



HAL
open science

Non-classical nucleation of zinc oxide from a physically-motivated machine-learning approach

Jacek Goniakowski, Sarath Menon, Gaétan Laurens, Julien Lam

► To cite this version:

Jacek Goniakowski, Sarath Menon, Gaétan Laurens, Julien Lam. Non-classical nucleation of zinc oxide from a physically-motivated machine-learning approach. *Journal of Physical Chemistry C*, 2022, 126 (40), pp.17456-17469. <10.1021/acs.jpcc.2c06341>. <hal-03805114>

HAL Id: hal-03805114

<https://hal.science/hal-03805114v1>

Submitted on 7 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Non-classical nucleation of zinc oxide from a physically-motivated machine-learning approach.

Jacek Goniakowski,[†] Sarath Menon,[‡] Gaétan Laurens,[¶] and Julien Lam^{*,§}

[†]*CNRS, Sorbonne Université, Institut des NanoSciences de Paris, UMR 7588, 4 Place Jussieu, F-75005 Paris, France*

[‡]*Computational Materials Design, Max-Planck-Institut für Eisenforschung GmbH, 40237 Düsseldorf, Germany*

[¶]*Institut Lumière Matière, UMR5306 Université Lyon 1-CNRS, Université de Lyon, 69622 Villeurbanne Cedex, France*

[§]*Centre d'élaboration des Matériaux et d'Etudes Structurales, CNRS (UPR 8011), 29 rue Jeanne Marvig, 31055 Toulouse Cedex 4, France*

E-mail: julien.lam@cnr.fr

Abstract

Observing non-classical nucleation pathways remains challenging in simulations of complex materials with technological interests. This is because it requires very accurate force fields that can capture the whole complexity of their underlying interatomic interactions and an advanced structural analysis able to discriminate between competing crystalline phases. Here, we first report the construction and a particularly thorough validation of a machine-learning force field for zinc oxide interactions using the Physical LassoLars Interaction Potentials approach which allows us to be predictive even for high temperature dynamical systems such as ZnO melt. Then, we carried out several types of crystallization simulations and followed the formation of ZnO crystal with atomistic precision. Our results, which were analyzed using a data-driven approach based on bond order parameters, demonstrate the presence of both prenucleation clusters and two-step nucleation scenarios, thus retrieving seminal predictions of non-classical nucleation pathways made on much simpler models. Dedicated calculations of high temperature ZnO free energy within a newly developed automated nonequi-

librium thermodynamic integration method revealed the existence of a thermodynamic bias for the predicted non-classical nucleation scenarios.

Introduction

While crystals in material science are ubiquitous, the mechanisms of their formation which spans from nucleation to crystal growth remain one of the most intriguing processes in nature. Better understanding crystallization would allow for a rational control of material engineering and possibly for the development of novel functional materials and technological applications. From the fundamental point of view, numerous works have been dedicated to elucidating the emergence of the nucleation core¹⁻⁵ and its role in controlling the final crystal structure.⁶⁻¹² For instance, it is now possible to observe the crystal birth with electron microscopy,^{6,7,13-15} and colloidal science has also provided numerous experimental results on nucleation.¹⁶⁻¹⁹ Yet, numerical simulations remain the principal instrument to investigate crystallization at the atomistic level.²⁰ In this context, most of these works require large scale simulations in order to observe the phase transition

and have therefore only focused on materials for which the interactions are very simple, including hard-spheres,^{3,21–25} Lennard-Jones,^{1,5,9,11,26} water,^{23,27–34} as well as metallic potentials, like embedded-atom model (EAM).^{8,10,35}

Prompted by this large body of fundamental achievements, it becomes timely to reach the same level of understanding for crystallization in more complex materials in order to target more diverse technological applications. So far, the need for large scale simulations have prevented from using quantum accurate modeling including density functional theory (DFT), and the research field dedicated to constructing novel interaction potentials to bridge this computational gap has been ever expanding.^{36–43} In particular, the past decade has seen the emergence of innovative types of interaction potentials based on machine-learning algorithms.⁴⁴ Various approaches have been proposed such as Artificial Neural Networks,⁴⁴ Gaussian approximation potentials,⁴⁵ Linearized potentials,^{46–51} Spectral Neighbor Analysis Potential,^{52,53} Symmetric Gradient Domain Machine learning,^{54,55} and Moment Tensor Potentials.^{56,57} The success of those machine-learning interaction potentials (MLIPs) is seen through the large variety of studied materials, namely pure metals,^{46,58–61} organic molecules,^{62–65} water,^{66–70} amorphous materials,^{71–76} and hybrid perovskites.⁷⁷

The oxide materials has been much less investigated using MLIPs.^{78–85} With those materials, the complexity lies in the emergence of long-range electrostatic effects difficult to capture with traditional machine-learning approaches.^{86,87} In addition, oxide materials often exhibit a rich structural landscape composed of numerous polymorphs.^{88–91} Being able to model those polymorphs with a unique set of atomic interaction remains challenging even for advanced machine-learning methods.

In this article we focus on modeling ZnO crystallization from bulk ZnO melt, based on a newly developed approach named Physical LassoLars Interaction Potential (PLIP), which employs a physically-motivated mathematical formulation and relies on a constrained linear regression scheme for parameter adjustment. Beyond the exiting state-of-the-art interaction

models, such as ReactFF, Tersoff potentials, or artificial neural network potential,^{79,81} we explicitly show that our PLIP satisfactorily accounts not only for the six most stable bulk ZnO polymorphs and their low index surfaces, for the vibrational characteristics and the free energy behaviour of the most stable ZnO crystalline phases, but also for the structural characteristics of amorphous ZnO and of its high temperature melt.

Having demonstrated the ability of the PLIP to model ZnO in most various situations and, in particular, in dynamic high temperature limit, we used it in molecular dynamics (MD) simulations of ZnO crystallization by either freezing ZnO melt or by letting it occur spontaneously in an undercooled ZnO liquid. With help of a data-driven approach of phase recognition, we were able to access the nucleation of zinc oxide with an atomic precision and to reveal a dominant role of a non-classical nucleation scenarios, involving the less common BCT ZnO polymorph. Moreover, with dedicated calculations of ZnO free energy within a newly developed automated nonequilibrium thermodynamic integration method we showed the existence of a stability reversal between wurtzite and BCT polymorphs at high temperature and thus of a clear thermodynamic bias for such the predicted nucleation pathways.

From a material point of view, since finite size zinc oxide crystals are found in numerous applications, *e.g.* photocatalysis, piezoelectricity, drug delivery, antibacterial, and gas detection,⁹² our results provide tools and lay grounds for future studies on more complex, finite size and surface- and/or interface phenomena which are involved in formation of ZnO nanostructures.

Methods

Machine-learning interaction potential using the PLIP approach

As in most MLIP approaches, the total potential energy E is decomposed as the sum of independent atomic energies E_i : $E = \sum_{i=0}^N E_i$,

where N corresponds to the number of atoms of the considered configuration. Then, we employ a linear model which consists in approximating E_i as a linear combination of descriptors X_n^i : $E_i = \sum_n \omega_n X_n^i$ where ω_n are the linear coefficients that must be determined. More specifically, the PLIP model is made of three types of descriptors which explicitly follow a many-body order expansion:

$$[2B]_n^i = \sum_j f_n(R_{ij}) \times f_c(R_{ij}), \quad (1)$$

$$[3B]_{n,l}^i = \sum_j \sum_k f_n(R_{ij}) f_c(R_{ij}) f_n(R_{ik}) f_c(R_{ik}) \cos^l(\theta_{ijk}), \quad (2)$$

$$[NB]_{n,m}^i = \left(\sum_j f_n(R_{ij}) \times f_c(R_{ij}) \times f_s(R_{ij}) \right)^m, \quad (3)$$

where R_{ij} is the distance between atoms i and j , θ_{ijk} is the angle centered around the atom i , and l and m are two positive integers. $f_c(R_{ij})$ is the same cut-off function as in the seminal work of Behler and Parinello,⁴⁴ namely $f_c(R_{ij}) = \frac{1}{2} (1 + \cos(\pi(R_{ij}/R_{cut})))$, where the cut-off distance R_{cut} is set at 6 Å. The shift function is chosen in the form: $f_s(u) = 6u^5 - 15u^4 + 10u^3$, where $u = (R_{ij} - r_1)/(r_2 - r_1)$, and r_1 (resp. r_2) is defined as 95 % (resp. 105 %) of a short distance equal to 1.1 Å. Regarding the basis functions $f_n(R_{ij})$, we previously demonstrated the advantage of coupling different classical functions at the same time.⁵¹ Yet, in this particular study on zinc oxide where numerical efficiency is crucial, we work only with Gaussian functions for which the width and the central positions are respectively listed as follows: [0.5, 1.0, 1.5] and [0.5, 1.0, 1.5 ... 5.5, 6.0]. In addition, we vary the integers l and m respectively from 0 to 5, and from 4 to 7. Altogether, when taking into account the binary nature of the zinc oxide system, our model is made of 1981 available descriptors.

In order to match the first-principle database, the model is fitted using the LassoLars approach which allows for a well-informed selection of the most preponderant descriptors and a reduction in the complexity of the obtained

potential.^{51,93} In this study, each of the PLIP models selects approximately 150 descriptors among an order of magnitude more that are available.

First principle calculations and training database

Reference GGA-DFT calculations are performed with VASP,^{94,95} using PW91 exchange-correlation functional,⁹⁶ and the projector augmented wave method.⁹⁷ Standard zinc and soft oxygen (energy cutoff of 270 eV) pseudopotentials provided by VASP are used in all calculations, enabling an efficient structural relaxation of systems composed of several hundreds of atoms. GGA results obtained with the soft and the full (energy cutoff of 400 eV) oxygen pseudopotentials show satisfactory agreement (differences of bulk parameters a and c smaller than 0.01 Å, and cohesion energy differences ΔE_{coh} below 0.01 eV/ZnO). The present results on six ZnO polymorphs [wurtzite (WRZ), zinc blend (ZBL), body centered tetragonal (BCT), sodalite (SOD), h -BN (HBN), and cubane (CUB) crystallographic structures] coincide very well with the existing hybrid HSE03 reference⁹⁰ [See Tab. SI1], in particular concerning the relative stability of the considered polymorphs. Such agreement between the two types of modeling indicates that the chosen GGA-DFT approach leads to sufficiently accurate results while enabling for larger scale calculations.

Since the training database includes both ordered and disordered structures, in DFT calculations we systematically use relatively large supercells (16-19 Å large cuboids), containing 320-480 atoms, and sample the Brillouin zone with a single Γ point. In ordered structures, the atomic coordinates of all ions are relaxed until residual forces dropped below 0.01 eV/Å and, in the bulk structures, all components of the stress tensor are smaller than 0.01 eV/Å³. The disordered structures are collected along short high-temperature MD runs with no further relaxation as to produce non-vanishing forces on all ions.

In order to include information on the characteristics of low coordinated ions, for each of

the six ZnO polymorphs we also consider at least one low index non-polar surface, whereas three surface orientations are considered for the most stable (WRZ) polymorph. In this latter case, beyond the most stable (10-10) and (11-20) non-polar orientations, we also include the polar (0001)/(000-1) surfaces. As to impede the emergence of a macroscopic dipole moment, we use an asymmetric slab with the conventional (2×2) reconstructed surfaces at which one surface oxygen (zinc) ion of each four is removed at the oxygen (zinc) termination.⁹⁸ As a consequence, in this case the calculated surface energy corresponds to an average value of the zinc- and oxygen-terminated surfaces. In all calculations, slabs of 6-12 atomic layers are separated by about 15 Å of vacuum, and the atomic coordinates of all ions are fully relaxed. The calculated surface energies, Tab. SI2, are fully consistent with the existing computational evidence.^{90,99–103}

Three PLIPs are developed with different training datasets. Indeed, we train a first PLIP on a database composed of the 6 bulk polymorphs melted up to 5000 K. For those first MD runs, we use a classical Buckingham ZnO potential¹⁰⁴ which was previously employed in numerous studies of ZnO.^{105–108} We would like to emphasize that using a classical potential instead of first-principle calculations at this stage favors a more rapid sampling of a larger variety of configurations. Then, for the second database, we considered ZnO surfaces heated up to 2000 K using the first PLIP model as well as structures of the first database. Finally, for the third model, additional amorphous bulk obtained after rapidly cooling liquid structures from 2000 K to room temperature are added to the previous database. There, we employed the second PLIP model for the interaction potential. During the MD runs, we extracted respectively 20 and 10 for the bulk and the surface structures as well as 15 more amorphous structures. Thus, we sequentially increase the size and the variety of the database always using the previous model for sampling. Those three PLIPs are respectively denoted V1, V2, and V3 in the following. To validate the PLIPs, we perform a series of calculations where bulk

and surface properties are computed for the six ZnO polymorphs, and compared to DFT results. We employ the conjugate gradient algorithm, with a force convergence criterion of 10^{-9} eV/Å. Moreover, we also investigate the efficiency of the PLIPs in the context of disordered ZnO structures, such as liquid and amorphous phases. For this purpose, simulations are carried out by starting with 500 atoms randomly disposed and heated at 3000 K, and then cooled to 300 K with a rate of 1800 K/ns.

Molecular dynamics simulations of crystallization

After developing the PLIP models, we perform two complementary types of molecular dynamics simulations of crystallization.

On the one hand, in freezing simulations the temperature is progressively lowered as to initiate the crystallization. In this case, the system is initialized with 1000 randomly positioned atoms of zinc and the same number of oxygen atoms. Then, atomic positions are first optimized to remove large forces due to overlapping atoms. We further proceed to a first equilibration at 5000 K during 1 ps in the NVT ensemble with a Nose-Hoover thermostat. From preliminary freezing simulations (not shown) we found out that within the considered conditions crystallization occurs in the temperature range of 1500 K to 1000 K. Therefore, in the reported simulations, the freezing is operated in two steps. In the first step, the system is quenched from 2500 K to 1500 K during 100 ps which allows us to quickly obtain a first liquid structure. In the second step, the system is slowly cooled down from 1500 K to 1000 K during 10 ns. These two cooling steps are both performed in the NPT ensemble at $P = 1$ bar.

On the other hand, we also performed constant temperature simulations. For that purpose, liquid structures are generated just as in the freezing simulations. Then, a thermostat and a barostat are employed to maintain the temperature and pressure until spontaneous crystallization is observed.

The Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) package¹⁰⁹ with

a custom-made pair style for PLIP modeling is used in all the simulations.

Data-driven approach for the structural analysis

Analysis of the obtained structures and the crystallization dynamics requires the use of novel order parameters which are able to distinguish between each of the six crystal polymorphs of zinc oxide. Such a challenging task is clearly out of the reach of classical order parameter including Common neighbor analysis, Voronoi cell topology, or Polyhedral template matching which are mostly designed for monodisperse systems and for classical crystal structures like face-centered cubic and body centered cubic. In the case of zinc oxide crystallization, the very rich structural landscape requires a structurally agnostic order parameter, namely one that is transferable for all kinds of crystal structures.

In this context, we develop a data-driven approach for the structural analysis. In practice, averaged Steinhardt parameters denoted \bar{q}_l with $l \in [3 : 6]$ are computed for zinc and oxygen atoms when considering the entire system (zinc and oxygen atoms) and only inter-species distances (Zn-Zn and O-O). We use the `pyscal` Python package¹¹⁰ with the adaptive methods to compute the neighbor lists. This calculation leads to a Steinhardt vector denoted \mathbf{Q} of 12 components (6 for the entire system and 6 for inter-species distances) for each atom. Then, we compute the same vector in the six bulk polymorphs to construct a reference database averaged over all the zinc and oxygen atoms in those crystal structures. If a studied atom has the same local structure as one in the bulk polymorph, its Steinhardt vector must be close to that of the database which can be measured using Euclidean distance in vector space denoted D_{Y-X} . In Fig. 1, we test those Euclidean distances on nanoparticles of known structures and show that the method is able to identify the correct structure. In addition, the method retrieves that the center of the nanoparticle is more similar to the bulk than its edges.

Finally, in order to quantify these Euclidean

distances, we compute a likelihood parameter denoted S measured along a path between two reference structures named A and B using:

$$S = \frac{e^{-\lambda D_{AX}} + 2e^{-\lambda D_{BX}}}{e^{-\lambda D_{AX}} + e^{-\lambda D_{BX}}} \quad (4)$$

where λ is adjusted so that $D_{AB} = 2.3/\lambda$. As such, we obtain a score that quantifies the local structure: an atom which is similar to structure A (resp. B) gives $S \sim 1$ (resp. $S \sim 2$). Therefore, we consider that the value of 1.5 discriminates between the two types of structure.

In retrospective, our approach combines the use of (1) a vector of Steinhardt parameters which was already suggested for colloidal systems and amorphous models^{111,112}, (2) a similarity measurement with the identity function as Kernel, and (3) a path collective variable which was previously employed for the construction of order parameters based on permutation invariant vectors.¹¹³

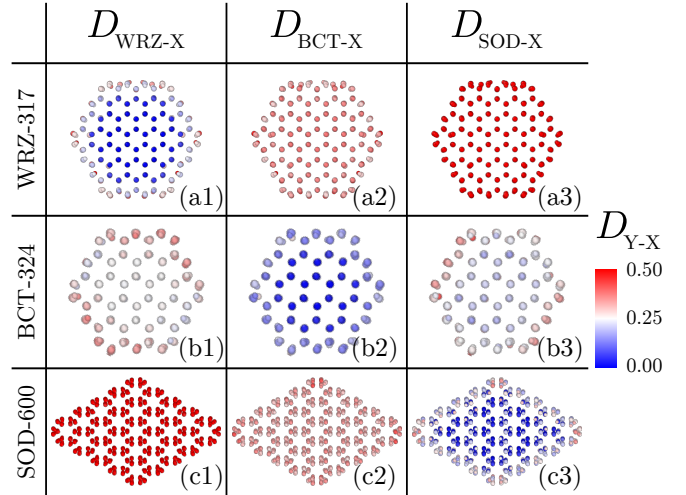


Figure 1: Structural analysis of three different nanoparticles obtained from Viñes et al.¹¹⁴ that are respectively in the WRZ (a), BCT (b) and SOD (c) crystal phases. We measured the distances with respect to the same three different crystal phases in bulk.

Results

Validation of the PLIP models

Bulk crystals

As a first step to validate the developed PLIPs, the crystal structures of the six polymorphs were optimized using the three PLIPs, V1, V2, and V3. The resulting lattice parameter a and cohesion energy E_{coh} relative to the lowest-energy WRZ phase were compared to those obtained using DFT in order to evaluate the accuracy of the PLIPs. Errors with respect to DFT values for these two quantities are displayed in Fig. 2.(a) and (b), respectively, along with results obtained with three classical force-field potentials, namely the Buckingham,¹⁰⁴ Tersoff,¹¹⁵ and ReaxFF¹¹⁶ potentials.

Regarding the three PLIPs, they give approximately the same results, with the V2 and V3 PLIPs being slightly more accurate than V1. The already accurate results obtained with V1 indicate that training with only bulk ZnO structures is sufficient to reach good model for the bulk crystals. Except for the Buckingham potential for which the errors are considerably higher, the two other classical potentials present comparable errors for all the polymorphs with respect to the PLIP models.

Accuracy on ZnO surfaces was also investigated and the errors on the surface energy are computed both using PLIPs and the classical potentials [See Fig. 2.(c)]. When comparing each PLIP version, we note that while V2 and V3 are accurate by construction, V1 in most cases leads to very good results in terms of surface energy although it was not trained with any surface structures. Contrary to bulk parameters, the classical potentials are less accurate at predicting the ZnO surface energies compared to the PLIPs. Surprisingly, the Buckingham potential, while the simplest in its mathematical formulation, exhibits the best overall results among the classical potentials. However, the PLIP errors are generally much lower than what is obtained with any of the classical potentials with an improvement of a factor between 3 and 8.5.

In order to further test the accuracy of PLIPs

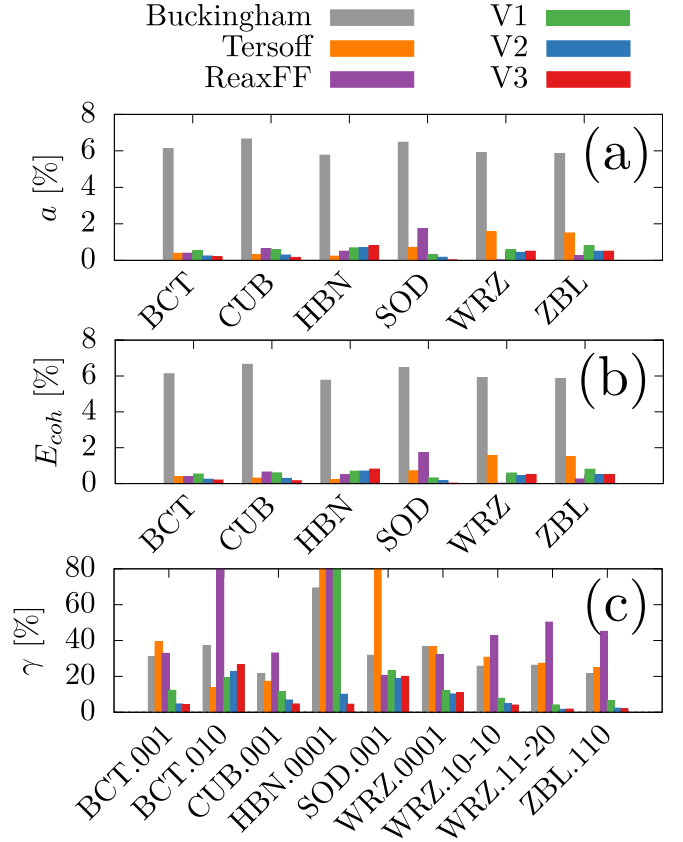


Figure 2: Absolute errors (%) with respect to the DFT values of (a) bulk lattice parameter a , (b) cohesion energy E_{coh} relative to the wurtzite polymorph, and (c) relative surface energy γ determined from the relaxed structures obtained with the three PLIPs as well as with three classical potentials. For signed values of the error, please refer to Fig. S1.

on bulk crystals, we computed the phonon density of states (DOS) for the three most stable crystal structures ie. WRZ, ZBL and BCT. For that purpose, the phonon frequencies are computed by diagonalizing of the dynamical matrix obtained through measuring the variation of atomic forces due to finite atomic displacements. To increase the calculation accuracy, we expanded the conventional unit cells following: WRZ ($5 \times 5 \times 3$), ZBL ($3 \times 3 \times 3$) and BCT ($3 \times 3 \times 5$). The phonon calculations were performed using the PHONOPY code.^{117,118} From Fig. 3, one can see that although the tested classical potentials provided acceptable results for static properties like lattice parameters and cohesive energies, they are extremely inaccurate for dynamical properties as shown by the DOS.

From a qualitative picture, ReaxFF which was the most accurate classical potentials already shows additional peaks in the DOS especially in the case of WRZ and ZBL. Quantitatively, the error of the three PLIPs is at least two times smaller than any of the classical potentials. Along with the previous measurement on surface energy, this results using DOS provides an additional argument for the necessity of constructing PLIP models for the zinc oxide system.

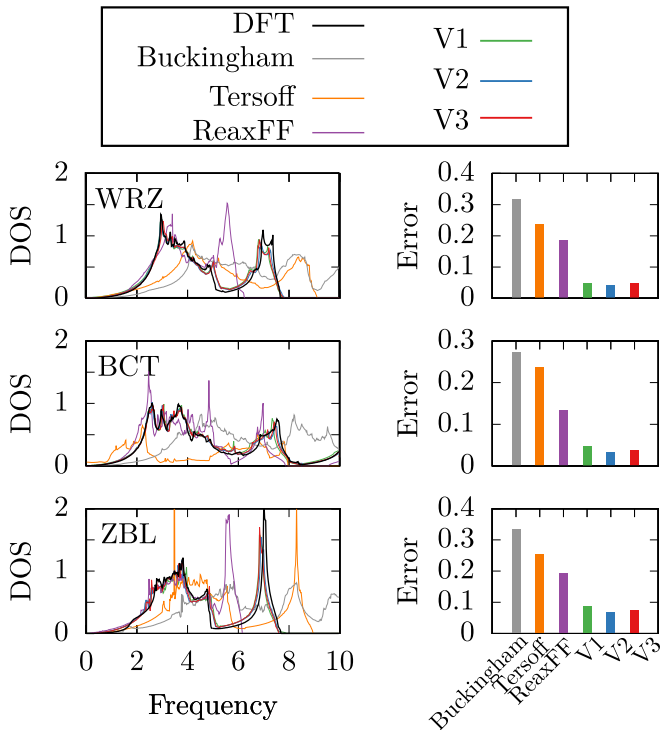


Figure 3: Phonon density of states in three crystal structures. Quantitative comparison is calculated as the difference of the obtained curves with respect to the DFT reference.

Finally, we note that we evaluated the melting temperature using the phase coexistence approach and obtained 1750 ± 50 K for the PLIP V3 model which is about 500 K lower than the experimental value. The excellent matching of PLIP and DFT vibrational characteristics for the key ZnO crystalline phases [See Fig. 3], suggests that this discrepancy originates from deficiencies not of the PLIP itself but of the DFT approximation.

Disordered regimes

After the crystal properties of ZnO, we were interested in measuring the accuracy of the model in the case of disordered structures. For this purpose, we performed ab initio MD simulations using three different initial amorphous structures in the NVT ensemble at 1500 K. The total duration of each simulation was equal to 4 ps with a timestep of 1 fs. The obtained structures, were analyzed by calculating the partial RDFs. In Fig. 4, we display results obtained with the DFT calculations, classical potentials and PLIP models. Quantitative errors based on the difference between RDF curves and DFT ones are also indicated for a better comparative evaluation. The RDFs coming from the PLIP structures fit perfectly those of DFT structures (visually and with the error measurements), while a mismatch is clearly visible between the DFT and the classical potential curves. Indeed, the main peak of the Buckingham potential curves is misaligned in all the partial RDFs, and both Tersoff and ReaxFF curves are completely different for the two homonuclear curves. In the case of O-O, the ReaxFF and Tersoff curves also predict an erroneous small peak at around 1.3 Å. Regarding the errors, the PLIPs are undoubtedly closer to the DFT results than all of the classical potentials, with errors generally 2 to 3 times smaller. Moreover, even if all the PLIP RDFs are similar, a slight advantage is given to the V3 curve, especially at the first peak. It remains that the training of V1 and V2 models on hot liquid structures seems sufficient to describe the amorphous structures stabilized at room temperature. Results on the bond angle distributions shown in Fig. S3 also displays a much better agreement with DFT calculations for all the PLIP models. We would like to emphasize that even if none of the DFT liquids were included in the database, the PLIP models are still able to extrapolate towards such high temperatures regime.

Prompted by those results on liquid structures, we further tested the obtained PLIP in the amorphous regime. In particular, To generate amorphous structures, a hot liquid was cooled until reaching the ambient temperature

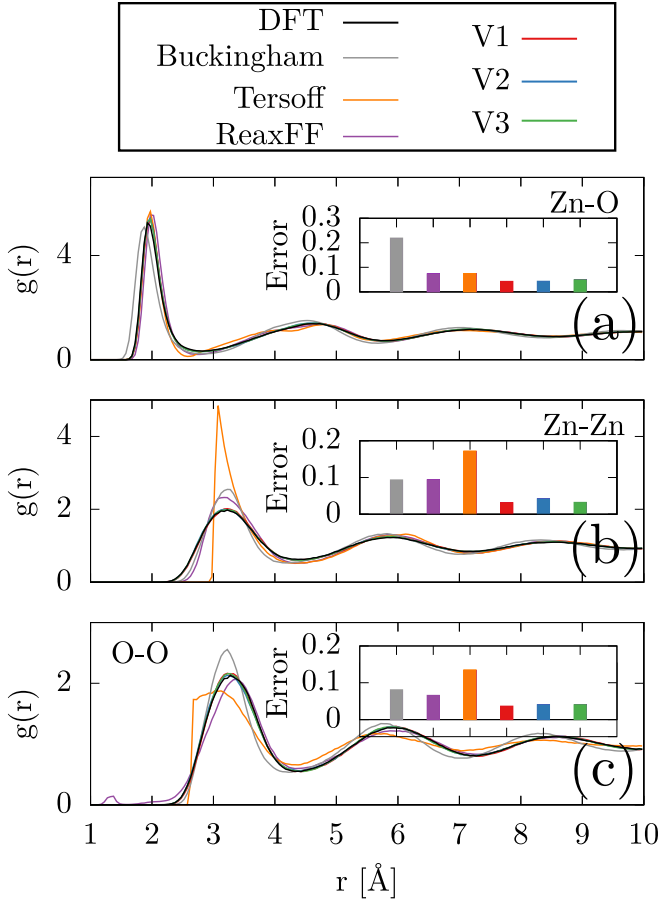


Figure 4: Partial radial distribution functions $g(r)$ of the liquid structures obtained at 1500 K for (a) Zn-O, (b) Zn-Zn and (c) O-O. Structures calculated with the three PLIPs are compared with those computed using DFT and the three potentials. Quantitative comparison is obtained with the absolute errors displayed in insets and calculated as the difference of the RDF curves with respect to the DFT reference.

with a cooling rate equal to 1.8×10^6 K/ns which allows for the formation of the amorphous form of ZnO. Finally, the obtained structures were further optimized in energy and forces down to respectively 10^{-4} eV and 10^{-6} eV/Å. As such, twenty different amorphous structures made of 360 atoms were constructed using different initial seeds and were analyzed by calculating the partial RDFs. In Fig. SI4 and SI5, we display results obtained with the DFT calculations, classical potentials and PLIP models. Results are very similar to what we obtained in the liquid regime. Indeed, while more simple in its mathematical formulation, the Buckingham potential

is generally the best among the classical force field with both Tersoff and ReaxFF showing nonphysical behavior respectively for the Zn-Zn and the O-O first neighbor. Then, all of the obtained PLIP exhibit a very good agreement with DFT calculations.

Nanostructures

For this last test of the PLIP models, we employed clusters of ZnO obtained by Viñes et al.¹¹⁴ using DFT. They focused on three polymorphs, namely BCT, WRZ and SOD, and also worked with single-caged (SC) and multi-caged (MC) structures and reported the most probable low energy structures for systems made of tens to hundreds of atoms. In our test, we additionally optimized these structures using our own DFT flavor, the PLIP models and the three potentials. With the obtained structures, we computed the formation energy difference ΔE with respect to DFT results. Moreover, as to assess the quality of cluster structures geometries, we measured the root-mean-square deviation (RMSD) based on the difference of atomic positions with respect to the DFT reference and extracted the maximum value obtained when examining every atom of the nanostructure. Fig. 5 shows that the classical potentials are not able to provide accurate results on these nanostructures while the three versions of PLIP give reliable results when compared to DFT calculations.

Interestingly, the PLIP models are able to accurately reproduce the energy and the geometry even for the smallest clusters that are qualitatively different from the bulk-like structures of the largest clusters, which is an evidence for their transferability towards untrained structures. In particular, the single cage SC-108 and SC-90 do not show any bulk-like structure since these are shell clusters, and the PLIP models still give very small errors. For the three PLIPs, a slight advantage is still given for the V2 and V3 models for which the ΔE and RMSDs do not exceed 1 % and 0.5 Å, respectively.

Altogether, it is worth noticing that although the PLIP models were only trained with DFT

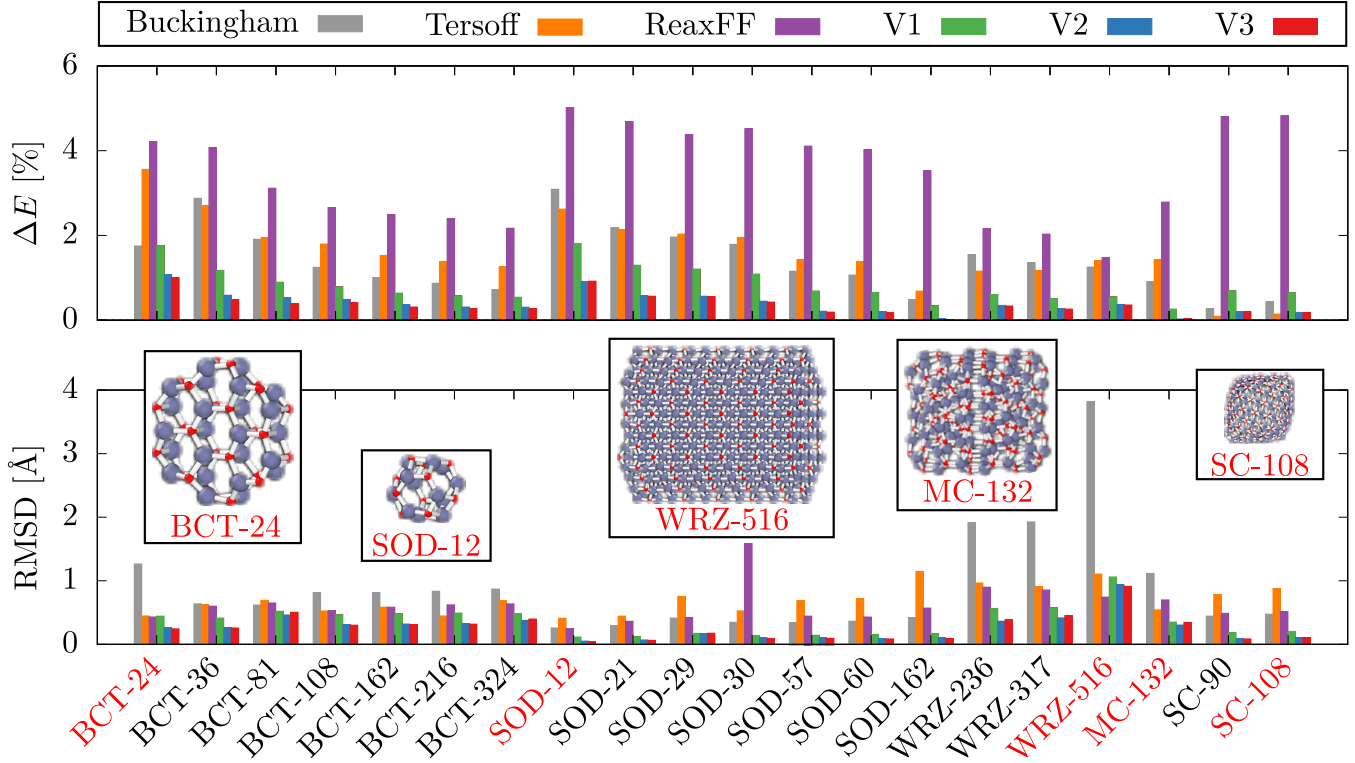


Figure 5: (Top panel) Relative absolute error on the clusters energy (per ZnO formula unit) relative to the Wurtzite bulk energy obtained with the PLIPs and the classical potentials. (Bottom panel) Corresponding RMSD based on comparison of atomic positions with those in DFT configurations obtained from structural optimization of geometries published by Viñes et al.¹¹⁴ Typical clusters are represented in the insets. For signed values of the errors on the energies, please refer to Fig. S2.

forces, they are still able to reach very accurate results for energetic properties (bulk, surface, and nanoparticles). From now on, results presented in the remaining article will consider only the calculations having employed the V3 potential because it is built with a more complete database.

Crystallization differences

In this last validation section, we focus on crystallization results. In particular, we performed the exact same freezing simulations using PLIP as well as the 3 classical force fields. First, we note that the Tersoff and the ReaxFF models do not lead to crystallization under the considered thermodynamics conditions. Indeed, in both cases, the difficulty in modeling the oxygen-oxygen short range interactions already raised in case of the disordered ZnO leads to a nonphysical behavior in the high temperature regime where freezing simulations are started. It is possible that those two classical potentials may also lead to crystallization under more finely tuned initial conditions, including, e.g., the initial distribution of atoms and temperature of the freezing runs. However, one may also expect that the erroneous description of the short range oxygen-oxygen interaction impacts the nucleation and the initial ZnO growth, where small grains are in contact with the liquid. Then, Fig. 6.(a,b) shows the structures obtained at the end of the freezing simulations with PLIP and the Buckingham model. A quantitative picture is obtained by using the Euclidean distances in Steinhardt vector space between the final structures and those in the bulk database. From Fig 6.c, it appears that the smallest distance for the PLIP results is obtained with the WRZ structure while for the Buckingham potential, it corresponds to the BCT structure. This quantitative observation suggests that this classical potential leads to a qualitatively different result, *i.e.* the growth of a BCT structure instead of the Wurtzite one, which is the most stable in both experiments and DFT calculations. This surprising result may be in part assigned to the fact that the classical potential overestimates stability of the

BCT phase. Indeed, according to bulk DFT results the BCT phase is less stable than the WRZ one by $0.048eV/ZnO$ but this energy difference is reduced to $0.017eV/ZnO$ only within the Buckingham potential. In contrast, the PLIP model is able to retrieve a much more satisfying value of $0.044eV/ZnO$. This analysis on the final structure obtained after freezing shows that PLIP is able to crystallize the most stable structure which is a strong justification for using the PLIP model over a classical potential.

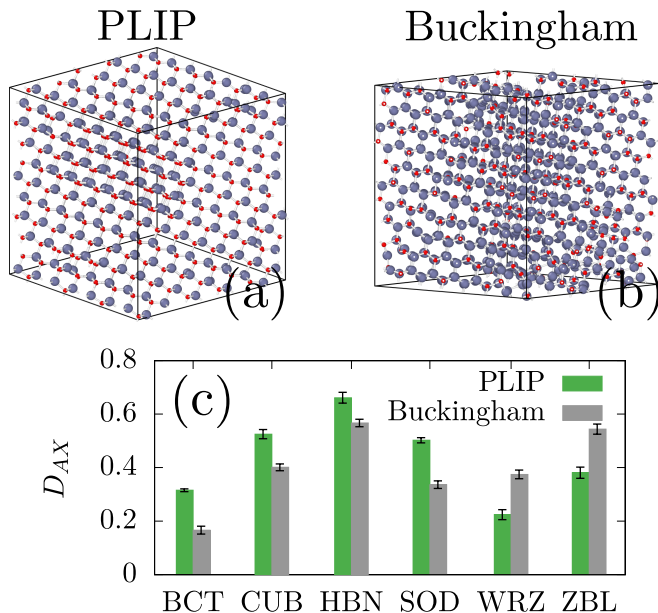


Figure 6: Structures obtained after freezing simulations using (a) the PLIP model and (b) the Buckingham potential, after a slow cooling during 10 ns. (c) Euclidean distances between the obtained structures and the bulk perfect crystals computed in the Steinhardt vector space. Results are averaged over 5 simulations obtained with different initial conditions.

Thus, one of the key advantages for using the PLIP model over classical interaction is that, along with the quantitatively better results for bulk, surface, disordered and nanoparticle properties, it also leads to qualitatively better outcomes of freezing simulations. In addition, when comparing to DFT, the main advantage is the possibility for larger scale simulations. This is crucial in studies of crystallization where freezing simulations have to cover more than a few nanoseconds because shorter ones

result in amorphous structures. Furthermore, much larger unit cells allowed by our PLIP help avoiding any nonphysical behavior due to interactions between nucleation cores in neighboring cells.

Computational efficiency

Before closing on this validation section, Fig. 7 shows the computational efficiency of different models of ZnO interactions. It appears that the PLIP potentials are between 4 and 6 orders of magnitude faster than DFT depending on the system size. While being visibly slower than the most simple potentials (Buckingham and Tersoff) their efficiency is comparable to the that of the comparatively less reliable ReaxFF model. Furthermore, we would like to quickly comment on how easy the methodology can be applied to other systems. A first bottleneck is the construction of the database. In our case, a sequential increase of the database size was carried out by using first a classical empirical potential and then the previous version of the machine-learning force fields. As such, this part of the methodology is not as CPU consuming as the usage of ab initio MD for the database sampling and it still allows us to rapidly obtain structures that are highly uncorrelated. In addition, a second key aspect of the overall methodology is the computational cost of the learning process. In our case, because we employ LassoLars which is a constrained linear regression methods, we can obtain a PLIP model in approximately 10 minutes using 16 CPU cores which is relatively a low expense compared to more advanced non-linear regression methods.

Temporal evolution during the ZnO crystallization

Following the section dedicated to the validation of the PLIP model, we will now describe results obtained for ZnO crystallization, and the insights they give into the nucleation scenarios for this material. Two complementary types of crystallization simulations were performed with settings described in details in the Methods Section.

Global observations

In freezing simulations the ZnO liquid was slowly cooled down from 1500 K to 1000 K during 10 ns as to initiate crystallization. Results of five individual runs performed under the same thermodynamic conditions but with different initial atom positions and velocities are shown in Fig. 8 (top). To complement these out-of-equilibrium simulations, we have also mimicked infinitely slow cooling rates with constant temperature MD runs. Figure 8 (bottom) reports the results obtained in six runs at $T = 1150$ K, initiated with different atom positions and velocities.

Regarding the structural evolution along the crystallization path, the behavior of the numbers of atoms in the predominant WRZ (green curve) and BCT (blue curve) environments reveals three growth stages in all the preformed runs: (1) an abrupt and quick increase of N_{WRZ} and N_{BCT} marks the nucleation and growth of WRZ and BCT structures from liquid ZnO; (2) a plateau more or less broad indicates the stabilization of coexisting BCT and WRZ phases; (3) a phase transformation of BCT into WRZ visible by a drop of the curve of the former, compensated by the increase of the latter. We stress that similar results are obtained in freezing simulations with even a faster cooling rate of 25 K/ns (See Fig. S4), with a larger periodic cells (4000 atoms) (See Fig. S5) and at different

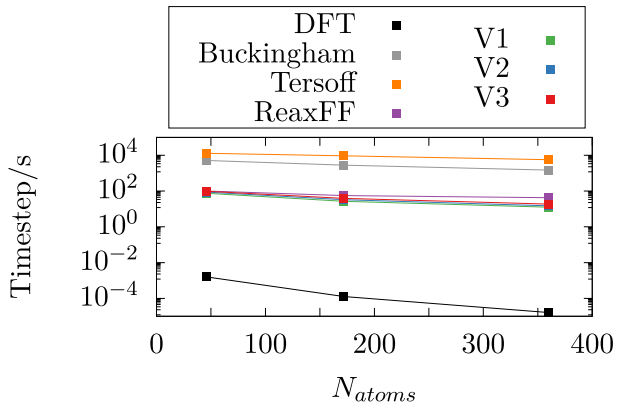


Figure 7: Computational efficiency of different models of ZnO interactions measured as the number of timesteps that can be reached in a second. Results were obtained on 16 CPU cores of Intel Skylake 6140 2.3 GHz.

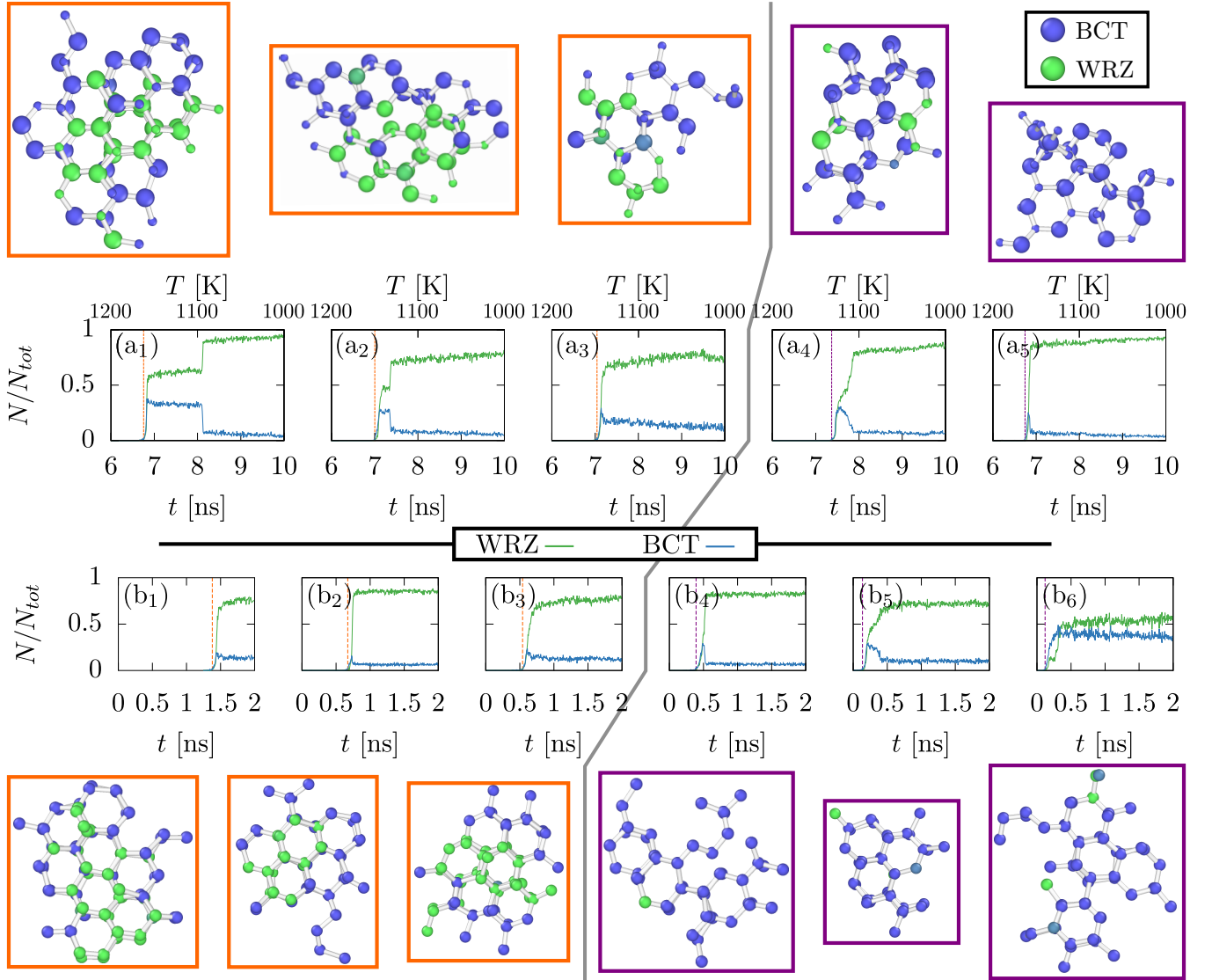


Figure 8: Number of atoms detected in WRZ or BCT phases, according to the likelihood parameter S , along the crystallization path for independent MD runs using 2000 atoms and initiated with different atom positions and velocities. Top panels: freezing simulations at cooling rate equal to 50 K/ns. Bottom panels: constant-temperature simulations at $T = 1150$ K. The red solid line discriminates between the two observed nucleation processes. Images of the typical nucleating structures show two nucleation scenarios: (A) WRZ surrounded by BCT (orange) and (B) Mostly BCT (violet).

constant temperature simulations ($T = 1100$ K and 1200 K) (See Fig. S6). This confirms that the observed nucleation and early ZnO growth characteristics are not biased by applied cooling rates (which are faster than these in most experimental situations) and are also not entangled by finite size effects.

Furthermore, Fig.8 shows typical pictures of the nucleation core during both the freezing and the equilibrium simulations. As such, we find two types of structures for the nucleating

clusters: (A) WRZ atoms surrounded by BCT phase and (B) Mostly BCT atoms. Among the reported results of 11 runs, we obtain 6 and 5 cases of scenario A and B, respectively. In the following, we will particularly focus on what happens during the first growth stage and we will further describe the two types of scenarios, exemplified in Figs. 9.

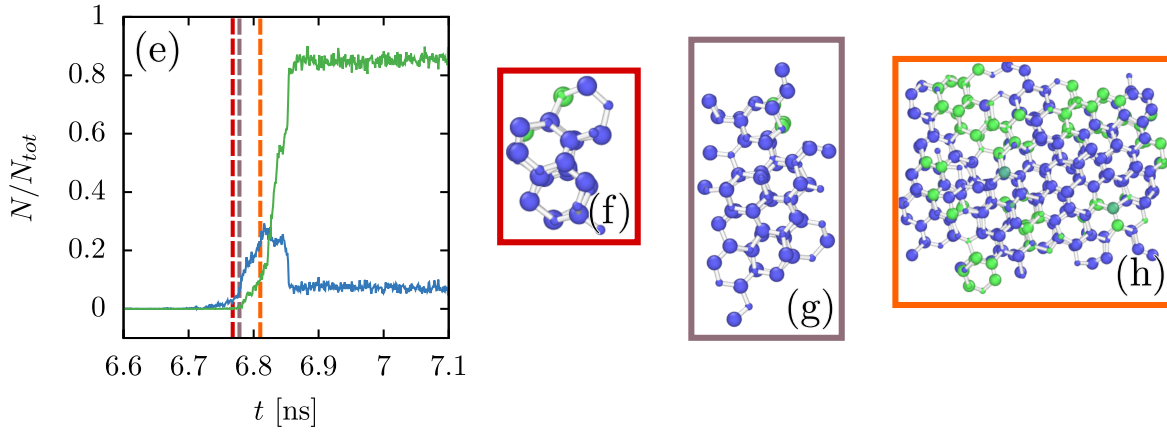
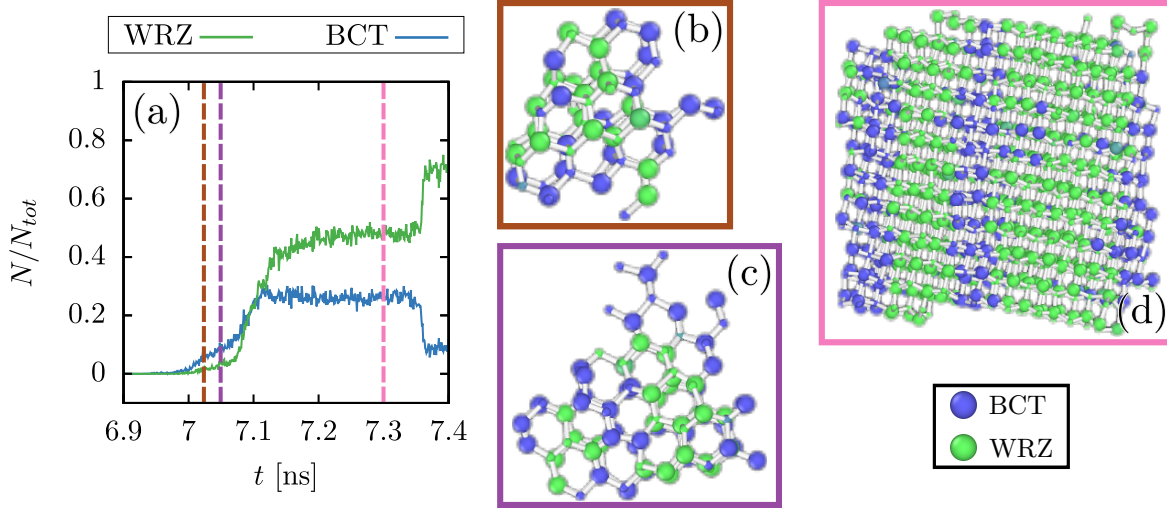


Figure 9: (a, e) Number of atoms detected in WRZ or BCT environments along the cooling path in typical first and second nucleation scenarios which correspond respectively to Fig.8(a₁,a₅). Colored dashed lines point out different times at which typical crystal structures shown in (b, c, d) and (f, g, h) were selected. Atoms identified as WRZ and BCT according to the likelihood parameter S are colored in green and blue, respectively.

Focus on the nucleation mechanisms

In the first scenario, Fig. 9 (top), the growth is initiated by a nucleation of a WRZ grain. The behavior of N_{WRZ} at the very beginning of the MD simulation, Fig. 9.a, shows that several small WRZ ZnO clusters form and dissolve in the ZnO liquid before the actual growth starts. From a structural point of view, the stability of the first grain seems to be enhanced by the formation of three-membered ZnO rings [See Fig.9.b], characteristic of the Wurtzite (0001) surface. Indeed, in absence of such rings, similarly-sized ZnO clusters tend to dissolve. Most interestingly, the emerging WRZ clusters are systematically surrounded by a BCT phase taking the form of relatively thin and less crys-

tallized structures [See Fig.9.c]. Then, during the growth, the grain becomes significantly elongated, with virtually parallel (0001) facets at its opposite sides thus suggesting that crystallization along the (0001) atomic planes is favored at this growth stage [See Fig.9.d]. Altogether, the first nucleation scenario comprises: (1) Stochastic appearance of unstable WRZ clusters that do not exhibit three-membered rings, (2) Emergence of a stable WRZ cluster made of three-membered rings and surrounded by a shell of BCT phase, and (3) Continuous growth of WRZ with a reminiscent BCT shell.

In contrast, the second scenario, Fig.9 (bottom), shows quite a different nucleation path. Indeed, while N_{WRZ} and N_{BCT} increase quasi-simultaneously in the first scenario, here, the

BCT phase emerges sooner than the WRZ one. Figs. 9.(f,g) clearly show that the emerging crystal is initially made of mostly BCT phase. Only once the BCT crystal is large enough, a solid-solid transition is observed with BCT turning locally into WRZ. As a consequence, the BCT structure becomes porous, with multiple embedded WRZ clusters and is not stable enough to continue persisting. It is remarkable that while the initial BCT structure is much larger than in the first scenario, it can not survive as long as in the previous scenario because of this growth of embedded WRZ grains. The second scenario of nucleation can be summarized as a two-step mechanisms: (1) Metastable BCT crystals constitute the nucleation core and (2) The stable WRZ crystal only appears within large BCT crystals. An interesting feature of this second scenario is that we no longer observe growth of elongated WRZ grains since WRZ phase originates from multiple embedded smaller clusters. The elongated grains should therefore be favored by the presence of BCT shells.

Discussions

Both these scenarios highlight the essential role played by the metastable BCT phase in ZnO nucleation and early growth stages. Indeed, either it appears in the ZnO liquid simultaneously with the stable WRZ phase in form of a shell surrounding the growing WRZ clusters, or it nucleates first and the embedded WRZ clusters appear inside the metastable BCT matrix as a result of a solid-solid transition. From a fundamental perspective, both of these mechanisms contradict the classical nucleation scenario in which the first emerging crystal is made entirely of the most stable phase. Such non-classical nucleation scenarios have been already observed in simulations for model systems described by simple hard spheres^{24,25} and Lennard-Jones interactions^{11,26} but, to our knowledge, the present prediction is among the first ones for systems with more complex, ionic-covalent bonding thus confirming seminal experimental observations.^{119–121}

While a thorough analysis of physical reasons of BCT involvement in the nucleation scenarios goes beyond the scope of the present study, it is worth pointing out that thermodynamic and kinetic effects may tend to interplay. On the one hand, from a kinetic viewpoint, the involvement of BCT phase in ZnO crystallization may result from a kinetic preference for an attachment pattern of ZnO₄ tetrahedral units during crystal growth which favors the BCT phase. On the other hand, from a thermodynamic viewpoint, taking into account the small stability difference at 0 K between the WRZ and BCT phases (0.044 eV/ZnO), the effect of lattice vibrations of differently arranged tetrahedral ZnO₄ units in the two lattices (corner-connected only in WRZ, whereas some units are also edge-connected in BCT) may impact the relative stability of the two polymorphs at higher temperatures. In fact, our constant-temperature simulations, Fig. S6, tend to show that the tendency to form long-lasting BCT increases with the temperature, which may suggest that BCT is less unstable at higher temperature. In order to better understand the thermodynamic influence over the crystallization results, we first used the phonon dispersion curves shown in Fig. 3.a to deduce the free energy in the harmonic approximation. Fig. 10.a shows that PLIP model is able to retrieve DFT values for the free energy curves which advocate furthermore for the accuracy of our MLIP approach. However, we must note that these calculations were obtained under the harmonic approximation and as such they can not be used to investigate the thermodynamic picture at high temperatures. To go beyond the harmonic approximation, we use a newly developed automated nonequilibrium thermodynamic integration method.¹²² In particular, we constructed two BCT and WRZ crystals made of 2880 atoms. We used thermodynamic integration going from the Einstein crystal to the PLIP model during 50 ps. Three cycles are carried out to determine the error which is equal to 0.004 eV/ZnO. Because of its large computational cost, such calculations are beyond the reach of DFT calculations but the PLIP excellent agreement with DFT within the

harmonic approximation supports the extrapolation towards this other regime. Fig. 10.b shows that at such higher level of approximation, it exists a stability reversal between WRZ and BCT phases as the temperatures increases. This points towards thermodynamics as one of key factors at the origin of the reported non-classical nucleation mechanisms.

Further understanding of this thermodynamic picture would involve the measurement of the solid–liquid interfacial energies for both BCT and WRZ as well as rare-event sampling to compute the free energy barriers between the liquid phase and both WRZ and BCT crystals and also between the two crystals. However, the appropriate choice of collective variables for free energy calculations is already a fundamental challenge because of the very complex structural landscape. Altogether, these calculations require an enormous computational effort that can hardly be reached even with the PLIP model.

Unfortunately, regarding experimental findings, the existing evidence for the bulk BCT phase of ZnO is relatively small since most studies focused on its appearance in ZnO ultra-thin films or axially strained nano-objects. Indeed, the BCT phase has been systematically predicted as the most stable ZnO polymorph in the thin film geometry,^{123,124} due to the more favorable energies of its low index surfaces which, in contrast to the basal WRZ(0001) surface, are

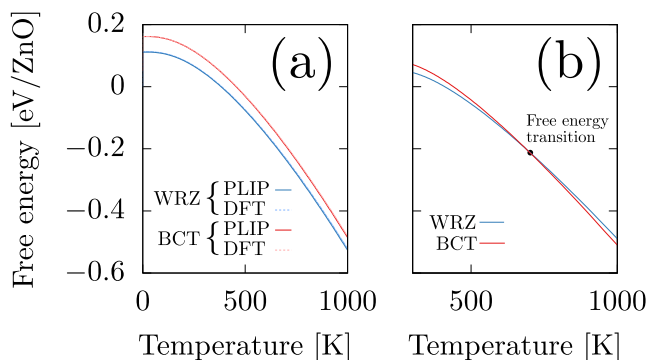


Figure 10: Free energy curves obtained under the harmonic approximation in DFT and PLIP V3 (a) and using the automated nonequilibrium thermodynamic integration method and the PLIP V3 (b).

non-polar. Also, a reversible WRZ-BCT reconstruction in the outmost layers of ZnO has been observed at some of ZnO surfaces.¹²⁵ These findings suggest that the shape/morphology of the crystallizing ZnO grains may also be an important factor.

An analogy can still be drawn with the case of titania where rutile and anatase phases tend to compete as observed with BCT and WRZ in zinc oxide.¹²⁶ Anatase is considered as a metastable phase which can be kinetically stabilized at lower temperatures, owing to its less constrained structure and consequently enhanced kinetics of formation.¹²⁷ It has been pointed out that the more rapid crystallization of anatase may also be due to the lower surface free energy of this polymorph compared to that of rutile.¹²⁸ From a computational viewpoint, zero-temperature DFT-based calculations predicted bulk anatase to be more stable than rutile and it is only most recently, that fine Diffusion Monte Carlo (DMC) simulations has suggested that bulk rutile may become more stable than anatase at higher temperatures, where the effect of lattice vibrations becomes crucial.¹²⁹ Similar conclusions were drawn from recent Self-Interaction-Error-corrected DFT calculations.¹³⁰ However, the complete nucleation picture drawn at the atomic scale remains unclear in the case of titania since very few finite temperature MD simulations were performed^{131–133} and no a single one focused on polymorph selection because short ab initio MD or classical empirical force fields were employed.

Conclusions

Despite the key technological importance of ZnO in many applications, its crystallization features are at present relatively poorly understood. This is due to overwhelming computational effort necessary for dedicated simulations, which excludes a direct use of DFT-based methods and requires fine interatomic potentials, able to correctly account for the structural diversity of ZnO as well as for its behaviour at finite and high temperatures.

With the goal to observe crystal nucleation

in a homogeneous ZnO liquid phase and to follow the subsequent crystal growth, we have constructed and validated a new robust machine learning interatomic potential suitable for large-scale simulations (nanosecond molecular dynamics of several thousands of atoms). The potential is based on physically-motivated mathematical formulation and a constrained LassoLars method is used to identify and adjust its pertinent parameters. The training database was composed of DFT-GGA results on a variety of ordered crystalline structures while also including surface and disordered configurations. We show that the new interaction potential successfully reproduces the delicate structural and energetic DFT-GGA characteristics of six of the most stable ZnO bulk polymorphs as well as of their low-index surfaces. The potential was also successfully tested on completely untrained structures made of nanoparticles in different crystal phases and was shown to reproduce the vibrational characteristics and free energy behaviour of the most stable bulk polymorphs. Importantly, it also correctly accounts for the simulated structural characteristics of high temperature liquid ZnO such that they become quasi-indistinguishable from the DFT ones. These extensive testings demonstrate the reliability and the transferability of the PLIP model which was then employed to study the crystallization of zinc oxide.

In particular, we have performed molecular dynamics simulations on freezing of bulk ZnO liquid as well as equilibrium calculations of undercooled ZnO liquid crystallization under a constant temperature and pressure conditions. Analysis of the crystalline phase in such complex structural landscape was made possible by employing a data-driven approach based on Steinhardt’s parameters which enabled us to access the nucleation of the initial zinc oxide crystal at the atomistic level and to follow its behavior upon further growth. We have systematically observed two different scenarios which are both in contradiction to classical nucleation theory: (1) Prenucleation clusters made of two coexisting phases, WRZ in the core and BCT in the shell and (2) Two-step nucleation process with metastable BCT emerging first and

turning into stable WRZ afterwards. Similar nucleation scenarios were previously observed in much simpler systems including Lennard-Jones, hard-sphere, and pure metals. In our case, the combination of the Physical LassoLars interaction potential that allows for a quantum-accurate modeling of zinc oxide with our data-driven approach for a refined structural analysis of crystal phases leads us to confirm previous predictions of non-classical nucleation processes.^{119–121} Moreover, dedicated calculations of ZnO free energy within a newly developed automated nonequilibrium thermodynamic integration method showed the existence of a stability reversal between wurtzite and BCT polymorphs at high temperatures and thus of a clear thermodynamic bias for such non-classical nucleation pathways.

As a perspective, we note that the proposed work paves the way for more complex numerical studies that would account the influence of pressure, finite liquid size and/or presence of surfaces, substrate, and/or temperature gradient, relevant for fabrication of technologically important ZnO nanostructures and should help understanding the relationship between their characteristics and the conditions of their early growth stages.

Acknowledgement

JL acknowledges financial support of the Fonds de la Recherche Scientifique - FNRS. Computational resources have been provided by the Consortium des Equipements de Calcul Intensif (CECI), by the Fédération Lyonnaise de Modélisation et Sciences Numériques (FLMSN) and by the Regional Computer Center CALMIP in Toulouse. JL thanks David Mora Fonz, Sarath Menon and Anup Pandey for fruitful discussions.

Data availability

The data generated in this study are available on Github [link will be given upon acceptance]

Supporting Information Available

The Supplementary information contains: (1) DFT results associated with cohesive and surface energies of the studied crystals, (2) Signed errors for bulk lattice parameter, cohesion energy and surface energy as well as for cluster energies (3) Liquid and amorphous structural properties (4) Additional crystallization results obtained with freezing and equilibrium simulations.

References

- (1) Ten Wolde, P. R.; Ruiz-Montero, M. J.; Frenkel, D. Numerical Evidence for bcc Ordering at the Surface of a Critical fcc Nucleus. *Phys. Rev. Lett.* **1995**, *75*, 2714–2717.
- (2) Wolde, P. R. t.; Frenkel, D. Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations. *Science* **1997**, *277*, 1975–1978.
- (3) Auer, S.; Frenkel, D. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. *Nature* **2001**, *409*, 1020–1023.
- (4) Lutsko, J. F.; Lam, J. Long-wavelength density fluctuations as nucleation precursors. *Phys. Rev. E* **2020**, *101*, 052122.
- (5) Lutsko, J. F. How crystals form: A theory of nucleation pathways. *Sci. Adv.* **2019**, *5*, 7399.
- (6) Sleutel, M.; Van Driessche, A. E. S. Role of clusters in nonclassical nucleation and growth of protein crystals. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E546–E553.
- (7) Van Driessche, A. E. S.; Van Gerven, N.; Bomans, P. H. H.; Joosten, R. R. M.; Friedrich, H.; Gil-Carton, D.; Sommerdijk, N. A. J. M.; Sleutel, M. Molecular nucleation mechanisms and control strategies for crystal polymorph selection. *Nature* **2018**, *556*, 89–94.
- (8) Menon, S.; Díaz Leines, G.; Drautz, R.; Rogal, J. Role of pre-ordered liquid in the selection mechanism of crystal polymorphs during nucleation. *J. Chem. Phys.* **2020**, *153*, 104508.
- (9) Lam, J.; Lutsko, J. F. Lattice induced crystallization of nanodroplets: the role of finite-size effects and substrate properties in controlling polymorphism. *Nanoscale* **2018**, *10*, 4921–4926.
- (10) Amodeo, J.; Pietrucci, F.; Lam, J. Out-of-Equilibrium Polymorph Selection in Nanoparticle Freezing. *J. Phys. Chem. Lett.* **2020**, *11*, 8060–8066.
- (11) Desgranges, C.; Delhommelle, J. Controlling Polymorphism during the Crystallization of an Atomic Fluid. *Phys. Rev. Lett.* **2007**, *98*, 235502.
- (12) Desgranges, C.; Delhommelle, J. Insights into the Molecular Mechanism Underlying Polymorph Selection. *J. Am. Chem. Soc.* **2006**, *128*, 15104–15105.
- (13) Yamazaki, T.; Kimura, Y.; Vekilov, P. G.; Furukawa, E.; Shirai, M.; Matsumoto, H.; Van Driessche, A. E. S.; Tsukamoto, K. Two types of amorphous protein particles facilitate crystal nucleation. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 2154–2159.
- (14) Sleutel, M.; Lutsko, J.; Van Driessche, A. E. S.; Durán-Olivencia, M. A.; Maes, D. Observing classical nucleation theory at work by monitoring phase transitions with molecular precision. *Nat. Commun.* **2014**, *5*, 1–8.
- (15) Ogata, A. F.; Rakowski, A. M.; Carpenter, B. P.; Fishman, D. A.; Merham, J. G.; Hurst, P. J.; Patterson, J. P. Direct Observation of Amorphous Precursor Phases in the Nucleation of Protein–Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2020**, *142*, 1433–1442.

- (16) van Blaaderen, A.; Ruel, R.; Wiltzius, P. Template-directed colloidal crystallization. *Nature* **1997**, *385*, 321–324.
- (17) Leunissen, M. E.; Christova, C. G.; Hynninen, A.-P.; Royall, C. P.; Campbell, A. I.; Imhof, A.; Dijkstra, M.; van Roij, R.; van Blaaderen, A. Ionic colloidal crystals of oppositely charged particles. *Nature* **2005**, *437*, 235–240.
- (18) Velikov, K. P.; Christova, C. G.; Dulens, R. P. A.; van Blaaderen, A. Layer-by-Layer Growth of Binary Colloidal Crystals. *Science* **2002**, *296*, 106–109.
- (19) de Nijs, B.; Dussi, S.; Smalenburg, F.; Meeldijk, J. D.; Groenendijk, D. J.; Fillion, L.; Imhof, A.; van Blaaderen, A.; Dijkstra, M. Entropy-driven formation of large icosahedral colloidal clusters by spherical confinement. *Nat. Mater.* **2015**, *14*, 56–60.
- (20) Sosso, G. C.; Chen, J.; Cox, S. J.; Fitzner, M.; Pedevilla, P.; Zen, A.; Michaelides, A. Crystal Nucleation in Liquids: Open Questions and Future Challenges in Molecular Dynamics Simulations. *Chem. Rev.* **2016**, *116*, 7078–7116.
- (21) Coli, G. M.; Dijkstra, M. An Artificial Neural Network Reveals the Nucleation Mechanism of a Binary Colloidal AB13 Crystal. *ACS Nano* **2021**, *15*, 4335–4346.
- (22) Russo, J.; Tanaka, H. The microscopic pathway to crystallization in supercooled liquids. *Sci. Rep.* **2012**, *2*, 1–8.
- (23) Leoni, F.; Russo, J. Nonclassical Nucleation Pathways in Stacking-Disordered Crystals. *Phys. Rev. X* **2021**, *11*, 031006.
- (24) Pusey, P. N.; Zaccarelli, E.; Valeriani, C.; Sanz, E.; Poon, W. C. K.; Cates, M. E. Hard spheres: crystallization and glass formation. *Philos. Trans. Royal Soc. A* **2009**, *367*, 4993–5011.
- (25) Sanz, E.; Valeriani, C.; Zaccarelli, E.; Poon, W. C. K.; Pusey, P. N.; Cates, M. E. Crystallization Mechanism of Hard Sphere Glasses. *Phys. Rev. Lett.* **2011**, *106*, 215701.
- (26) Trudu, F.; Donadio, D.; Parrinello, M. Freezing of a Lennard-Jones Fluid: From Nucleation to Spinodal Regime. *Phys. Rev. Lett.* **2006**, *97*, 105701.
- (27) Lin, C.; Corem, G.; Godsi, O.; Alexandrowicz, G.; Darling, G. R.; Hodgson, A. Ice Nucleation on a Corrugated Surface. *J. Am. Chem. Soc.* **2018**, *140*, 15804–15811.
- (28) Metya, A. K.; Molinero, V. Is Ice Nucleation by Organic Crystals Nonclassical? An Assessment of the Monolayer Hypothesis of Ice Nucleation. *J. Am. Chem. Soc.* **2021**, *143*, 4607–4624.
- (29) Qiu, Y.; Molinero, V. Why Is It So Difficult to Identify the Onset of Ice Premelting? *J. Phys. Chem. Lett.* **2018**, *9*, 5179–5182.
- (30) Qiu, Y.; Hudait, A.; Molinero, V. How Size and Aggregation of Ice-Binding Proteins Control Their Ice Nucleation Efficiency. *J. Am. Chem. Soc.* **2019**, *141*, 7439–7452.
- (31) Lupi, L.; Hudait, A.; Molinero, V. Heterogeneous Nucleation of Ice on Carbon Surfaces. *J. Am. Chem. Soc.* **2014**, *136*, 3156–3164.
- (32) Moore, E. B.; Molinero, V. Structural transformation in supercooled water controls the crystallization rate of ice. *Nature* **2011**, *479*, 506–508.
- (33) Lupi, L.; Hudait, A.; Peters, B.; Grünwald, M.; Gotchy Mullen, R.; Nguyen, A. H.; Molinero, V. Role of stacking disorder in ice nucleation. *Nature* **2017**, *551*, 218–222.
- (34) Russo, J.; Romano, F.; Tanaka, H. New metastable form of ice and its role in

- the homogeneous crystallization of water. *Nat. Mater.* **2014**, *13*, 733–739.
- (35) Desgranges, C.; Delhommelle, J. Molecular Insight into the Pathway to Crystallization of Aluminum. *J. Am. Chem. Soc.* **2007**, *129*, 7012–7013.
- (36) Combettes, S.; Lam, J.; Benzo, P.; Ponchet, A.; Casanove, M.-J.; Calvo, F.; Benoit, M. How interface properties control the equilibrium shape of core–shell Fe–Au and Fe–Ag nanoparticles. *Nanoscale* **2020**, *12*, 18079–18090.
- (37) Laurens, G.; Amans, D.; Lam, J.; Al-louche, A.-R. Comparison of aluminum oxide empirical potentials from cluster to nanoparticle. *Phys. Rev. B* **2020**, *101*, 045427.
- (38) Calvo, F.; Combe, N.; Morillo, J.; Benoit, M. Modeling Iron–Gold Nanoparticles Using a Dedicated Semi-Empirical Potential: Application to the Stability of Core–Shell Structures. *J. Phys. Chem. C* **2017**, *121*, 4680–4691.
- (39) Zipoli, F.; Curioni, A. Reactive potential for the study of phase-change materials: GeTe. *New J. Phys.* **2013**, *15*, 123006.
- (40) Rajasekaran, G.; Kumar, R.; Parashar, A. Tersoff potential with improved accuracy for simulating graphene in molecular dynamics environment. *Mater. Res. Express* **2016**, *3*, 035011.
- (41) Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; Verstraelen, T.; Grama, A.; van Duin, A. C. T. The ReaxFF reactive force-field: development, applications and future directions. *npj Comput. Mater.* **2016**, *2*, 1–14.
- (42) Brommer, P.; Kiselev, A.; Schopf, D.; Beck, P.; Roth, J.; Trebin, H.-R. Classical interaction potentials for diverse materials from ab initio data: a review of potfit. *Model. Simul. Mater. Sci. Eng.* **2015**, *23*, 074002.
- (43) Vashishta, P.; Kalia, R. K.; Nakano, A.; Rino, J. P. Interaction potentials for alumina and molecular dynamics simulations of amorphous and liquid alumina. *J. Appl. Phys.* **2008**, *103*, 083504.
- (44) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- (45) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.
- (46) Seko, A.; Takahashi, A.; Tanaka, I. First-principles interatomic potentials for ten elemental metals via compressed sensing. *Phys. Rev. B* **2015**, *92*, 054113.
- (47) Seko, A.; Takahashi, A.; Tanaka, I. Sparse representation for a potential energy surface. *Phys. Rev. B* **2014**, *90*, 024101.
- (48) Takahashi, A.; Seko, A.; Tanaka, I. Conceptual and practical bases for the high accuracy of machine learning interatomic potentials: Application to elemental titanium. *Phys. Rev. Materials* **2017**, *1*, 063801.
- (49) Seko, A.; Togo, A.; Tanaka, I. Group-theoretical high-order rotational invariants for structural representations: Application to linearized machine learning interatomic potential. *Phys. Rev. B* **2019**, *99*, 214108.
- (50) Goryaeva, A. M.; Maillet, J.-B.; Marinica, M.-C. Towards better efficiency of interatomic linear machine learning potentials. *Comput. Mater. Sci.* **2019**, *166*, 200–209.
- (51) Benoit, M.; Amodeo, J.; Combettes, S.; Khaled, I.; Roux, A.; Lam, J. Measuring

- transferability issues in machine-learning force fields: the example of gold–iron interactions with linearized potentials. *Mach. Learn.: Sci. Technol.* **2020**, *2*, 025003.
- (52) Thompson, A. P.; Swiler, L. P.; Trott, C. R.; Foiles, S. M.; Tucker, G. J. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **2015**, *285*, 316–330.
- (53) Wood, M. A.; Thompson, A. P. Extending the accuracy of the SNAP interatomic potential form. *J. Chem. Phys.* **2018**, 241721.
- (54) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, e1603015.
- (55) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 1–10.
- (56) Shapeev, A. V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Simul.* **2016**,
- (57) Novikov, I. S.; Gubaev, K.; Podryabinkin, E. V.; Shapeev, A. V. The MLIP package: moment tensor potentials with MPI and active learning. *Mach. Learn.: Sci. Technol.* **2020**, *2*, 025002.
- (58) Novoselov, I. I.; Yanilkin, A. V.; Shapeev, A. V.; Podryabinkin, E. V. Moment tensor potentials as a promising tool to study diffusion processes. *Comput. Mater. Sci.* **2019**, *164*, 46–56.
- (59) Takahashi, A.; Seko, A.; Tanaka, I. Linearized machine-learning interatomic potentials for non-magnetic elemental metals: Limitation of pairwise descriptors and trend of predictive power. *J. Chem. Phys.* **2018**, *148*, 234106.
- (60) Zeni, C.; Rossi, K.; Glielmo, A.; Fekete, Á.; Gaston, N.; Baletto, F.; De Vita, A. Building machine learning force fields for nanoclusters. *J. Chem. Phys.* **2018**, *148*, 241739.
- (61) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C* **2017**, *121*, 511–522.
- (62) Bereau, T.; DiStasio, R. A.; Tkatchenko, A.; von Lilienfeld, O. A. Non-covalent interactions across organic and biological subsets of chemical space: Physics-based potentials parametrized from machine learning. *J. Chem. Phys.* **2018**, 241706.
- (63) Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *J. Chem. Phys.* **2019**, *150*, 114102.
- (64) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **2017**, *3*, e1701816.
- (65) Veit, M.; Jain, S. K.; Bonakala, S.; Rudra, I.; Hohl, D.; Csányi, G. Equation of State of Fluid Methane from First Principles with Machine Learning Potentials. *J. Chem. Theory Comput.* **2019**, *15*, 2574–2586.
- (66) Nguyen, T. T.; Székely, E.; Imbalzano, G.; Behler, J.; Csányi, G.; Ceriotti, M.; Götz, A. W.; Paesani, F. Comparison of permutationally invariant

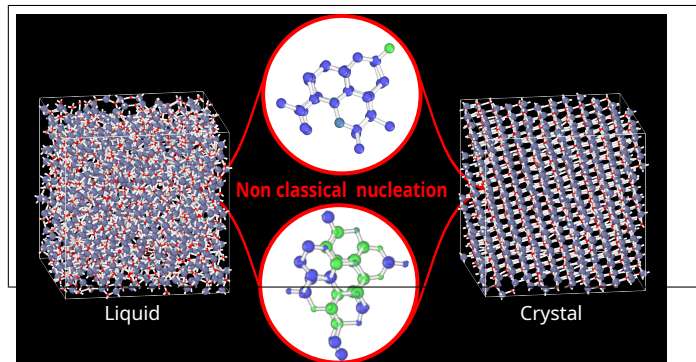
- polynomials, neural networks, and Gaussian approximation potentials in representing water interactions through many-body expansions. *J. Chem. Phys.* **2018**, 241725.
- (67) Bartók, A. P.; Gillan, M. J.; Manby, F. R.; Csányi, G. Machine-learning approach for one- and two-body corrections to density functional theory: Applications to molecular and condensed water. *Phys. Rev. B* **2013**, 88, 054104.
- (68) Morawietz, T.; Sharma, V.; Behler, J. A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges. *J. Chem. Phys.* **2012**, 136, 064103.
- (69) Natarajan, S. K.; Behler, J. Neural network molecular dynamics simulations of solid-liquid interfaces: water at low-index copper surfaces. *Phys. Chem. Chem. Phys.* **2016**, 18, 28704–28725.
- (70) Morawietz, T.; Behler, J. A Density-Functional Theory-Based Neural Network Potential for Water Clusters Including van der Waals Corrections. *J. Phys. Chem. A* **2013**, 117, 7356–7366.
- (71) Deringer, V. L.; Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* **2017**, 95, 094203.
- (72) Bartók, A. P.; Kermode, J.; Bernstein, N.; Csányi, G. Machine Learning a General-Purpose Interatomic Potential for Silicon. *Phys. Rev. X* **2018**, 8, 041048.
- (73) Caro, M. A.; Deringer, V. L.; Koskinen, J.; Laurila, T.; Csányi, G. Growth Mechanism and Origin of High sp^3 Content in Tetrahedral Amorphous Carbon. *Phys. Rev. Lett.* **2018**, 120, 166101.
- (74) Deringer, V. L.; Bernstein, N.; Bartók, A. P.; Cliffe, M. J.; Kerber, R. N.; Marbella, L. E.; Grey, C. P.; Elliott, S. R.; Csányi, G. Realistic Atomistic Structure of Amorphous Silicon from Machine-Learning-Driven Molecular Dynamics. *J. Phys. Chem. Lett.* **2018**, 9, 2879–2885.
- (75) Deringer, V. L.; Caro, M. A.; Jana, R.; Aarva, A.; Elliott, S. R.; Laurila, T.; Csányi, G.; Pastewka, L. Computational Surface Chemistry of Tetrahedral Amorphous Carbon by Combining Machine Learning and Density Functional Theory. *Chem. Mater.* **2018**, 30, 7438–7445.
- (76) Sosso, G. C.; Deringer, V. L.; Elliott, S. R.; Csányi, G. Understanding the thermal properties of amorphous solids using machine-learning-based interatomic potentials. *Mol. Simul.* **2018**, 44, 866–880.
- (77) Jinnouchi, R.; Lahnsteiner, J.; Karsai, F.; Kresse, G.; Bokdam, M. Phase Transitions of Hybrid Perovskites Simulated by Machine-Learning Force Fields Trained on the Fly with Bayesian Inference. *Phys. Rev. Lett.* **2019**, 122, 225701.
- (78) Sivaraman, G.; Krishnamoorthy, A. N.; Baur, M.; Holm, C.; Stan, M.; Csányi, G.; Benmore, C.; Vázquez-Mayagoitia, Á. Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide. *npj Comput. Mater.* **2020**, 6, 1–8.
- (79) Artrith, N.; Hiller, B.; Behler, J. Neural network potentials for metals and oxides – First applications to copper clusters at zinc oxide. *Phys. Status Solidi B* **2013**, 250, 1191–1203.
- (80) Eckhoff, M.; Lausch, K. N.; Blöchl, P. E.; Behler, J. Predicting oxidation and spin states by high-dimensional neural networks: Applications to lithium manganese oxide spinels. *J. Chem. Phys.* **2020**, 153, 164107.
- (81) Artrith, N.; Morawietz, T.; Behler, J. High-dimensional neural-network poten-

- tials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B* **2011**, *83*, 153101.
- (82) Sundararaman, S.; Huang, L.; Ispas, S.; Kob, W. New optimization scheme to obtain interaction potentials for oxide glasses. *J. Chem. Phys.* **2018**, *148*, 194504.
- (83) Eckhoff, M.; Schönwald, F.; Risch, M.; Volkert, C. A.; Blöchl, P. E.; Behler, J. Closing the gap between theory and experiment for lithium manganese oxide spinels using a high-dimensional neural network potential. *Phys. Rev. B* **2020**, *102*, 174102.
- (84) Artrith, N.; Urban, A.; Ceder, G. Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species. *Phys. Rev. B* **2017**, *96*, 014112.
- (85) Artrith, N.; Urban, A. An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO₂. *Comput. Mater. Sci.* **2016**, *114*, 135–150.
- (86) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **2021**, *12*, 1–11.
- (87) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. General-Purpose Machine Learning Potentials Capturing Nonlocal Charge Transfer. *Acc. Chem. Res.* **2021**, *54*, 808–817.
- (88) Navrotsky, A.; Mazeina, L.; Majzlan, J. Size-Driven Structural and Thermodynamic Complexity in Iron Oxides. *Science* **2008**, *319*, 1635–1638.
- (89) Machala, L.; Tuček, J.; Zbořil, R. Polymorphous Transformations of Nanometric Iron(III) Oxide: A Review. *Chem. Mater.* **2011**, *23*, 3255–3272.
- (90) Sponza, L.; Goniakowski, J.; Noguera, C. Structural, electronic, and spectral properties of six ZnO bulk polymorphs. *Phys. Rev. B* **2015**, *91*, 075126.
- (91) Tsybulya, S. V.; Kryukova, G. N. Nanocrystalline transition aluminas: Nanostructure and features of x-ray powder diffraction patterns of low-temperature Al₂O₃ polymorphs. *Phys. Rev. B* **2008**, *77*, 024112.
- (92) Primo, J. d. O.; Bittencourt, C.; Acosta, S.; Sierra-Castillo, A.; Colomer, J.-F.; Jaeger, S.; Teixeira, V. C.; Anaissi, F. J. Synthesis of Zinc Oxide Nanoparticles by Ecofriendly Routes: Adsorbent for Copper Removal From Wastewater. *Front. Chem.* **2020**, *8*.
- (93) Zeni, C.; Rossi, K.; Glielmo, A.; de Gironcoli, S. Compact atomic descriptors enable accurate predictions via linear models. *J. Chem. Phys.* **2021**, *154*, 224112.
- (94) Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **1993**, *47*, 558–561.
- (95) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.
- (96) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **1996**, *77*, 3865.
- (97) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758–1775.
- (98) Goniakowski, J.; Finocchi, F.; Noguera, C. Polarity of oxide surfaces and nanostructures. *Reports on Progress in Physics* **2007**, *71*, 016501.

- (99) Claeysens, F.; Freeman, C. L.; Allan, N. L.; Sun, Y.; Ashfold, M. N. R.; Harding, J. H. Growth of ZnO thin films—experiment and theory. *J. Mater. Chem.* **2005**, *15*, 139–148.
- (100) Freeman, C. L.; Claeysens, F.; Allan, N. L.; Harding, J. H. Graphitic Nanofilms as Precursors to Wurtzite Films: Theory. *Phys. Rev. Lett.* **2006**, *96*, 066102.
- (101) Goniakowski, J.; Noguera, C.; Giordano, L. Prediction of Uncompensated Polarity in Ultrathin Films. *Phys. Rev. Lett.* **2007**, *98*, 205701.
- (102) Morgan, B. J. Preferential stability of the d-BCT phase in ZnO thin films. *Phys. Rev. B* **2009**, *80*, 174105.
- (103) Demiroglu, I.; Bromley, S. T. Nanofilm versus Bulk Polymorphism in Wurtzite Materials. *Phys. Rev. Lett.* **2013**, *110*, 245501.
- (104) Binks, D. J.; Grimes, R. W. Incorporation of Monovalent Ions in ZnO and Their Influence on Varistor Degradation. *J. Am. Ceram. Soc.* **1993**, *76*, 2370–2372.
- (105) Wang, J.; Xiao, P.; Zhou, M.; Wang, Z. R.; Ke, F. J. Wurtzite-to-tetragonal structure phase transformation and size effect in ZnO nanorods. *J. Appl. Phys.* **2010**, *107*, 023512.
- (106) Lin, K.-H.; Sun, S.-J.; Ju, S.-P.; Tsai, J.-Y.; Chen, H.-T.; Hsieh, J.-Y. Observation of the amorphous zinc oxide recrystalline process by molecular dynamics simulation. *J. Appl. Phys.* **2013**, *113*, 073512.
- (107) Kulkarni, A. J.; Zhou, M.; Ke, F. J. Orientation and size dependence of the elastic properties of zinc oxide nanobelts. *Nanotechnology* **2005**, *16*, 2749–2756.
- (108) Kulkarni, A. J.; Zhou, M.; Sarasamak, K.; Limpijumnong, S. Novel Phase Transformation in ZnO Nanowires under Tensile Loading. *Phys. Rev. Lett.* **2006**, *97*, 105502.
- (109) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
- (110) Menon, S.; Leines, G. D.; Rogal, J. pycscal: A python module for structural analysis of atomic environments. *Journal of Open Source Software* **2019**, *4*, 1824.
- (111) Boattini, E.; Marín-Aguilar, S.; Mitra, S.; Foffi, G.; Smallenburg, F.; Filion, L. Autonomously revealing hidden local structures in supercooled liquids. *Nat. Commun.* **2020**, *11*, 1–9.
- (112) Boattini, E.; Dijkstra, M.; Filion, L. Unsupervised learning for local structure detection in colloidal systems. *J. Chem. Phys.* **2019**, *151*, 154901.
- (113) Pipolo, S.; Salanne, M.; Ferlat, G.; Klotz, S.; Saitta, A. M.; Pietrucci, F. Navigating at Will on the Water Phase Diagram. *Phys. Rev. Lett.* **2017**, *119*, 245701.
- (114) Viñes, F.; Lamiel-Garcia, O.; Illas, F.; Bromley, S. T. Size dependent structural and polymorphic transitions in ZnO: from nanocluster to bulk. *Nanoscale* **2017**, *9*, 10067–10074.
- (115) Erhart, P.; Juslin, N.; Goy, O.; Nordlund, K.; Müller, R.; Albe, K. Analytic bond-order potential for atomistic simulations of zinc oxide. *Journal of Physics: Condensed Matter* **2006**, *18*, 6585–6605.
- (116) Raymand, D.; van Duin, A. C.; Baudin, M.; Hermansson, K. A reactive force field (ReaxFF) for zinc oxide. *Surface Science* **2008**, *602*, 1020–1031.
- (117) Togo, A.; Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **2015**, *108*, 1–5.

- (118) Togo, A.; Oba, F.; Tanaka, I. First-principles calculations of the ferroelastic transition between rutile-type and CaCl₂-type SiO₂ at high pressures. *Phys. Rev. B* **2008**, *78*, 134106.
- (119) Jehannin, M.; Rao, A.; Cölfen, H. New Horizons of Nonclassical Crystallization. *J. Am. Chem. Soc.* **2019**, *141*, 10120–10136.
- (120) Lee, J.; Yang, J.; Kwon, S. G.; Hyeon, T. Nonclassical nucleation and growth of inorganic nanoparticles - Nature Reviews Materials. *Nat. Rev. Mater.* **2016**, *1*, 1–16.
- (121) De Yoreo, J. *Crystallization via Nonclassical Pathways Volume 1: Nucleation, Assembly, Observation & Application*; 2020; Chapter 1, pp 1–17.
- (122) Menon, S.; Lysogorskiy, Y.; Rogal, J.; Drautz, R. Automated free-energy calculation from atomistic simulations. *Phys. Rev. Mater.* **2021**, *5*, 103801.
- (123) Morgan, B. J. Preferential stability of the d-BCT phase in ZnO thin films. *Phys. Rev. B* **2009**, *80*, 174105.
- (124) Demiroglu, I.; Bromley, S. T. Nanofilm versus Bulk Polymorphism in Wurtzite Materials. *Phys. Rev. Lett.* **2013**, *110*, 245501.
- (125) He, M.-R.; Yu, R.; Zhu, J. Reversible Wurtzite–Tetragonal Reconstruction in ZnO(10 $\bar{1}$ 0) Surfaces. *Angew. Chem. Int. Ed.* **2012**, *51*, 7744–7747.
- (126) Manuputty, M. Y.; Dreyer, J. A. H.; Sheng, Y.; Bringley, E. J.; Botero, M. L.; Akroyd, J.; Kraft, M. Polymorphism of nanocrystalline TiO₂ prepared in a stagnation flame: formation of the TiO₂-II phase. *Chem. Sci.* **2019**, *10*, 1342–1350.
- (127) Hanaor, D. A. H.; Sorrell, C. C. Review of the anatase to rutile phase transformation. *J. Mater. Sci.* **2011**, *46*, 855–874.
- (128) Ranade, M. R.; Navrotsky, A.; Zhang, H. Z.; Banfield, J. F.; Elder, S. H.; Zaban, A.; Borse, P. H.; Kulkarni, S. K.; Doran, G. S.; Whitfield, H. J. Energetics of nanocrystalline TiO₂. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6476–6481.
- (129) Luo, Y.; Benali, A.; Shulenburger, L.; Krogel, J. T.; Heinonen, O.; Kent, P. R. C. Phase stability of TiO₂ polymorphs from diffusion Quantum Monte Carlo. *New J. Phys.* **2016**, *18*, 113049.
- (130) Zhang, Y.; Furness, J. W.; Xiao, B.; Sun, J. Subtlety of TiO₂ phase stability: Reliability of the density functional theory predictions and persistence of the self-interaction error. *J. Chem. Phys.* **2019**, *150*, 014105.
- (131) Yang, L.; Zhang, W.; He, L.; Li, H.; Zheng, S. Study on the growth and morphology evolution of titanium oxide clusters in molten iron with molecular dynamics simulation. *RSC Adv.* **2019**, *9*, 32620–32627.
- (132) Mavračić, J.; Mocanu, F. C.; Deringer, V. L.; Csányi, G.; Elliott, S. R. Similarity Between Amorphous and Crystalline Phases: The Case of TiO₂. *J. Phys. Chem. Lett.* **2018**, *9*, 2985–2990.
- (133) Alderman, O. L. G.; Skinner, L. B.; Benmore, C. J.; Tamalonis, A.; Weber, J. K. R. Structure of molten titanium dioxide. *Phys. Rev. B* **2014**, *90*, 094204.

Graphical TOC Entry



Supporting Information:

Non-classical nucleation of zinc oxide from a physically-motivated machine-learning approach.

Jacek Goniakowski,[†] Sarath Menon,[‡] Gaétan Laurens,[¶] and Julien Lam^{*,§}

[†]*CNRS, Sorbonne Université, Institut des NanoSciences de Paris, UMR 7588, 4 Place Jussieu, F-75005 Paris, France*

[‡]*Computational Materials Design, Max-Planck-Institut für Eisenforschung GmbH, 40237 Düsseldorf, Germany*

[¶]*Institut Lumière Matière, UMR5306 Université Lyon 1-CNRS, Université de Lyon, 69622 Villeurbanne Cedex, France*

[§]*Centre d'élaboration des Matériaux et d'Etudes Structurales, CNRS (UPR 8011), 29 rue Jeanne Marvig, 31055 Toulouse Cedex 4, France*

E-mail: julien.lam@cns.fr

Table S1: Calculated (present PW91 and reference HSE03¹) characteristics of the six bulk ZnO polymorphs: lattice parameters a and c (Å), and cohesion energy differences with respect to ground state (WRZ) structure ΔE_{coh} (eV/ZnO).

	PW91		HSE03	
	a, c	ΔE_{coh}	a, c	ΔE_{coh}
WRZ	3.28, 5.30	-0.00	3.25, 5.24	-0.00
ZBL	3.27, 3.27	-0.01	3.24, 3.24	-0.03
BCT	5.61, 3.29	-0.05	5.55, 3.25	-0.05
SOD	5.67, 5.67	-0.14	5.61, 5.61	-0.14
HBN	3.46, 4.59	-0.14	3.43, 4.50	-0.14
CUB	6.28, 6.28	-0.23	6.22, 6.22	-0.24

Table S2: Calculated (PW91) surface energies of low index surfaces of the six ZnO polymorphs E_{surf} (J/m²).

	orientation	E_{surf}
WRZ	(10-10)	0.89
WRZ	(11-20)	0.93
WRZ	(0001)	1.30
ZBL	(110)	0.92
BCT	(010)	0.49
BCT	(001)	1.14
SOD	(001)	1.00
HBN	(0001)	0.36
CUB	(001)	0.48

References

- (1) Sponza, L.; Goniakowski, J.; Noguera, C. Structural, electronic, and spectral properties of six ZnO bulk polymorphs. *Phys. Rev. B* **2015**, *91*, 075126.

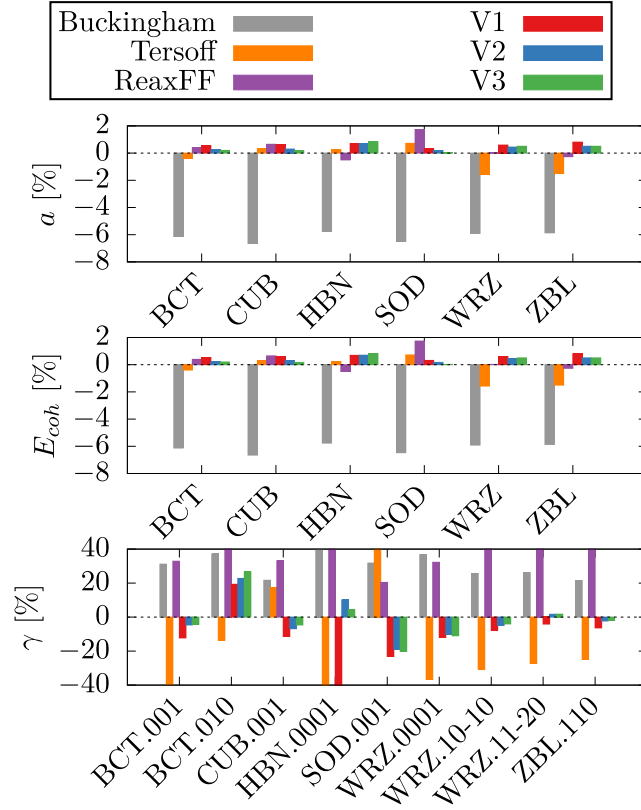


Figure S1: Signed errors (%) with respect to the DFT values of bulk lattice parameter a , cohesion energy E_{coh} relative to the wurtzite polymorph, and the relative surface energy γ determined from the relaxed structures obtained with the three PLIPs as well as with the three classical potentials.

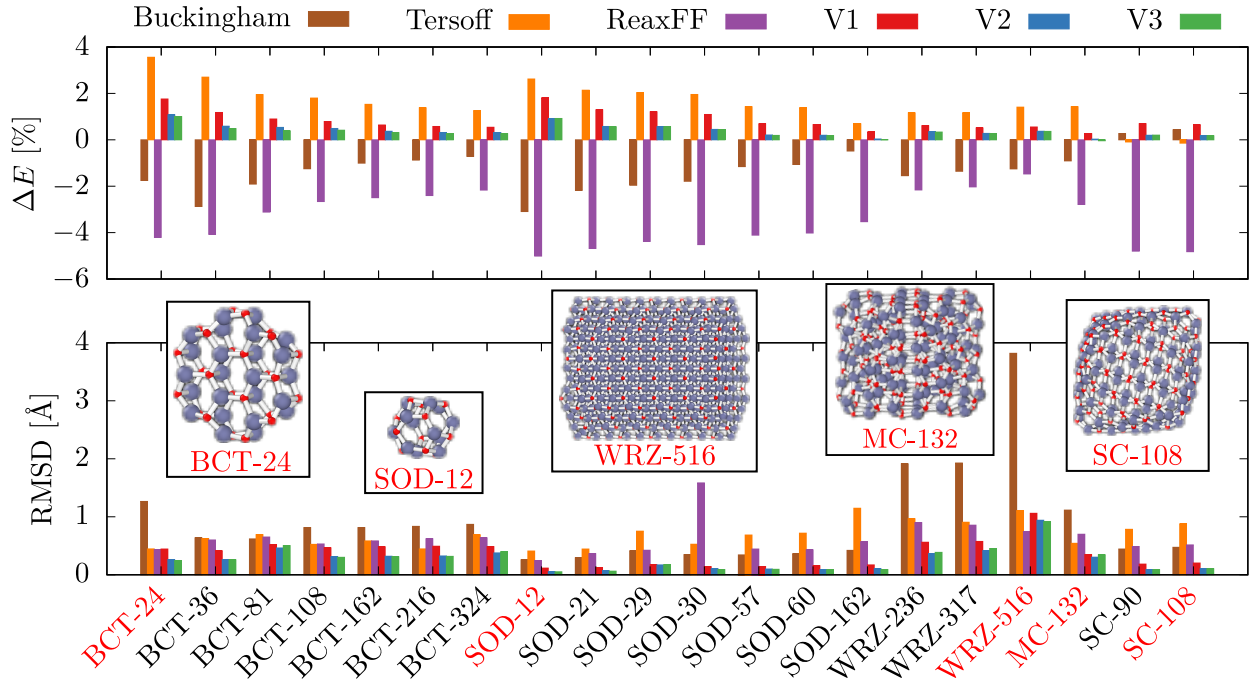


Figure S2: (Top panel) Relative signed error on the clusters's energy per unit of ZnO relative to the Wurtzite bulk energy and evaluated compared to DFT results. (Bottom panel) RMSD based on atomic positions of the studied structures. Snapshots representing some of those. Results were obtained with the PLIPs and the classical potentials.

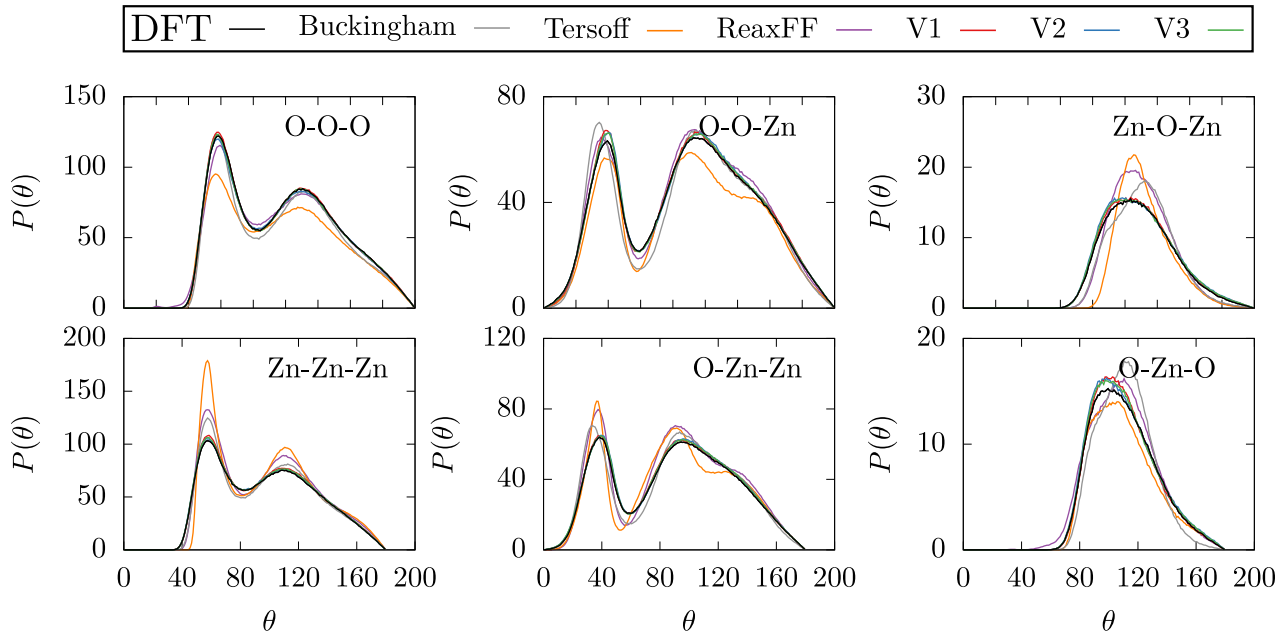


Figure S3: Bond angle distribution (BAD) $P(\theta)$ of the liquid obtained at 1500 K. Structures calculated with the three PLIPs are compared with those computed using DFT and the three potentials.

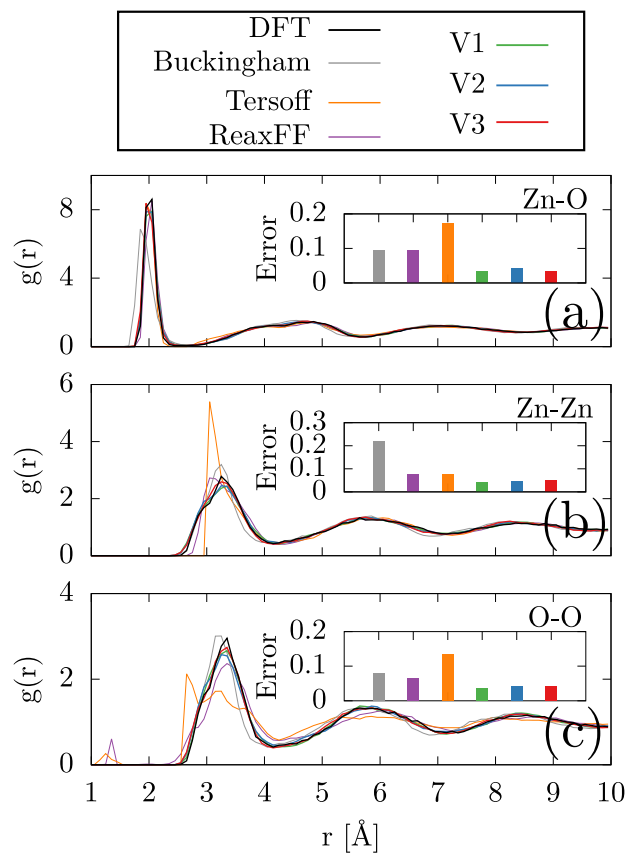


Figure S4: Radial distribution functions (RDF) $g(r)$ of final structures obtained after cooling a hot liquid. Structures calculated with the three PLIPs are compared with those computed using DFT and the three potentials. BAD deviations with respect to those of DFT are presented in insets.

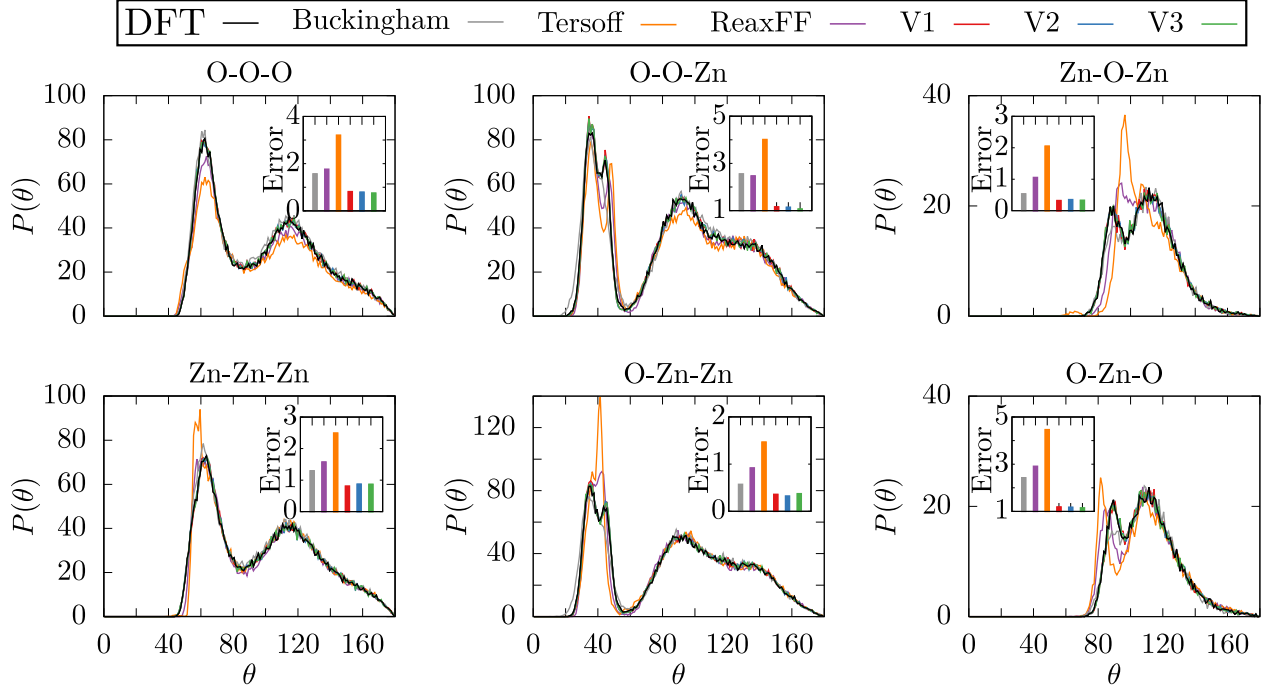


Figure S5: Bond angle distribution (BAD) $P(\theta)$ of final structures obtained after cooling a hot liquid. Structures calculated with the three PLIPs are compared with those computed using DFT and the three potentials. BAD deviations with respect to those of DFT are presented in insets.

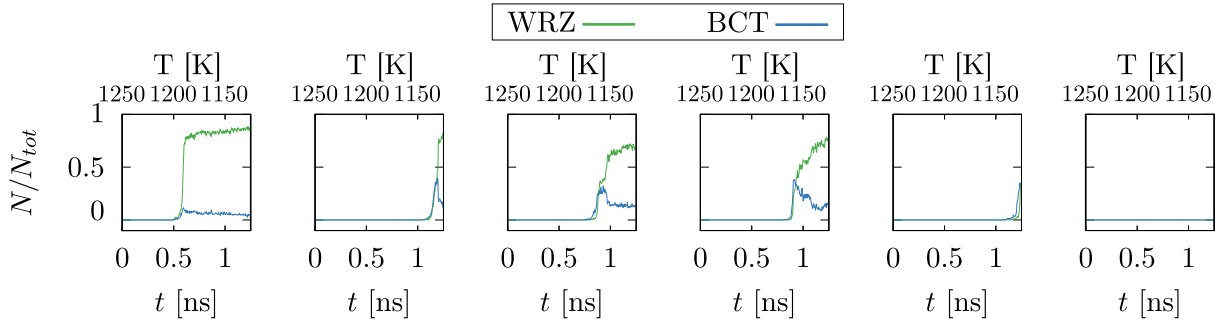


Figure S6: 5 independent MD runs using 2000 atoms and a cooling rate equal to 25 K/ns (in comparison with a 50 K/ns cooling rate presented in the main article). MD runs are analyzed in terms of number of atoms detected in WRZ or BCT phases, according to the likelihood parameter S , along the cooling path, *i.e.* with respect to time and temperature.

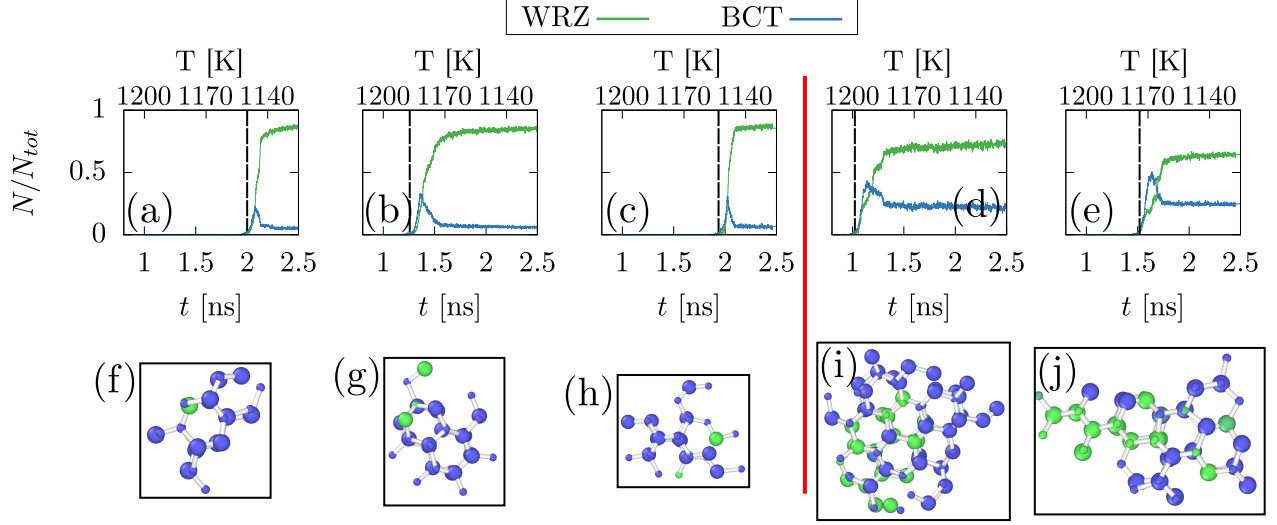


Figure S7: (a-e) Number of atoms detected in WRZ or BCT phases, according to the likelihood parameter S , along the cooling path, *i.e.* with respect to time and temperature, for 5 independent MD runs using 4000 atoms and a cooling rate equal to 50 K/ns. (f-j) Snapshots of intermediate structures obtained during the nucleation at the time highlighted by black dashed lines in the corresponding temporal evolution. Atoms identified as WRZ and BCT are colored in green and blue, respectively. The red solid line discriminates between the two observed nucleation processes.

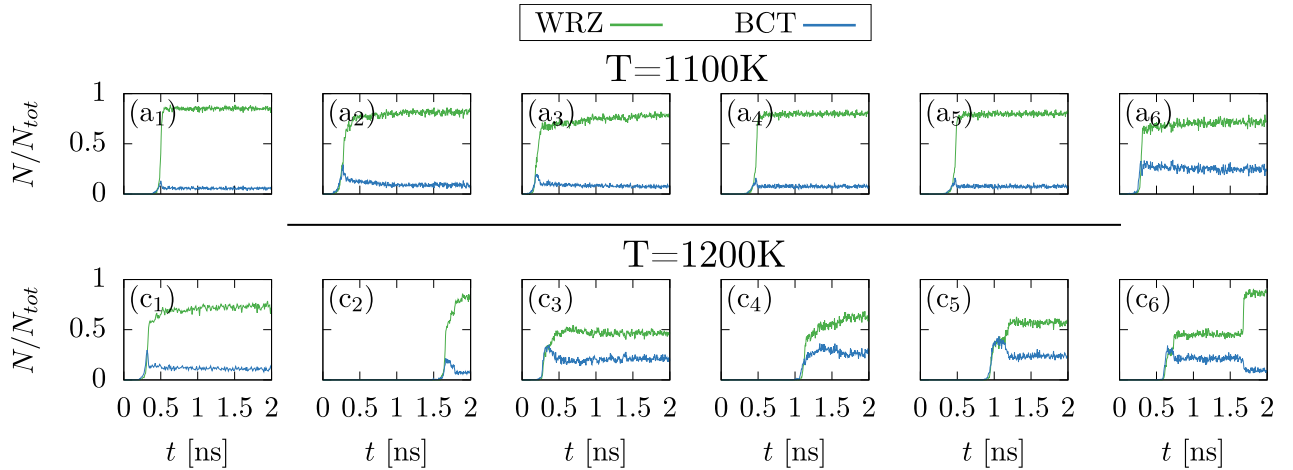


Figure S8: (a-e) Temporal evolution of the number of atoms detected in WRZ or BCT phases, according to the likelihood parameter S , for 6 independent MD runs using 2000 atoms at fixed temperatures.