathan Amodeo, Fabio Pietrucci, and Julien Lam. “Out-of-Equilibrium Polymorph Selection in Nanoparticle Freezing”. In: *The Journal of Physical Chemistry Letters* 11.19 (2020). PMID: 32880462, pp. 8060–8066. DOI: [10.1021/acs.jpclett.0c02129](https://doi.org/10.1021/acs.jpclett.0c02129).

- [11] Caroline Desgranges and Jerome Delhommelle. “Controlling Polymorphism during the Crystallization of an Atomic Fluid”. In: *Phys. Rev. Lett.* 98 (23 June 2007), p. 235502. DOI: [10.1103/PhysRevLett.98.235502](https://doi.org/10.1103/PhysRevLett.98.235502).
- [12] Caroline Desgranges and Jerome Delhommelle. “Insights into the Molecular Mechanism Underlying Polymorph Selection”. In: *Journal of the American Chemical Society* 128.47 (2006). PMID: 17117858, pp. 15104–15105. DOI: [10.1021/ja0670310](https://doi.org/10.1021/ja0670310).
- [13] Freddy Bouchet, Joran Rolland, and Jeroen Wouters. “Rare Event Sampling Methods”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.8 (Aug. 2019). 080402. ISSN: 1054-1500. DOI: [10.1063/1.5120509](https://doi.org/10.1063/1.5120509).
- [14] Pablo M. Piaggi and Michele Parrinello. “Calculation of phase diagrams in the multithermal-multibaric ensemble”. In: *The Journal of Chemical Physics* 150.24 (June 2019). 244119. ISSN: 0021-9606. DOI: [10.1063/1.5102104](https://doi.org/10.1063/1.5102104).
- [15] G.M. Torrie and J.P. Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *Journal of Computational Physics* 23.2 (1977), pp. 187–199. ISSN: 0021-9991. DOI: [https://doi.org/10.1016/0021-9991\(77\)90121-8](https://doi.org/10.1016/0021-9991(77)90121-8).
- [16] Dominic C Rose, Jamie F Mair, and Juan P Garrahan. “A reinforcement learning approach to rare trajectory sampling”. In: *New Journal of Physics* 23.1 (Jan. 2021), p. 013013. DOI: [10.1088/1367-2630/abd7bd](https://doi.org/10.1088/1367-2630/abd7bd).
- [17] Xian-Ming Bai and Mo Li. “Differences between solid superheating and liquid supercooling”. In: *The Journal of Chemical Physics* 123.15 (Oct. 2005). 151102. ISSN: 0021-9606. DOI: [10.1063/1.2085147](https://doi.org/10.1063/1.2085147).
- [18] Xian-Ming Bai and Mo Li. “Calculation of solid-liquid interfacial free energy: A classical nucleation theory based approach”. In: *The Journal of Chemical Physics* 124.12 (Mar. 2006). 124707. ISSN: 0021-9606. DOI: [10.1063/1.2184315](https://doi.org/10.1063/1.2184315).
- [19] J. Sethna. *Statistical Mechanics: Entropy, Order Parameters and Complexity*. Oxford Master Series in Physics. OUP Oxford, 2006. ISBN: 9780191566219.
- [20] Jim De Yoreo. “A Perspective on Multistep Pathways of Nucleation”. In: *Crystallization via Nonclassical Pathways Volume 1: Nucleation, Assembly, Observation & Application*. Chap. 1, pp. 1–17. DOI: [10.1021/bk-2020-1358.ch001](https://doi.org/10.1021/bk-2020-1358.ch001).
- [21] S. Karthika, T. K. Radhakrishnan, and P. Kalaichelvi. “A Review of Classical and Nonclassical Nucleation Theories”. In: *Crystal Growth & Design* 16.11 (2016), pp. 6663–6681. DOI: [10.1021/acs.cgd.6b00794](https://doi.org/10.1021/acs.cgd.6b00794).
- [22] R. A. Van Santen. “The Ostwald step rule”. In: *The Journal of Physical Chemistry* 88.24 (1984), pp. 5768–5769. DOI: [10.1021/j150668a002](https://doi.org/10.1021/j150668a002).

- [23] Nguyen T. K. Thanh, N. Maclean, and S. Mahiddine. “Mechanisms of Nucleation and Growth of Nanoparticles in Solution”. In: *Chemical Reviews* 114.15 (2014). PMID: 25003956, pp. 7610–7630. doi: 10.1021/cr400544s.
- [24] Gabriele C. Sosso, Ji Chen, Stephen J. Cox, Martin Fitzner, Philipp Pedevilla, Andrea Zen, and Angelos Michaelides. “Crystal Nucleation in Liquids: Open Questions and Future Challenges in Molecular Dynamics Simulations”. In: *Chemical Reviews* 116.12 (2016). PMID: 27228560, pp. 7078–7116. doi: 10.1021/acs.chemrev.5b00744.
- [25] Eric R. Beyerle, Ziyue Zou, and Pratyush Tiwary. *Recent advances in describing and driving crystal nucleation using machine learning and artificial intelligence*. 2023.
- [26] Frank H. Stillinger and Thomas A. Weber. “Hidden structure in liquids”. In: *Phys. Rev. A* 25 (2 Feb. 1982), pp. 978–989. doi: 10.1103/PhysRevA.25.978.
- [27] Sung-Hoon Lee, Jongseob Kim, Sae-Jin Kim, Sungjin Kim, and Gyeong-Su Park. “Hidden Structural Order in Orthorhombic Ta₂O₅”. In: *Phys. Rev. Lett.* 110 (23 June 2013), p. 235502. doi: 10.1103/PhysRevLett.110.235502.
- [28] Hua Tong and Hajime Tanaka. “Revealing Hidden Structural Order Controlling Both Fast and Slow Glassy Dynamics in Supercooled Liquids”. In: *Phys. Rev. X* 8 (1 Mar. 2018), p. 011041. doi: 10.1103/PhysRevX.8.011041.
- [29] R. L. McGreevy and L. Pusztai. “Reverse Monte Carlo Simulation: A New Technique for the Determination of Disordered Structures”. In: *Molecular Simulation* 1.6 (1988), pp. 359–367. doi: 10.1080/08927028808080958.
- [30] Fang Li and Jeffrey S. Lannin. “Radial distribution function of amorphous carbon”. In: *Phys. Rev. Lett.* 65 (15 Oct. 1990), pp. 1905–1908. doi: 10.1103/PhysRevLett.65.1905.
- [31] N. La čevi č, F. W. Starr, T. B. Schröder, V. N. Novikov, and S. C. Glotzer. “Growing correlation length on cooling below the onset of caging in a simulated glass-forming liquid”. In: *Phys. Rev. E* 66 (3 Sept. 2002), p. 030101. doi: 10.1103/PhysRevE.66.030101.
- [32] Daniel Faken and Hannes Jónsson. “Systematic analysis of local atomic structure combined with 3D computer graphics”. In: *Computational Materials Science* 2.2 (1994), pp. 279–286. ISSN: 0927-0256. doi: [https://doi.org/10.1016/0927-0256\(94\)90109-0](https://doi.org/10.1016/0927-0256(94)90109-0).
- [33] Cynthia L. Kelchner, S. J. Plimpton, and J. C. Hamilton. “Dislocation nucleation and defect structure during surface indentation”. In: *Phys. Rev. B* 58 (17 Nov. 1998), pp. 11085–11088. doi: 10.1103/PhysRevB.58.11085.

- [34] P.J.P. Espitia, C.G. Otoni, and N.F.F. Soares. “Chapter 34 - Zinc Oxide Nanoparticles for Food Packaging Applications”. In: *Antimicrobial Food Packaging*. Ed. by Jorge Barros-Velázquez. San Diego: Academic Press, 2016, pp. 425–431. ISBN: 978-0-12-800723-5. DOI: <https://doi.org/10.1016/B978-0-12-800723-5.00034-6>.
- [35] Jacek Goniakowski, Sarath Menon, Gaétan Laurens, and Julien Lam. “Nonclassical Nucleation of Zinc Oxide from a Physically Motivated Machine-Learning Approach”. In: *The Journal of Physical Chemistry C* 126.40 (2022), pp. 17456–17469. DOI: 10.1021/acs.jpcc.2c06341.
- [36] Xin-Wei Wang, Xiao-Wei Sun, T. Song, Jun-Hong Tian, and Zi-Jiang Liu. “Structural transition, mechanical properties and electronic structure of the ZnO under high pressure via first-principles investigations”. In: *Applied Physics A* 128 (July 2022). DOI: 10.1007/s00339-022-05845-x.
- [37] Wolfgang Lechner and Christoph Dellago. “Accurate determination of crystal structures based on averaged local bond order parameters”. In: *The Journal of Chemical Physics* 129.11 (Sept. 2008). 114707. ISSN: 0021-9606. DOI: 10.1063/1.2977970.
- [38] Philipp Geiger and Christoph Dellago. “Neural networks for local structure detection in polymorphic systems”. In: *The Journal of Chemical Physics* 139.16 (Oct. 2013). 164105. ISSN: 0021-9606. DOI: 10.1063/1.4825111.
- [39] Pablo M. Piaggi and Michele Parrinello. “Entropy based fingerprint for local crystalline order”. In: *The Journal of Chemical Physics* 147.11 (Sept. 2017). 114112. ISSN: 0021-9606. DOI: 10.1063/1.4998408.
- [40] Miguel A. Caro, Anja Aarva, Volker L. Deringer, Gábor Csányi, and Tomi Laurila. “Reactivity of Amorphous Carbon Surfaces: Rationalizing the Role of Structural Motifs in Functionalization Using Machine Learning”. In: *Chemistry of Materials* 30.21 (2018). PMID: 30487663, pp. 7446–7455. DOI: 10.1021/acs.chemmater.8b03353.
- [41] Emanuele Boattini, Michel Ram, Frank Smallenburg, and Laura Filion. “Neural-network-based order parameters for classification of binary hard-sphere crystal structures”. In: *Molecular Physics* 116.21-22 (2018), pp. 3066–3075. DOI: 10.1080/00268976.2018.1483537.
- [42] Emanuele Boattini, Marjolein Dijkstra, and Laura Filion. “Unsupervised learning for local structure detection in colloidal systems”. In: *The Journal of Chemical Physics* 151.15 (Oct. 2019). 154901. ISSN: 0021-9606. DOI: 10.1063/1.5118867.
- [43] Emanuele Boattini, Susana Marín-Aguilar, Saheli Mitra, Giuseppe Foffi, Frank Smallenburg, and Laura Filion. “Autonomously revealing hidden local structures in supercooled liquids”. In: *Nature communications* 11.1 (2020), p. 5479.

- [44] Ryo Tamura, Momo Matsuda, Jianbo Lin, Yasunori Futamura, Tetsuya Sakurai, and Tsuyoshi Miyazaki. “Structural analysis based on unsupervised learning: Search for a characteristic low-dimensional space by local structures in atomistic simulations”. In: *Phys. Rev. B* 105 (7 Feb. 2022), p. 075107. DOI: 10.1103/PhysRevB.105.075107.
- [45] Wesley F. Reinhart. “Unsupervised learning of atomic environments from simple features”. In: *Computational Materials Science* 196 (2021), p. 110511. ISSN: 0927-0256. DOI: <https://doi.org/10.1016/j.commatsci.2021.110511>.
- [46] Gareth A. Tribello, Michele Ceriotti, and Michele Parrinello. “Using sketch-map coordinates to analyze and bias molecular dynamics simulations”. In: *Proceedings of the National Academy of Sciences* 109.14 (2012), pp. 5196–5201. DOI: 10.1073/pnas.1201152109.
- [47] Michele Ceriotti, Gareth A. Tribello, and Michele Parrinello. “Simplifying the representation of complex free-energy landscapes using sketch-map”. In: *Proceedings of the National Academy of Sciences* 108.32 (2011), pp. 13023–13028. DOI: 10.1073/pnas.1108486108.
- [48] Paul J. Steinhardt, David R. Nelson, and Marco Ronchetti. “Bond-orientational order in liquids and glasses”. In: *Phys. Rev. B* 28 (2 July 1983), pp. 784–805. DOI: 10.1103/PhysRevB.28.784.
- [49] Ran Ni and Marjolein Dijkstra. “Crystal nucleation of colloidal hard dumbbells”. In: *The Journal of Chemical Physics* 134.3 (Jan. 2011). 034501. ISSN: 0021-9606. DOI: 10.1063/1.3528222.
- [50] W -S Xu, Z -Y Sun, and L -J An. “Dense packing in the monodisperse hard-sphere system: A numerical study”. In: *The European Physical Journal E* 31 (2010), pp. 377–382.
- [51] L.-C. Valdes, F. Affouard, M. Descamps, and J. Habasaki. “Mixing effects in glass-forming Lennard-Jones mixtures”. In: *The Journal of Chemical Physics* 130.15 (Apr. 2009). 154505. ISSN: 0021-9606. DOI: 10.1063/1.3106759.
- [52] Takeshi Kawasaki and Hajime Tanaka. “Structural origin of dynamic heterogeneity in three-dimensional colloidal glass formers and its link to crystal nucleation”. In: *Journal of Physics: Condensed Matter* 22.23 (May 2010), p. 232102. DOI: 10.1088/0953-8984/22/23/232102.
- [53] Hui Wang, Harvey Gould, and W. Klein. “Homogeneous and heterogeneous nucleation of Lennard-Jones liquids”. In: *Phys. Rev. E* 76 (3 Sept. 2007), p. 031604. DOI: 10.1103/PhysRevE.76.031604.
- [54] Aaron S. Keys and Sharon C. Glotzer. “How do Quasicrystals Grow?” In: *Phys. Rev. Lett.* 99 (23 Dec. 2007), p. 235503. DOI: 10.1103/PhysRevLett.99.235503.

- [55] Christopher R. Iacobella, Aaron S. Keys, Mark A. Horsch, and Sharon C. Glotzer. “Icosahedral packing of polymer-tethered nanospheres and stabilization of the gyroid phase”. In: *Phys. Rev. E* 75 (4 Apr. 2007), p. 040801. DOI: [10.1103/PhysRevE.75.040801](https://doi.org/10.1103/PhysRevE.75.040801).
- [56] Yanting Wang, S. Teitel, and Christoph Dellago. “Melting of icosahedral gold nanoclusters from molecular dynamics simulations”. In: *The Journal of Chemical Physics* 122.21 (June 2005). 214722. ISSN: 0021-9606. DOI: [10.1063/1.1917756](https://doi.org/10.1063/1.1917756).
- [57] Charusita Chakravarty, Pablo G. Debenedetti, and Frank H. Stillinger. “Linde-mann measures for the solid-liquid phase transition”. In: *The Journal of Chemical Physics* 126.20 (May 2007). 204508. ISSN: 0021-9606. DOI: [10.1063/1.2737054](https://doi.org/10.1063/1.2737054).
- [58] F. Calvo and D. J. Wales. “Stepwise melting of a model glass former under confinement”. In: *The Journal of Chemical Physics* 131.13 (Oct. 2009). 134504. ISSN: 0021-9606. DOI: [10.1063/1.3239468](https://doi.org/10.1063/1.3239468).
- [59] Jessica Hernández-Guzmán and Eric R. Weeks. “The equilibrium intrinsic crystal–liquid interface of colloids”. In: *Proceedings of the National Academy of Sciences* 106.36 (2009), pp. 15198–15202. DOI: [10.1073/pnas.0904682106](https://doi.org/10.1073/pnas.0904682106).
- [60] Walter Mickel, Sebastian C. Kapfer, Gerd E. Schröder-Turk, and Klaus Mecke. “Shortcomings of the bond orientational order parameters for the analysis of disordered particulate matter”. In: *The Journal of Chemical Physics* 138.4 (Jan. 2013). 044501. ISSN: 0021-9606. DOI: [10.1063/1.4774084](https://doi.org/10.1063/1.4774084).
- [61] Sebastian C. Kapfer, Walter Mickel, Klaus Mecke, and Gerd E. Schröder-Turk. “Jammed spheres: Minkowski tensors reveal onset of local crystallinity”. In: *Phys. Rev. E* 85 (3 Mar. 2012), p. 030301. DOI: [10.1103/PhysRevE.85.030301](https://doi.org/10.1103/PhysRevE.85.030301).
- [62] Anuraag R. Kansal, Salvatore Torquato, and Frank H. Stillinger. “Diversity of order and densities in jammed hard-particle packings”. In: *Phys. Rev. E* 66 (4 Oct. 2002), p. 041109. DOI: [10.1103/PhysRevE.66.041109](https://doi.org/10.1103/PhysRevE.66.041109).
- [63] Pieter Rein ten Wolde, Maria J. Ruiz-Montero, and Daan Frenkel. “Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling”. In: *The Journal of Chemical Physics* 104.24 (June 1996), pp. 9932–9947. ISSN: 0021-9606. DOI: [10.1063/1.471721](https://doi.org/10.1063/1.471721).
- [64] Troadec, J. P., Gervois, A., and Oger, L. “Statistics of Voronoi cells of slightly perturbed face-centered cubic and hexagonal close-packed lattices”. In: *Europhys. Lett.* 42.2 (1998), pp. 167–172. DOI: [10.1209/epl/i1998-00224-x](https://doi.org/10.1209/epl/i1998-00224-x).
- [65] Franz Aurenhammer, Rolf Klein, and Der-Tsai Lee. *Voronoi Diagrams and Delaunay Triangulations*. WORLD SCIENTIFIC, 2013. DOI: [10.1142/8685](https://doi.org/10.1142/8685).

- [66] Dimitrios Gunopulos. “Clustering Overview and Applications”. In: *Encyclopedia of Database Systems*. Ed. by LING LIU and M. TAMER ÖZSU. Boston, MA: Springer US, 2009, pp. 383–387. ISBN: 978-0-387-39940-9. DOI: 10.1007/978-0-387-39940-9_602.
- [67] Michele Ceriotti. “Unsupervised machine learning in atomistic simulations, between predictions and understanding”. In: *The Journal of Chemical Physics* 150.15 (Apr. 2019). 150901. ISSN: 0021-9606. DOI: 10.1063/1.5091842.
- [68] Sébastien Becker, Emilie Devijver, Rémi Molinier, and Noël Jakse. “Unsupervised topological learning for identification of atomic structures”. In: *Phys. Rev. E* 105 (4 Apr. 2022), p. 045304. DOI: 10.1103/PhysRevE.105.045304.
- [69] Piero Gasparotto, Robert Horst Meißner, and Michele Ceriotti. “Recognizing Local and Global Structural Motifs at the Atomic Scale”. In: *Journal of Chemical Theory and Computation* 14.2 (2018). PMID: 29298385, pp. 486–498. DOI: 10.1021/acs.jctc.7b00993.
- [70] John A. Richards. “Supervised Classification Techniques”. In: *Remote Sensing Digital Image Analysis: An Introduction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 247–318. ISBN: 978-3-642-30062-2. DOI: 10.1007/978-3-642-30062-2_8.
- [71] Gideon Schwarz. “Estimating the Dimension of a Model”. In: *The Annals of Statistics* 6.2 (1978), pp. 461–464. DOI: 10.1214/aos/1176344136.
- [72] C. Biernacki, G. Celeux, and G. Govaert. “Assessing a mixture model for clustering with the integrated completed likelihood”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.7 (2000), pp. 719–725. DOI: 10.1109/34.865189.
- [73] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum Likelihood from Incomplete Data Via the EM Algorithm”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977), pp. 1–22. DOI: <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.
- [74] A. P. Thompson et al. “LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales”. In: *Comp. Phys. Comm.* 271 (2022), p. 108171. DOI: 10.1016/j.cpc.2021.108171.
- [75] Wataru Shinoda, Motoyuki Shiga, and Masuhiro Mikami. “Rapid estimation of elastic constants by molecular dynamics simulation under constant stress”. In: *Phys. Rev. B* 69 (13 Apr. 2004), p. 134103. DOI: 10.1103/PhysRevB.69.134103.
- [76] T. Schneider and E. Stoll. “Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions”. In: *Phys. Rev. B* 17 (3 Feb. 1978), pp. 1302–1322. DOI: 10.1103/PhysRevB.17.1302.

- [77] Sarath Menon, Grisell Díaz Leines, and Jutta Rogal. “pyscal: A python module for structural analysis of atomic environments”. In: *Journal of Open Source Software* 4.43 (2019), p. 1824. doi: 10.21105/joss.01824.
- [78] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [79] Jorge R. Espinosa, Carlos Vega, Chantal Valeriani, and Eduardo Sanz. “The crystal-fluid interfacial free energy and nucleation rate of NaCl from different simulation methods”. In: *The Journal of Chemical Physics* 142.19 (May 2015). 194709. ISSN: 0021-9606. doi: 10.1063/1.4921185.
- [80] Brandon C. Knott, Valeria Molinero, Michael F. Doherty, and Baron Peters. “Homogeneous Nucleation of Methane Hydrates: Unrealistic under Realistic Conditions”. In: *Journal of the American Chemical Society* 134.48 (2012). PMID: 23148735, pp. 19544–19547. doi: 10.1021/ja309117d.
- [81] Rodolfo G. Pereyra, Igal Szleifer, and Marcelo A. Carignano. “Temperature dependence of ice critical nucleus size”. In: *The Journal of Chemical Physics* 135.3 (July 2011), p. 034508. ISSN: 0021-9606. doi: 10.1063/1.3613672.
- [82] E. Sanz, C. Vega, J. R. Espinosa, R. Caballero-Bernal, J. L. F. Abascal, and C. Valeriani. “Homogeneous Ice Nucleation at Moderate Supercooling from Molecular Simulation”. In: *Journal of the American Chemical Society* 135.40 (2013). PMID: 24010583, pp. 15008–15017. doi: 10.1021/ja4028814.
- [83] J. R. Espinosa, E. Sanz, C. Valeriani, and C. Vega. “Homogeneous ice nucleation evaluated for several water models”. In: *The Journal of Chemical Physics* 141.18 (Oct. 2014), p. 18C529. ISSN: 0021-9606. doi: 10.1063/1.4897524.
- [84] Nils E. R. Zimmermann, Bart Vorselaars, David Quigley, and Baron Peters. “Nucleation of NaCl from Aqueous Solution: Critical Sizes, Ion-Attachment Kinetics, and Rates”. In: *Journal of the American Chemical Society* 137.41 (2015). PMID: 26371630, pp. 13352–13361. doi: 10.1021/jacs.5b08098.
- [85] Jorge R. Espinosa, Carlos Vega, Chantal Valeriani, and Eduardo Sanz. “Seeding approach to crystal nucleation”. In: *The Journal of Chemical Physics* 144.3 (Jan. 2016). 034501. ISSN: 0021-9606. doi: 10.1063/1.4939641.
- [86] Alexander Stukowski. “Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool”. In: *MODELLING AND SIMULATION IN MATERIALS SCIENCE AND ENGINEERING* 18.1 (JAN 2010). ISSN: 0965-0393. doi: {10.1088/0965-0393/18/1/015012}.
- [87] Richard L Burden, J Douglas Faires, and Annette M Burden. *Numerical analysis*. Cengage learning, 2015.

Appendix

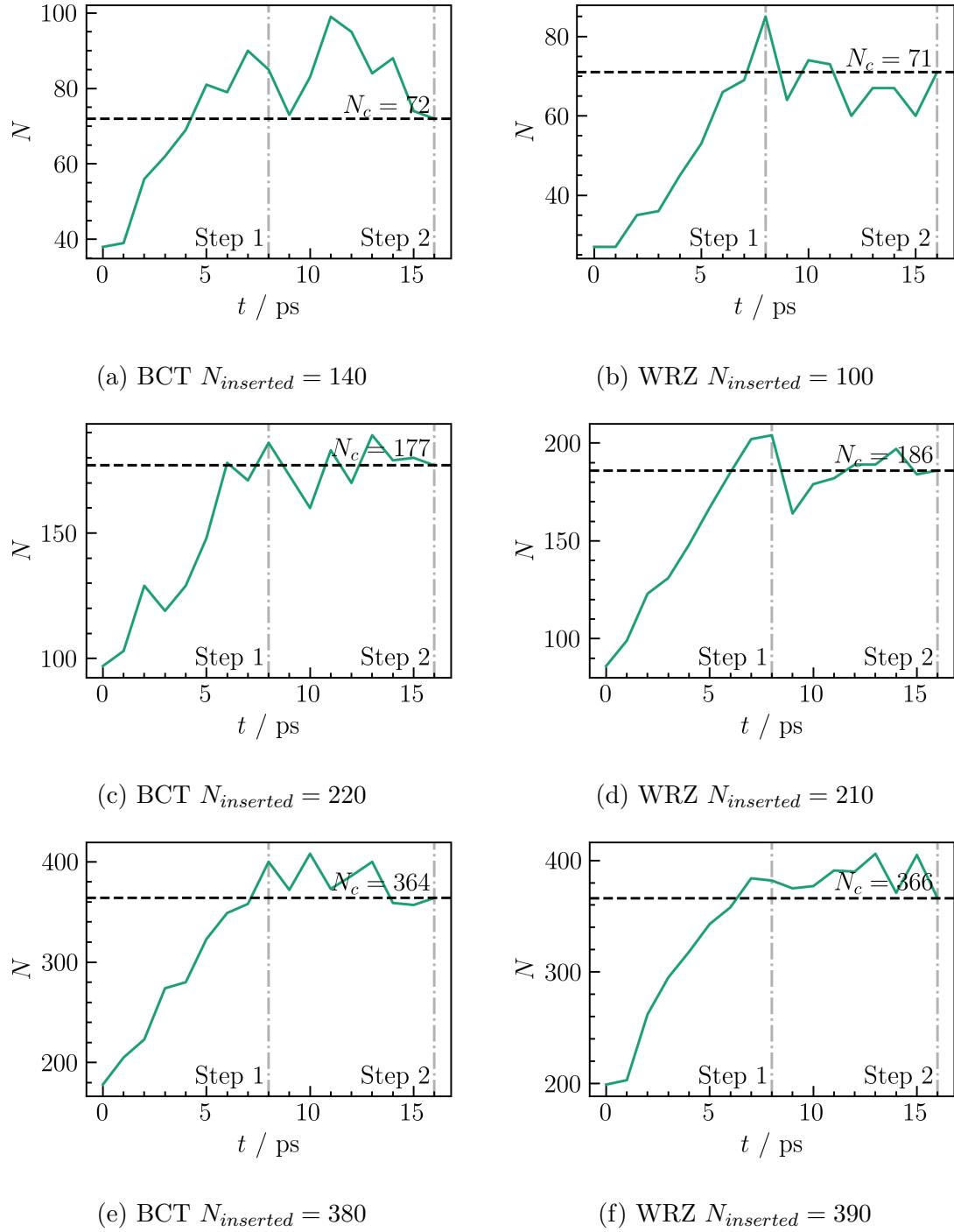
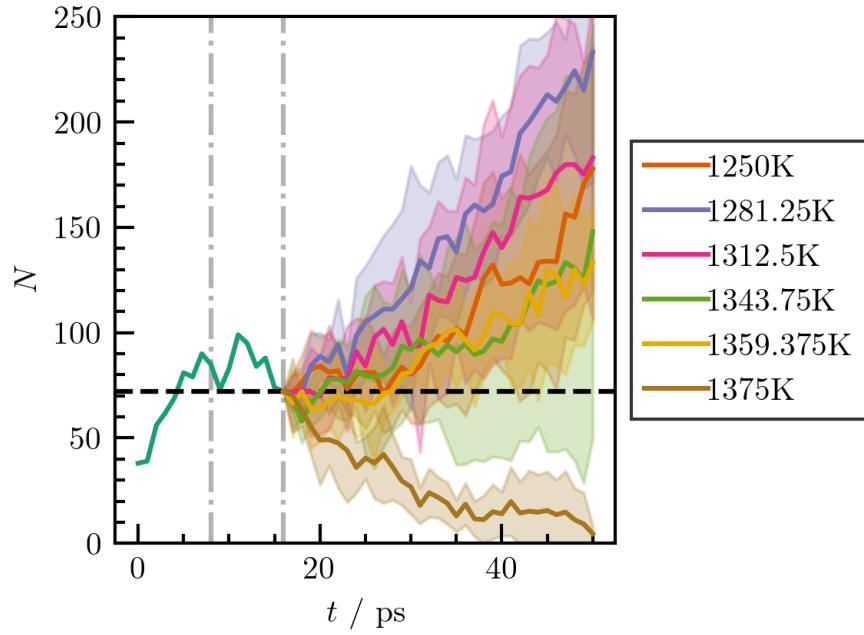
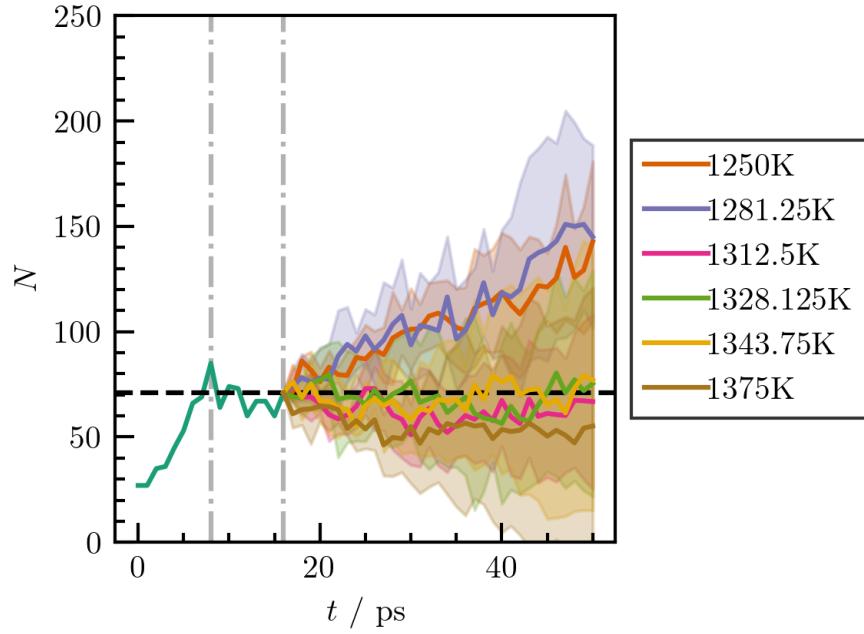


Figure A.1: Crystalline cluster size as a function of time during the relaxation step for different sizes of inserted crystal seeds.

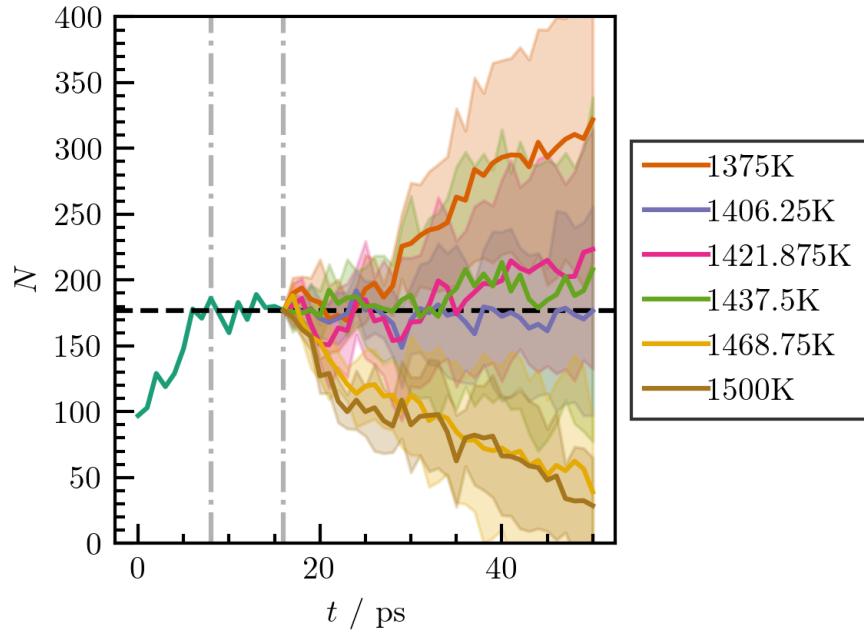


(a) BCT $N_c = 72$

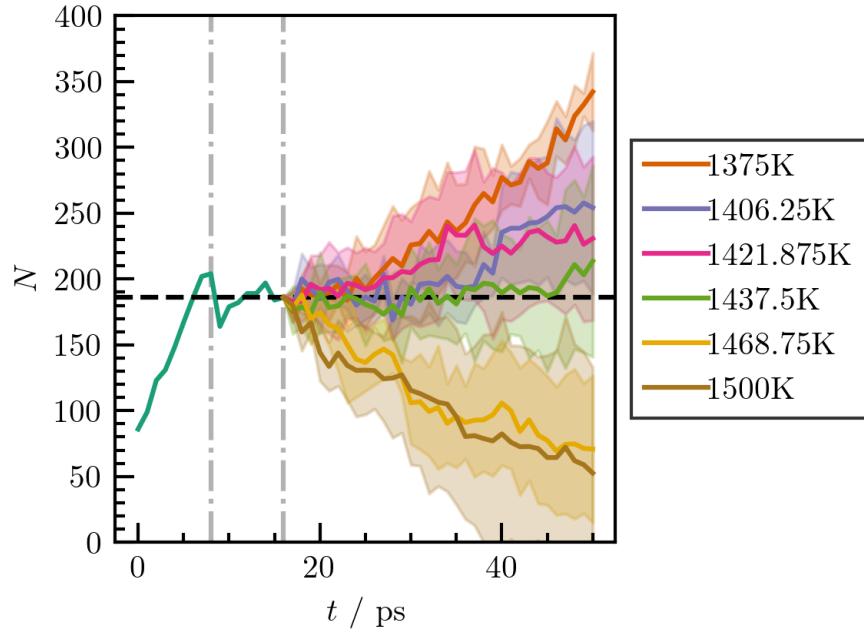


(b) WRZ $N_c = 71$

Figure A.2: Crystalline cluster size as a function of time during the relaxation and growth/melting simulations at different temperatures. Fig (a) shows the results for a BCT critical cluster of 72 atoms, while Fig (b) shows the results for a WRZ critical cluster of 71 atoms.

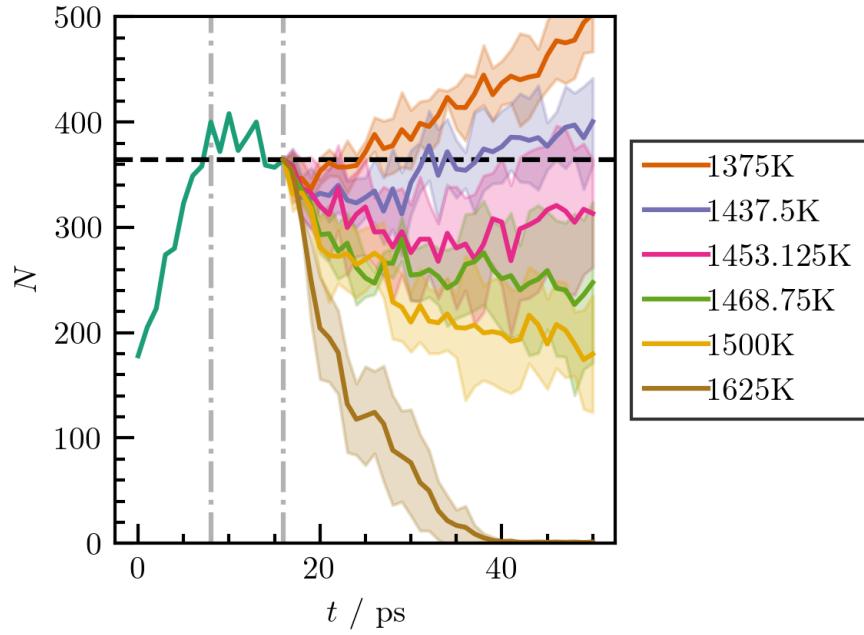


(a) BCT $N_c = 177$

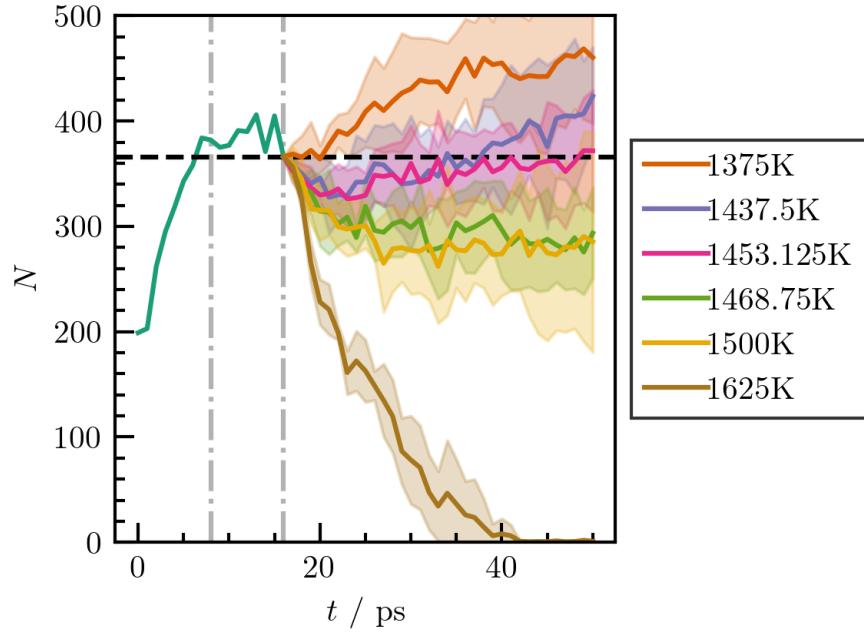


(b) WRZ $N_c = 186$

Figure A.3: Crystalline cluster size as a function of time during the relaxation and growth/melting simulations at different temperatures. Fig (a) shows the results for a BCT critical cluster of 177 atoms, while Fig (b) shows the results for a WRZ critical cluster of 186 atoms.



(a) BCT, $N_c = 364$



(b) WRZ, $N_c = 366$

Figure A.4: Crystalline cluster size as a function of time during the relaxation and growth/melting simulations at different temperatures. Fig (a) shows the results for a BCT critical cluster of 364 atoms, while Fig (b) shows the results for a WRZ critical cluster of 366 atoms.