

Development of a Machine Learning Interatomic Potential for Exploring Pressure-Dependent Kinetics of Phase Transitions in Germanium

A. Fantasia,^{1, a)} F. Rovaris,¹ O. Abou El Kheir,¹ A. Marzegalli,¹ D. Lanzoni,¹ L. Pessina,¹ P. Xiao,² C. Zhou,³ L. Li,³ G. Henkelman,⁴ E. Scalise,¹ and F. Montalenti^{1, b)}

¹⁾*Department of Materials Science, University of Milano-Bicocca, 20125 Milano, (Italy)*

²⁾*Dept. of Physics & Atmospheric Science, Dalhousie University, 1453 Lord Dalhousie Drive, B3H 4R2, Halifax, NS (Canada)*

³⁾*Dept. of Materials Science and Engineering, Southern University of Science and Technology, 1088 Xueyuan Avenue, 518055, Shenzhen, (China)*

⁴⁾*Dept. of Chemistry, The University of Texas at Austin, 105 East 24th Street STOP A5300, 78712, Austin, TX (USA)*

(Dated: 29 September 2024)

We introduce a data-driven potential aimed at the investigation of pressure-dependent phase transitions in bulk germanium, including the estimate of kinetic barriers. This is achieved by suitably building a database including several configurations along minimum energy paths, as computed using the solid-state nudged elastic band method. After training the model based on density functional theory (DFT)-computed energies, forces, and stresses, we provide validation and rigorously test the potential on unexplored paths. The resulting agreement with DFT calculations is remarkable in a wide range of pressures. The potential is exploited in large-scale NPT simulations, displaying local nucleation in the R8 to β -Sn pressure-induced phase transformation, here taken as an illustrative example.

^{a)}Electronic mail: a.fantasia1@campus.unimib.it

^{b)}Electronic mail: francesco.montalenti@unimib.it

I. INTRODUCTION

In recent years, there has been a remarkable transformation in how we approach atomistic simulations, and at the forefront of this change is the increase in the popularity of machine learning interatomic potentials (MLIPs). This innovative approach offers a way to significantly reduce the computational cost of ab initio calculations, typically based on density functional theory (DFT), with a great accuracy/efficiency trade-off allowing us to study larger systems and/or longer time-scale processes. The remarkable efficiency and precision of MLIPs have made them essential tools, gaining widespread recognition and adoption for exploring various atomistic systems¹⁻¹¹

Our objective in this work is the development of an interatomic potential suitable for investigating crystal phase transitions, taking germanium as a particularly interesting system due to its potential applications¹². The emphasis lies in showcasing and testing the developed potential, rather than conducting a comprehensive study of phase transitions in Ge. It's important to note that this endeavor necessitates the creation of an MLIP capable of accurately representing pressure-dependent kinetic barriers. Merely training the MLIP relying solely on configurations close to equilibrium and snapshots from molecular dynamics near metastable crystal structures would not provide adequate insight into transition mechanisms. This limitation arises because crucial saddle points defining kinetic barriers would be under-represented or excluded from the training dataset, leading to unreliable predictions. To address this, we incorporated into our dataset configurations along minimum energy paths obtained through solid-state nudged elastic band (ssNEB) calculations under varying stress conditions, to provide insight into transition barriers.

Empirical potentials, including Stillinger-Weber^{13,14}, Tersoff¹⁵, and modified-embedded-atom-method (MEAM)^{16,17}, have been developed to significantly reduce computational costs compared to DFT calculations, but may be unreliable when dealing with metastable phases¹⁸ and, more generally, with kinetics.

The interesting electronic properties of germanium, such as a high intrinsic electron mobility, make it useful for novel high-speed electronic and photonic devices. Additionally, its narrow band gap enables efficient infrared detection in optoelectronic devices, rendering Ge a valuable material for advancing modern technology. Ge is a well-known semiconductor that typically exhibits a cubic diamond (CD) structure (space group $Fd\bar{3}m1$) at ambient pressure and temperature. Through anvil press and nanoindentation experiments, CD transforms into a metallic phase with the β -Sn structure ($I4_1/amd$) upon non-hydrostatic pressure at around 10 GPa¹⁹⁻²¹, showing re-

markable stability over a wide range of pressures. Upon increasing pressure, β -Sn undergoes a series of transitions into other high-density metallic phases including simple hexagonal (SH) and HCP^{22,23}. Upon unloading from β -Sn, CD is not recovered and different metastable allotropes are reached instead. Experiments shown the presence of metastable BC8 (Ia $\bar{3}$), ST12 (P4 $\bar{3}$ 2 $\bar{1}$ 2 $\bar{2}$), and R8 (R $\bar{3}$) phases, and led to believe that the formation of such allotropes depends on stress conditions and rate of decompression from β -Sn^{24–28}. Sustained interest in ST12 stems from its optical properties²⁹, with recent advancements leading to the synthesis of a large pure bulk sample³⁰, the production of ST12 nanoparticles^{31,32} and also the fabrication of nanowires³³. BC8 and R8 are observed to transform to a hexagonal diamond (HD) phase under ambient pressure and temperature^{34,35}. This is of great technological interest given HD’s predicted narrow direct band gap, rendering it well-suited for optoelectronic applications compatible with silicon¹². It is also noteworthy that amorphous Ge (a-Ge) has been extensively documented in various experiments: observed as a byproduct during the unloading from β -Sn, as a pathway to achieve β -Sn through loading, and also through observed transitions from a low-density to a high-density form of a-Ge^{36–38}.

II. METHODS

A. Ab initio calculations and simulation details

The Vienna Ab Initio Simulation Package (VASP)^{39,40} was used to perform all DFT calculations. Ge $3d4s4p$ electrons were treated as valence electrons by the projector augmented-wave (PAW) method⁴¹. The Perdew-Burke-Ernzerhof (PBE)⁴² was used as exchange-correlation functional as it is a popular choice given its reasonable accuracy over a wide range of systems. Since metallic phases of Ge are demanding in terms of convergence, we conducted convergence tests to find the optimal balance between accuracy and efficiency. The truncation energy of the plane wave basis was set to 500 eV, and the k-point grid spacing was fixed at 0.14 \AA^{-1} . Formation energies were computed by taking the energy per atom difference between a given structurally optimized crystal phase and the most stable one, the CD phase. Specifically, using the selected k-point grid spacing, we observed that the formation energies for the 4H-HD, 2H-HD, BC8, ST12, and R8 phases have already reached convergence, showing variations of less than 1 meV/atom upon further parameter refinement. However, for denser metal structures like β -Sn, Fmmm, SH, HCP,

BCC, and FCC, there are still oscillations, with some reaching up to approximately 5 meV/atom in their computed energies. We still consider this result as satisfactory since different DFT functionals would predict formation energies with even larger discrepancies. Since we are interested in pressure-dependent phase transitions we also monitored the convergence in terms of pressure. Using the selected k-point grid spacing of 0.14 \AA^{-1} we observed that semiconducting phases have already reached a convergence of about 0.01 GPa. Some denser metallic structures, more demanding in terms of convergence, still present oscillations, with some limited examples reaching ≈ 1 GPa.

Molecular dynamics (MD) simulations were carried out in LAMMPS⁴³ using the DeePMD extension⁴⁴. To sample data, both the canonical (NVT) and the isothermal-isobaric (NPT) ensemble were used with an integrator time step of 1 fs while the temperature and pressure were set using a Nose-Hoover thermostat and barostat⁴⁵ with a relaxation time of 0.5 ps and 5.0 ps respectively.

To perform local structural optimizations, the Fast Inertial Relaxation Engine (FIRE) optimizer⁴⁶ within the Atomic Simulation Environment (ASE)⁴⁷ was employed by setting up the trained MLIP as a calculator to obtain potential energies, forces, and stresses.

When investigating transitions, a critical parameter of focus is the activation energy barrier associated with the reaction mechanism. This is significant because, following transition state theory⁴⁸, the transition rate is directly proportional to the exponential of the free energy barrier. Many different methods have been developed to accurately locate such transition states^{49–55}, but only a few take into account variable cell transformations. In this work, a generalized solid-state nudged elastic band (ssNEB) method⁵⁶ with climbing-image^{57,58} was employed for determining reaction pathways of crystal phase transitions involving both atomic and cell degrees of freedom. We utilized this well-established technique by means of its implementation for the ASE package in the TSASE code⁵⁹.

B. Dataset Generation

In order to build the MLIP for Ge, a reference dataset for the training process is generated using an iterative training procedure, exploiting the concept of active learning^{60–63}.

To thoroughly explore the kinetics of solid-solid phase transitions, we integrated configurations explored by performing ssNEB calculations into our dataset. This inclusion is pivotal, as relying solely on sampling configurations near metastable crystal structures would not be sufficient for

studying transition barriers.

The outlined procedure, schematically illustrated in Fig. 1, can be summarized as follows: We selected many different Ge-crystalline phases from the Materials Project (MP) database⁶⁴. These include the CD (mp-32), 4H-HD (mp-1091415), 2H-HD (mp-1007760), BC8 (mp-1080106), ST12 (mp-137), R8 (mp-128), β -Sn (mp-78), SH (mp-1224349), Imma (mp-1061054), Fmmm

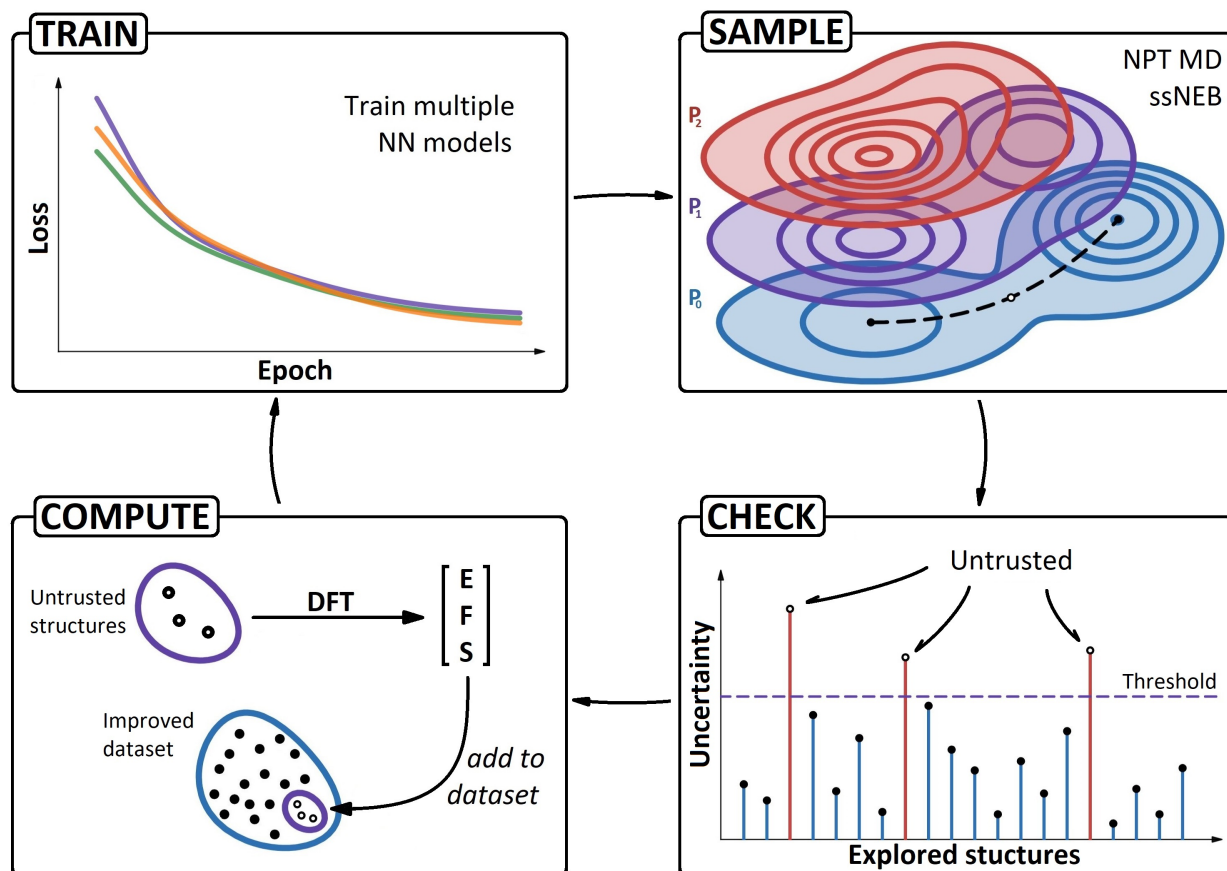


FIG. 1. Outline of the active learning approach utilized in this work. The procedure involves a cycle in which multiple NN models are trained. A model is tasked with sampling numerous configurations via NPT MD and ssNEB under varying stress conditions. A visual representation illustrating how the enthalpy landscape may shift under applied pressure is depicted in the top-right panel. The consistency in predicting energies (E), forces (F), and stresses (S) for these sampled structures is checked by comparison of the previously trained models' predictions. Configurations exhibiting significant deviations are identified as unreliable and subjected to DFT single-point calculations. Subsequently, these configurations are incorporated into the dataset, and the models undergo further refinement through additional training iterations, thus completing the cycle.

(mp-148), HCP (mp-1008733), BCC (mp-998883), FCC (mp-12093). After replicating these cells, the lattice parameters and the atomic positions are perturbed to obtain many different configurations. A starting incomplete dataset is built based on these structures that are labeled with energies, forces, and stresses computed through DFT calculations. An initial, coarse model trained on this dataset is used to explore more configurations. This is done by sampling snapshots from NVT and NPT MD, and intermediate images from ssNEBs. The energies, forces, and stresses of these structures are computed again through single-point DFT calculations and added to the existing starting dataset.

Multiple Neural Network (NN) models are then trained using identical training sets but with different random seeds. Consequently, these models would lead to diverse predictions for the same configuration. If predictions are consistent among the different models, it would suggest that the configuration aligns closely with the training set; conversely, discrepancies indicate unreliability in predictions. By employing this approach, we can select configurations exhibiting unreliable predicted energies, forces, and stresses, compute these observables again through single-point DFT calculations, and then iteratively expand the training set by incorporating such data deemed most relevant.

The sequential addition of inaccurately estimated configurations into the training dataset during iterations is a crucial aspect of constructing the MLIP.

The final dataset, upon which the actual model is trained, is composed of roughly 2700 structures making up a total of roughly 112 thousand atomic environments, and it's split into 87% training and 13% validation sets. The split was done randomly, except for certain structures we specifically included in the training set: phases at their minimum energy configuration and with lattice parameter perturbations close to the values corresponding to this minimum. Since many crystal-phase transitions are pressure-induced, structures covering a wide range of pressure have been sampled, reaching almost 100 GPa for some configurations. Most of the sampling was conducted within a pressure range of 5 GPa tensile to 30 GPa compressive strain. This range was chosen because the most technologically relevant crystal phase transformations in Ge, such as the formation of Hexagonal Ge and ST12, occur after loading of CD to β -Sn and subsequent unloading; these experiments are typically performed below 20 GPa. Structures containing interstitial atoms and vacancies were obtained by sampling MD starting from manually crafted configurations. Disordered structures were obtained by heating and then cooling some given crystal phases to get highly defective and amorphous-like configurations. These highly varied structures allow

Structure type	No. structures	No. environments
Cubic Diam.	253	9312
Hex. Diam. (4H)	158	6528
Hex. Diam. (2H)	185	8016
BC8	173	6000
ST12	274	8508
R8	177	6408
β -Sn	169	5648
Simple Hex.	150	4808
Other phases	151	7280
Interstitials	62	4092
Vacancies	63	3906
Disordered	278	26640
Transitions	597	14672
TOTAL	~ 2700	~ 112000

TABLE I. Summary of the dataset for the germanium model. The first column shows the number of structures in the database, while the second one shows the number of atoms (and, therefore, atomic environments) in the database for each structure type.

for a more comprehensive sampling of possible atomic environments, leading to a more robust interatomic potential. Transition structures were obtained by sampling intermediate images from ssNEB calculations between different crystal phases over a wide range of pressures. Also, after replicating these intermediate images, the lattice parameters and the atomic positions are perturbed to sample more different configurations close to the transition pathway found. This simple yet effective procedure is an alternative to other methods, such as Transition Path Sampling⁶⁵, Metadynamics⁶⁶, Umbrella Sampling⁶⁷, or many more recent ones⁶⁸⁻⁷², employed for an extensive sampling of the configuration space close to transition states and along a reaction coordinate.

The dataset resulting from our iterative procedure is detailed in Tab. I.

C. Machine Learning Model

Utilizing our dataset, the NN potential was trained using the Deep Potential Molecular Dynamic package (DeePMD-kit)^{73,74}. The deep potential (DP) model decomposes the total energy of the system into atomic contributions by employing a local geometric description of the atomic environment. In this approach, a multilayer NN is used as a regression model to describe the relationship between atomic configurations and energy. For each local system, an embedding network converts the atomic coordinate information into a descriptor matrix which is fed into a fitting network mapping the matrix to the local atomic energy. An in-depth explanation of the NN architecture is given in the original paper⁷³.

To accurately capture the structural information of many Ge crystal phases and different configurations, we used two types of descriptors constructed from all information (both angular and radial) in the so-called DeepPot-SE framework⁷⁴: a “se_e3” type embedding which takes angles between two neighboring atoms as input, and a “se_e2_a” type embedding which takes the distance between atoms as input. For the former descriptor, we used a cutoff value of $r_c = 3.30 \text{ \AA}$, a smoothing cutoff $r_s = 2.00 \text{ \AA}$, and a three-layer embedding network containing 8, 16, and 32 neurons. For the latter, $r_c = 6.60 \text{ \AA}$, $r_s = 3.30 \text{ \AA}$ were used along with a three-layer embedding network containing 16, 32, and 64 neurons with 32 axis neurons. For the fitting net, three hidden layers with 160, 120, and 80 neurons in each layer were employed. We trained and tested multiple models with various architectures and hyperparameters to empirically identify optimal values that balance accuracy and efficiency.

The family of loss functions \mathcal{L} of the model, minimized by using an Adam stochastic gradient descent optimizer⁷⁵ with an exponentially decreasing learning rate, takes into account energies, forces, and virials by

$$\mathcal{L} = \frac{p_e}{N} \Delta E^2 + \frac{p_f}{3N} \sum_i^N |\Delta \mathbf{F}_i|^2 + \frac{p_\xi}{9N} \|\Delta \Xi\|^2 \quad (1)$$

Here ΔE , $\Delta \mathbf{F}_i$ and $\Delta \Xi$ represent the difference in energy, force (on atom i) and virial, respectively, between the training data and the NNP prediction; N is the number of atoms; p_e , p_f , and p_ξ are tunable weights. Eq. (1) reports the loss contribution due to a single configuration in the training set, the total loss will be the sum over all the configurations. As suggested in the original paper⁷³ to reduce the total training time, we progressively increased p_e and p_ξ and decreased p_f during the training procedure, so that the force term dominates at the beginning, while energy and virial terms become important at the end. We decided to use also forces and virials along with energies in

the training process to significantly reduce the number of reference data needed to train accurate and robust models⁷⁴. Furthermore, a model that accurately predicts stresses is crucial for the investigation of pressure-induced crystal phase transitions.

Fig. 2 illustrates the regression plots for energies, forces, and stresses concerning both the training and validation datasets. Our model demonstrates satisfactory performance, with root mean square errors (RMSEs) of approximately 5 meV for energies and 0.1 eV/Å for forces. Stress prediction is also robust, exhibiting an RMSE of less than 0.5 GPa over a broad spectrum spanning approximately 100 GPa.

III. RESULTS AND DISCUSSION

A. Benchmarking of the model

To assess the reliability of the model, a diverse test set consisting of snapshots from MD at different temperatures and pressures was also prepared. Additionally, the main results concerning transition states are thoroughly discussed in the dedicated section III B. In detail, the test set comprises 125 diamond structures (CD, 4H-HD, and 2H-HD), corresponding to 7888 atomic environments, 74 low-density semiconductor structures (BC8, ST12, and R8) (3984 atomic environments), 97 high-density metal structures (6144 atomic environments), and 146 disordered or defective structures, encompassing amorphous configurations as well as those with interstitial atoms or vacancies (13401 atomic environments). Even on this test set, our model exhibits strong predictive capability, showing minimal deviations in the regression plots of Fig. 2.

Further evaluation focused on the prediction of formation energies. To determine the DFT formation energies, we computed the energy difference per atom between a specific structurally optimized crystalline configuration and the cubic diamond reference. Likewise, for estimating the DP formation energies, we followed the same procedure, but utilizing structures optimized for the DP. The comparison of the results obtained from such calculations is depicted in Fig. 3 and Tab. II.

The model performs remarkably well, with errors smaller than 5 meV for all the tested structures, except for the BCC phase. As expected, the predictive ability of the model on the 4H-HD, 2H-HD, BC8, ST12, R8 phases is slightly better than the other higher energy ones. Particularly noteworthy is the accurate prediction of the 6H-HD phase, despite its absence in the training set. As an intermediate phase positioned between CD and 4H-HD and characterized by distinct stack-

ing faults, the inclusion of CD, 4H-HD, and 2H-HD in the training set seems to be enough to adequately encapsulate the local atomic environments of the 6H-HD structure.

Additionally, we analyzed the accuracy of energy vs. volume curves, as shown in Fig. 4, which provide insights into the volumetric compression-expansion behavior of different phases. While the model performs admirably for diamond phases and the other semiconductor structures BC8, ST12, and R8, it exhibits slightly lower accuracy for metallic phases, especially in the volumetric expansion region. This is because, apart from the aforementioned issue concerning convergence, the training set emphasized sampling of these phases under pressure conditions rather than volu-

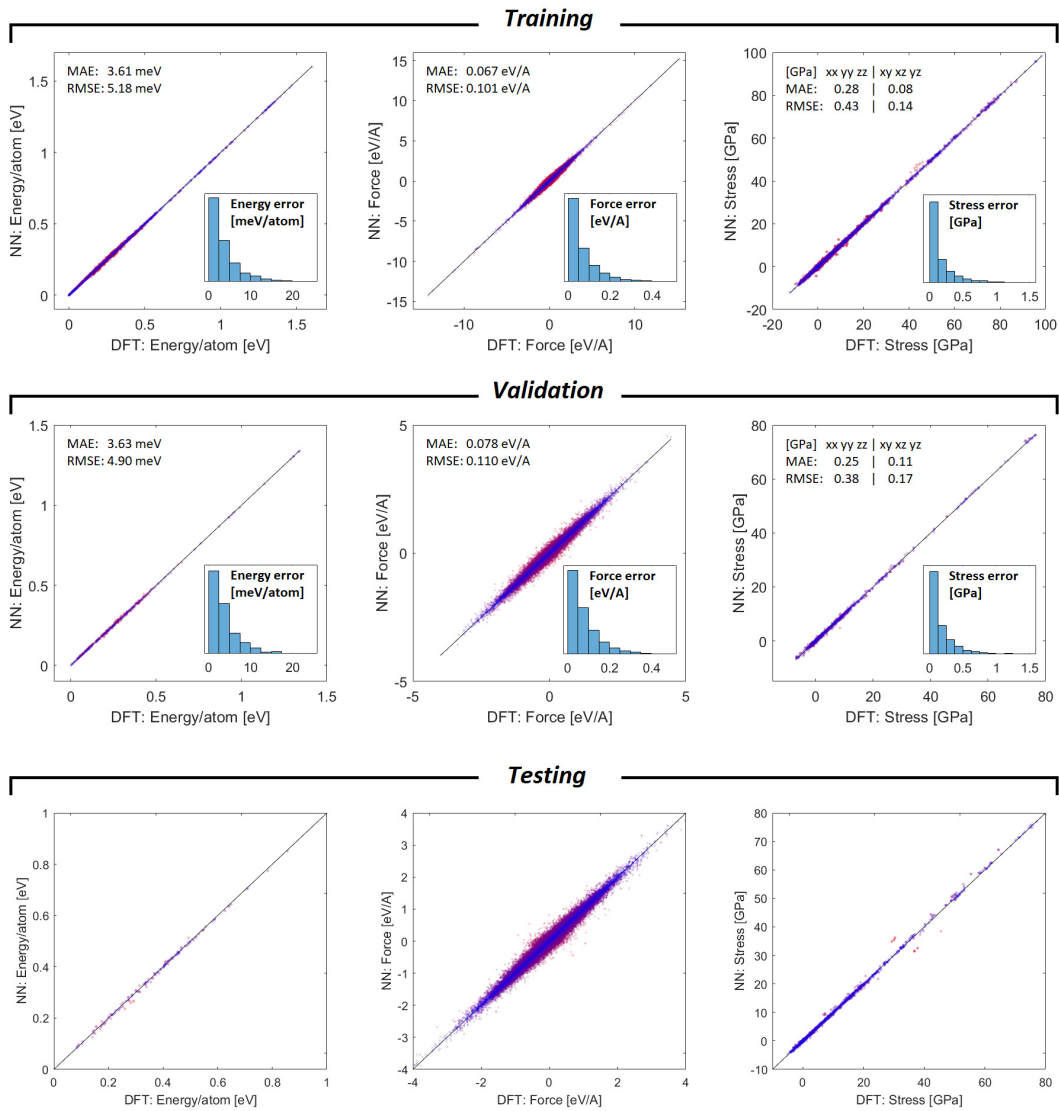


FIG. 2. Regression plots illustrating correlations of energies (left), forces (middle), and stresses (right) calculated by the trained deep neural network (NN) potential with corresponding DFT values.

metric dilation, and it is generally understood that NN models are interpolative, providing reliable results only within their training domain. This decision was made because these phases are only competitive under high pressures, rendering volumetric dilation less pertinent from a physical standpoint.

Three additional phases, which were not part of the training set, are also tested: Pbam, $P4_2/nm$, and $P4_12_12$. These structures, recently predicted theoretically, display significantly lower energies compared to the majority of metastable polymorphs observed in Ge. For this reason, there is speculation that these phases might emerge as metastable forms in experiments utilizing indentation or anvil cells^{76,77}. The comparison presented in Tab. II, between formation energies computed via DFT and those predicted by the DP, reveals a larger error for such crystal phases absent from the training set. To address this, any investigation aimed at these structures would necessitate their inclusion in the training set as well.

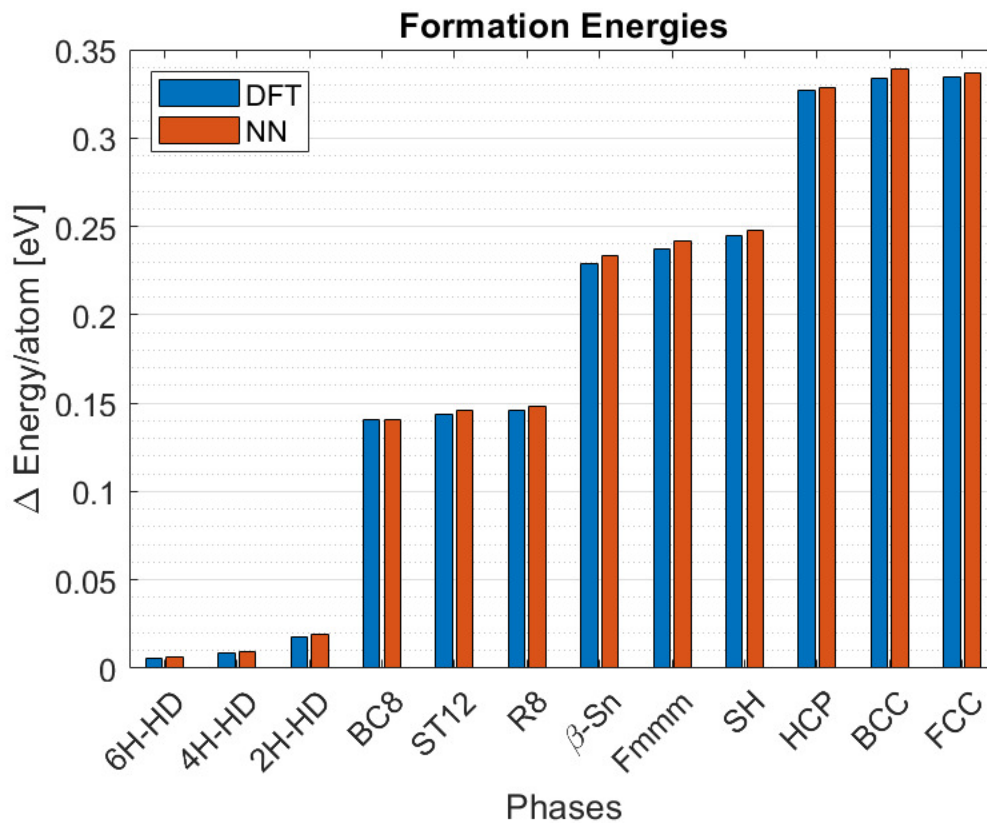


FIG. 3. Formation energies for the different crystalline phases of Ge computed both through DFT (in blue) and the developed MLIP (in red) are compared.

B. Transition Pathways

Crystal phase transitions in germanium were extensively investigated utilizing ssNEB calculations and NPT MD. Our research revealed several pathways between different crystal phases, helping to understand the phase transition mechanism and kinetics. As our objective in this paper is to showcase and assess the developed potential, rather than conduct an exhaustive study of phase

Phase	ϵ_{DFT} [meV]	ϵ_{DP} [meV]	$\Delta\epsilon$ [meV]
6H-HD	5.7	6.2	0.5
4H-HD	8.6	9.5	0.9
2H-HD	18.0	19.4	1.4
<hr/>			
BC8	140.5	140.7	0.2
ST12	143.7	145.9	2.2
R8	146.3	148.4	2.1
<hr/>			
β -Sn	229.3	233.2	3.9
Fmmm	237.4	241.5	4.1
SH	244.8	247.6	2.8
<hr/>			
HCP	326.8	328.8	2.0
BCC	333.6	339.0	5.4
FCC	334.6	336.9	2.3
<hr/>			
Pbam	31.8	56.7	24.9
P4 ₂ /ncm	35.8	32.7	-2.9
P4 ₁ 2 ₁ 2	38.7	73.9	35.2

TABLE II. Formation energies (ϵ) and absolute error ($\Delta\epsilon$) computed by DFT and the DP model for different crystalline phases of Ge. Three configurations not present in the training-set are included in the table and separated by a horizontal line.

transitions in Ge, here below we delve into a few illustrative examples.

Firstly, a collective transformation involving a 16-atom cell was observed, connecting the CD and BC8 phases. Multiple cell geometries have been selected as the starting and ending points for the ssNEB calculation. We chose the geometries that required the least amount of atomic displacement and bond breaking. The minimum energy path for this transition, illustrated in Fig. 5a, reveals a relatively high predicted activation energy of approximately 0.28 eV/atom. It is indeed commonly acknowledged that the reaction barrier during solid phase transition can often be significant, as numerous chemical bonds are simultaneously formed or broken as the crystal transitions from one phase to another.

Similarly, another pathway connecting the CD and β -Sn phases was identified, showing a remarkable agreement between ab initio and NN prediction, as depicted in Fig. 5b. This transfor-

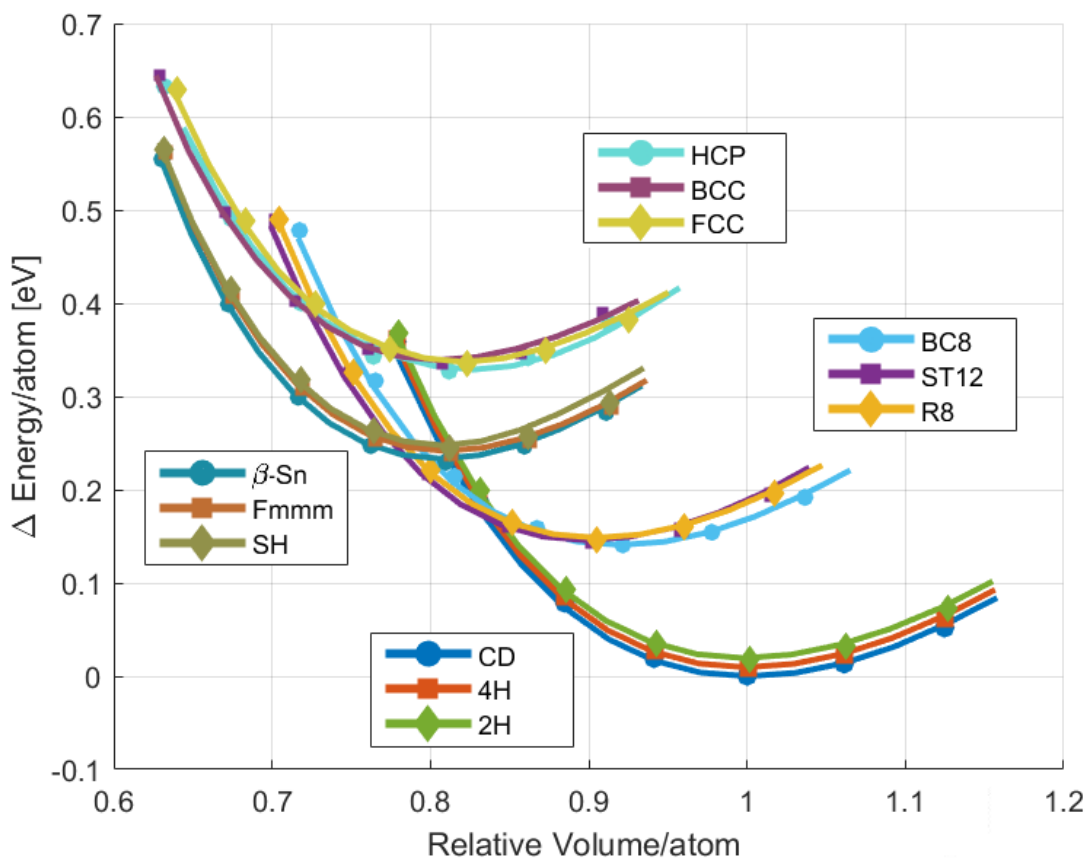


FIG. 4. The relationship between energy and volume per atom across different crystalline phases of Ge is depicted. The lines represent predictions from the MLIP, while the markers represent calculations from DFT calculations.

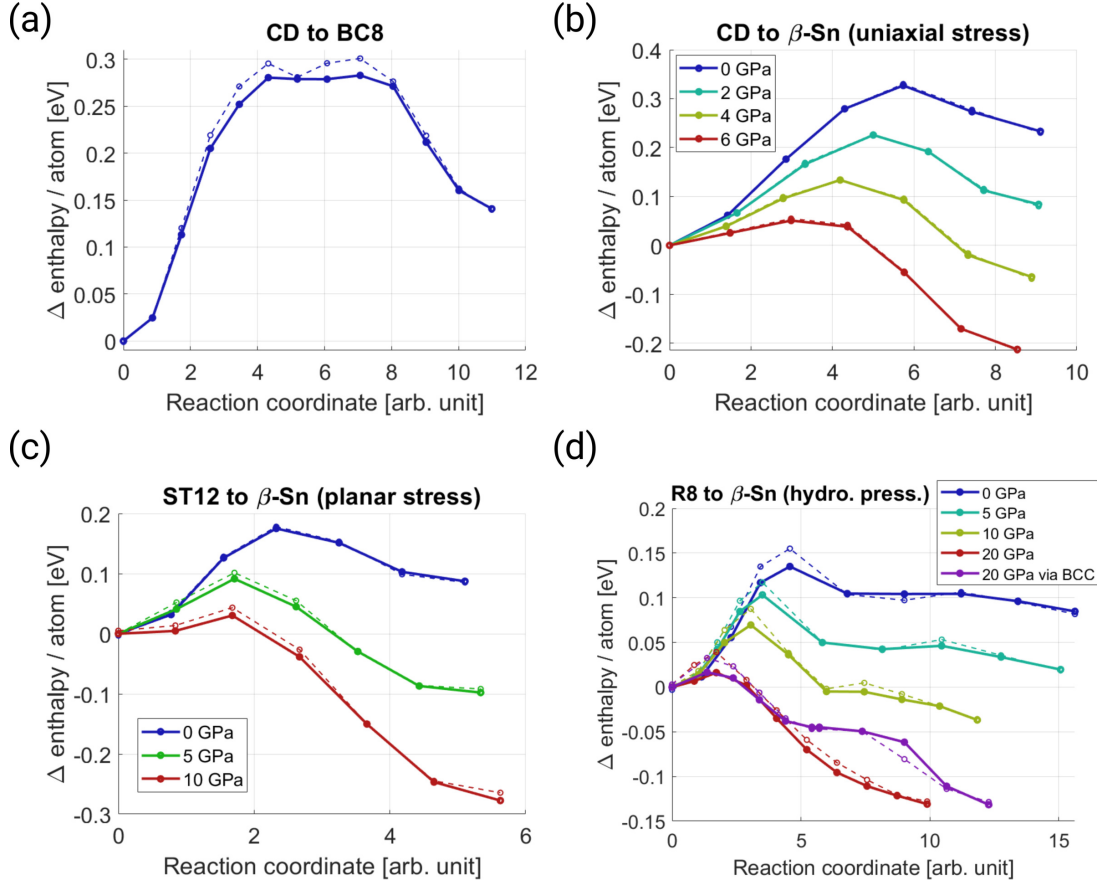


FIG. 5. Minimum energy paths of the phase transitions: (a) CD to BC8, (b) CD to β -Sn under uniaxial stress, (c) ST12 to β -Sn under planar stress, and (d) R8 to β -Sn under hydrostatic pressure. A comparison between the ssNEBs calculations performed with the DP model (solid lines and filled circles) and single-point DFT calculations of such images (dashed lines and unfilled circles) is shown.

mation is quite easy to identify because the relative positions of atoms in the cell remain constant during the transformation, only the lattice parameters change. This pathway, previously reported for Silicon⁷⁸, occurs similarly in Ge. The activation barrier, estimated to be approximately 0.32 eV/atom, presents a significant obstacle. However, the transformation is not hindered because stress can be applied to modify the MEP and the transition energy barrier, which directly affects the rate of phase transition. When subjected to external stress, the energy landscape is transformed into an enthalpy landscape, accounting for the work done by external stresses. To explore this phenomenon, we conducted multiple calculations using the ssNEB method with various levels of uniaxial stress. Through this procedure, we determined that at approximately 7 GPa of compressive uniaxial stress, the transition is expected to be barrierless. To our knowledge such an

agreement between ab initio and NN on the calculation of pressure-dependent energy barriers is unprecedented.

NPT MD simulations were employed to investigate such pressure-induced transformation. To examine the structural transformation, we utilized a simulation cell measuring approximately $56 \times 56 \times 76 \text{ \AA}$, containing 10368 atoms. The temperature was set at 300 K (room temperature), with uniaxial compressive stress along the z -direction at 7 GPa. Snapshots from such simulated transition are shown in Fig. 6a.

To ensure confidence in the model’s accuracy, we calculated the deviation in the predicted force for each atom at each step using an ensemble of three neural networks. We then computed the mean deviation at each step and recorded the maximum deviation value among all atoms. As shown in Fig. 6b, force predictions are considered satisfactory, with mean and maximum deviations below 0.05 eV/\AA and 0.15 eV/\AA , respectively, even at the transition midpoint where deviations are higher.

We then identified a pathway connecting the ST12 and β -Sn phases, illustrated by its corresponding MEP in Fig. 5(c). This transformation pathway, involving a 12-atom cell, aligns with a previously proposed atomistic mechanism for such transition⁸⁰. The activation energy barrier here is approximately 0.18 eV/atom , but it can be reduced by applying stress. Specifically, we observed that the application of compressive planar stress, in this case, effectively reduces the barrier, leading to a transition becoming barrierless at approximately 12 GPa of applied planar stress.

Additionally, the impact of hydrostatic pressure on the R8 to β -Sn transition was investigated. Results indicate a gradual reduction in the activation barrier with increasing pressure, as shown in Fig. 5d. Interestingly, a transition pathway via the BCC phase emerges above 20 GPa. In this case, more intermediate images have been added to the pathway which has been broken up into separate NEB calculations between all local minima.

It is noteworthy that, as illustrated in Fig. 5, a comparison of the ssNEB calculations performed using the MLIP with those conducted using DFT reveals that the prediction of the energy of intermediate images in the transition from R8 to β -Sn is quite poor, with errors reaching nearly 30 meV. This discrepancy arises because this transition process was examined as a test after the development of the MLIP, thus lacking intermediate structures in the training set.

NPT MD simulations were once again utilized to explore this pressure-induced transformation. A rhomboidal prism-shaped simulation cell housing 12288 atoms was employed, with the temperature maintained at 200 K and hydrostatic pressure at 23 GPa. The pressure valued was chosen by conducting a ss-NEB calculation on a simulation cell like the one used for the results reported

in Fig. 5d. We estimated a kinetic barrier at this pressure value to be about 8 meV/atom. Such a low barrier would allow us to see the transition directly during MD in a short simulation.

Snapshots capturing during MD simulation are depicted in Fig. 7a. The nucleation of the

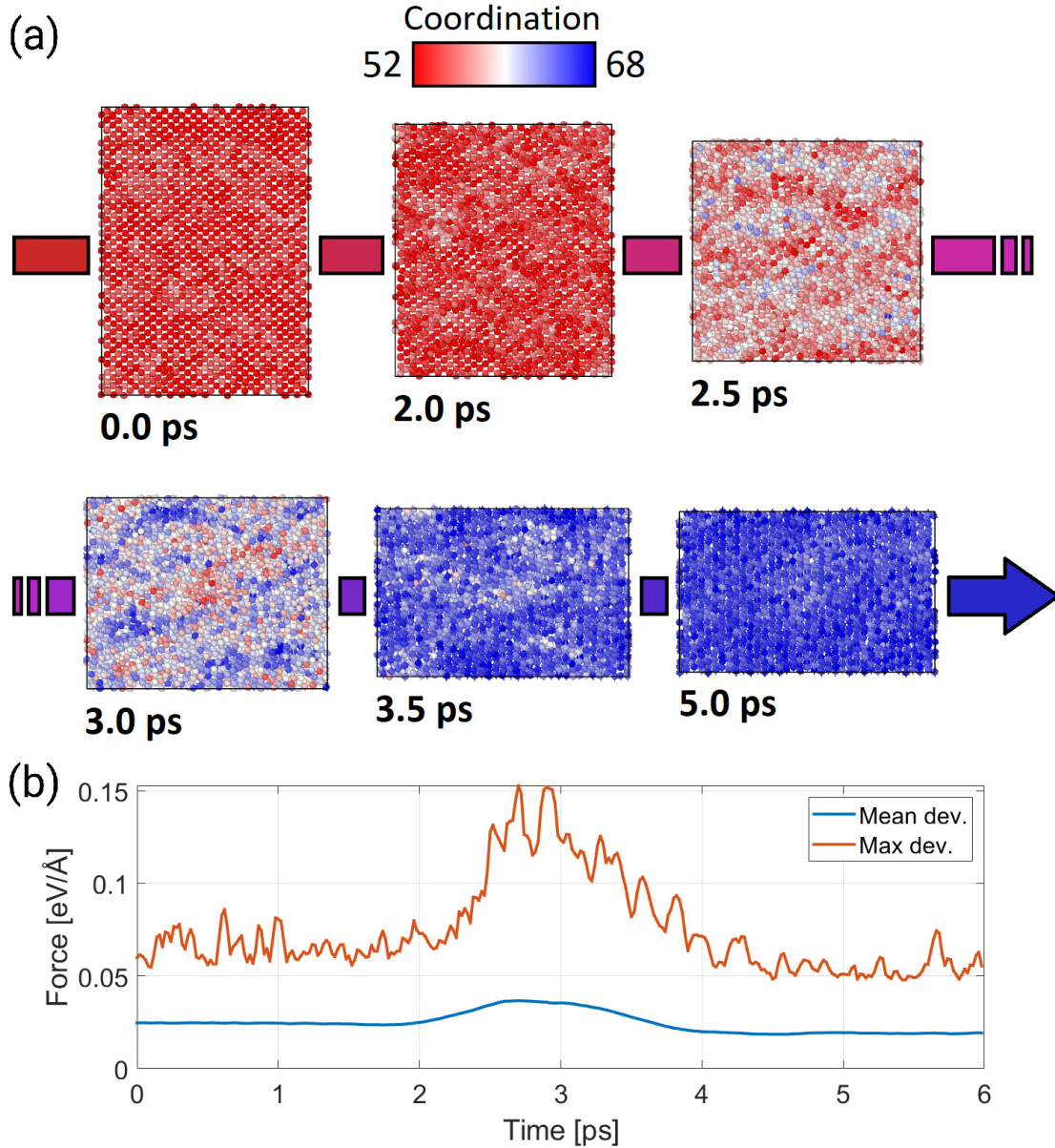


FIG. 6. (a) Snapshot images of slices taken from the trajectories of NPT MD simulation for the pressure-induced transformation mechanism from CD to β -Sn, performed at 300 K and with uniaxial compressive stress along the z -direction at 7 GPa. Color coding is done based on coordination analysis in Ovito⁷⁹, with the CD and β -Sn structures depicted in red and blue respectively. (b) Average and peak force deviations during the MD simulation.

BCC structure can be observed emerging from the R8 bulk, and subsequently transform into poly-

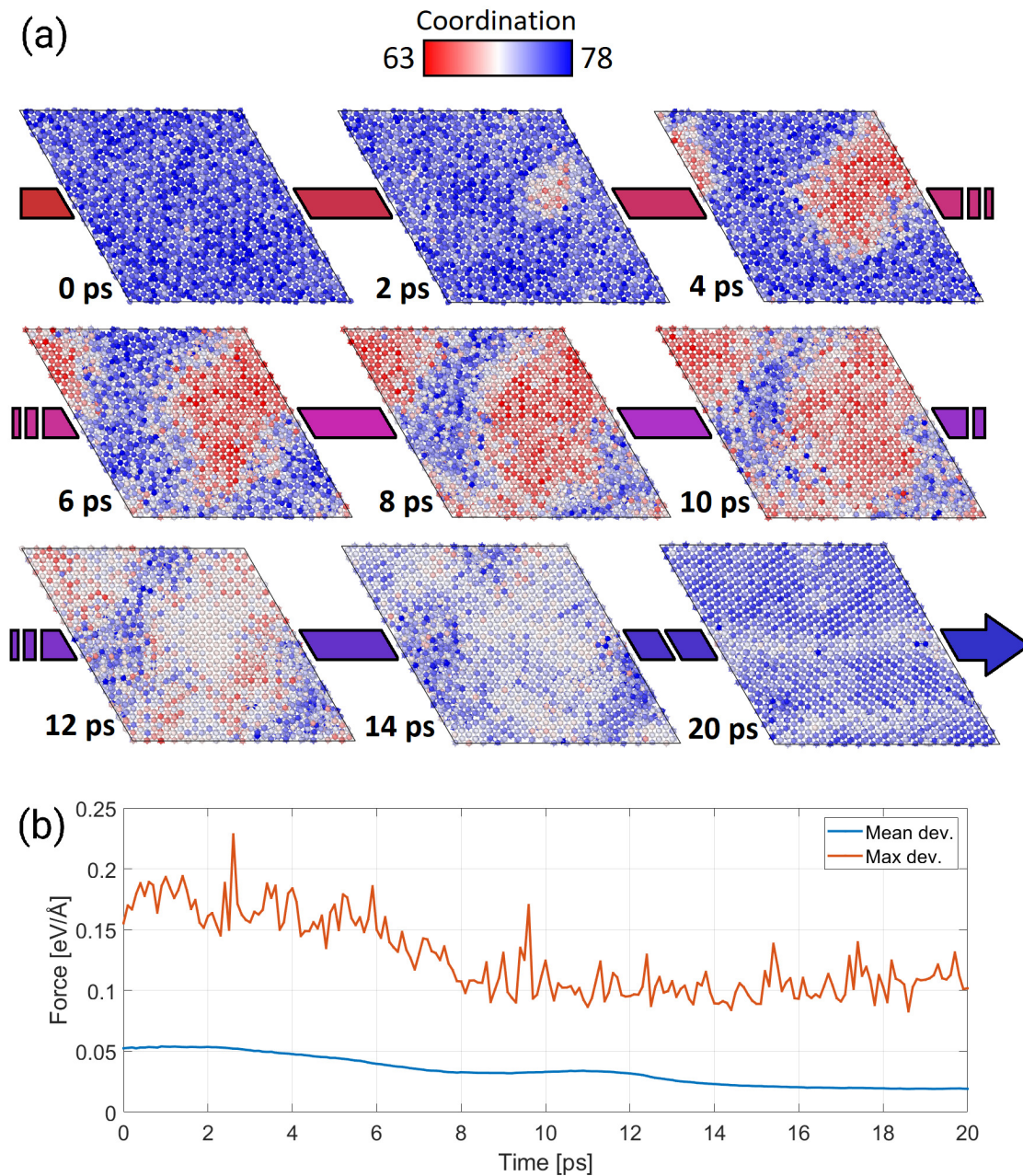


FIG. 7. (a) Snapshot images of slices taken from the trajectories of NPT MD simulation of the pressure-induced transformation mechanism from R8 to β -Sn via the BCC intermediate phase, performed at 200 K and with hydrostatic pressure at 23 GPa. Color coding is done based on coordination analysis in Ovito, and it is used to clearly show the R8 bulk (in blue) in which the BCC nucleation process occurs (in red) and finally transforms to get β -Sn (in blue again). (b) Average and peak force deviations during the MD simulation.

crystalline β -Sn with defects. As shown in Fig. 7b, mean deviations are approximately 0.05 eV/\AA and decrease below 0.02 eV/\AA as the transition progresses and more β -Sn forms.

The transition from β -Sn to R8, in reverse order with respect to the one investigated here, has been experimentally observed to occur in nanoindentation upon unloading at specific rates and stress conditions. However, the R8 to β -Sn transition under loading investigated in this work has been less studied, primarily because obtaining R8 from β -Sn is a prerequisite. Therefore, for such a transition there are no existing values to compare our estimation with. Still, our findings could be valuable for better understanding cyclic nanoindentation experiments, such as those described in Ref²⁸.

IV. CONCLUSIONS AND PERSPECTIVES

In this work, an accurate and efficient MLIP tailored for investigating pressure-dependent crystal phase transitions in germanium has been developed. The resulting DP is several orders of magnitude faster than DFT calculations and scales linearly with the number of atoms. Regression plots comparing NN predicted energies, forces, and stresses with DFT calculations performed on a suitable test set, together with the shown accurate prediction of formation energies and energy-volume curves of many relevant crystal phases, demonstrate the robustness of the model in describing such metastable allotropes of Ge. Moreover, the iterative refinement of the model through insertion in the training set of transition structures sampled from ssNEB calculations performed with the model itself has been shown as a simple yet effective practice to get the model to predict accurate activation barriers.

Our MLIP, which can be downloaded together with the full database⁸¹, will serve as a valuable tool facilitating further advancements in the study of Ge crystal phase transitions and their applications.

ACKNOWLEDGEMENTS

A.F. expresses gratitude to Naman Katyal and Jiyoung Lee for their insightful feedback. A.F., F.M., and E.S. acknowledge financial support from ICSC – Centro Nazionale di Ricerca in High-Performance Computing, Big Data and Quantum Computing, funded by the European Union – NextGenerationEU. A.F., F.R., F.M., and E.S. acknowledge the CINECA consortium under the

ISCRA initiative for the availability of high-performance computing resources and support. Work in Austin was supported by the National Science Foundation (CHE-2102317) and the Texas Advanced Computing Center. E.S. acknowledges financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 104 published on 2.2.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union – NextGenerationEU– Project Title "SiGe Hexagonal Diamond Phase by nanoIndenTation (HD- PIT)" – CUP H53D23000780001 - Grant Assignment Decree No. 957 adopted on 30.06.2023 by the Italian Ministry of Ministry of University and Research (MUR).

REFERENCES

- ¹J. Behler and M. Parrinello, *Physical Review Letters* **98**, 146401 (2007).
- ²J. Behler, *Physical Chemistry Chemical Physics* **13**, 17930 (2011).
- ³A. V. Shapeev, *Multiscale Modeling & Simulation* **14**, 1153 (2016).
- ⁴A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, *Physical Review X* **8**, 041048 (2018).
- ⁵Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, and S. P. Ong, *The Journal of Physical Chemistry A* **124**, 731 (2020).
- ⁶Y. Lysogorskiy, C. V. D. Oord, A. Bochkarev, S. Menon, M. Rinaldi, T. Hammerschmidt, M. Mrovec, A. Thompson, G. Csányi, C. Ortner, and R. Drautz, *npj Computational Materials* **7**, 97 (2021).
- ⁷O. T. Unke and M. Meuwly, *Journal of Chemical Theory and Computation* **15**, 3678 (2019).
- ⁸T. Devergne, T. Magrino, F. Pietrucci, and A. M. Saitta, *Journal of Chemical Theory and Computation* **18**, 5410 (2022).
- ⁹A. Mondal, D. Kussainova, S. Yue, and A. Z. Panagiotopoulos, *Journal of Chemical Theory and Computation* **19**, 4584 (2023).
- ¹⁰S. Stocker, H. Jung, G. Csányi, C. F. Goldsmith, K. Reuter, and J. T. Margraf, *Journal of Chemical Theory and Computation* **19**, 6796 (2023).
- ¹¹M. Bertani, T. Charpentier, F. Faglioni, and A. Pedone, *Journal of Chemical Theory and Computation* **20**, 1358 (2024).
- ¹²E. M. T. Fadaly, A. Dijkstra, J. R. Suckert, D. Ziss, M. A. J. Van Tilburg, C. Mao, Y. Ren, V. T. Van Lange, K. Korzun, S. Kölling, M. A. Verheijen, D. Busse, C. Rödl, J. Furthmüller, F. Bechstedt, J. Stangl, J. J. Finley, S. Botti, J. E. M. Haverkort, and E. P. A. M. Bakkers, *Nature*

- 580**, 205 (2020).
- ¹³F. H. Stillinger and T. A. Weber, *Physical Review B* **31**, 5262 (1985).
- ¹⁴K. Ding and H. C. Andersen, *Physical Review B* **34**, 6987 (1986).
- ¹⁵J. Tersoff, *Physical Review B* **39**, 5566 (1989).
- ¹⁶M. I. Baskes, J. S. Nelson, and A. F. Wright, *Physical Review B* **40**, 6085 (1989).
- ¹⁷E. H. Kim, Y.-H. Shin, and B.-J. Lee, *Calphad* **32**, 34 (2008).
- ¹⁸G. Ge, F. Rovaris, D. Lanzoni, L. Barbisan, X. Tang, L. Miglio, A. Marzegalli, E. Scalise, and F. Montalenti, *Acta Materialia* **263**, 119465 (2024).
- ¹⁹G. Voronin, C. Pantea, T. Zerda, J. Zhang, L. Wang, and Y. Zhao, *Journal of Physics and Chemistry of Solids* **64**, 2113 (2003).
- ²⁰A. Mujica, A. Rubio, A. Muñoz, and R. J. Needs, *Reviews of Modern Physics* **75**, 863 (2003).
- ²¹F. Coppari, J. C. Chervin, A. Congeduti, M. Lazzeri, A. Polian, E. Principi, and A. Di Cicco, *Physical Review B* **80**, 115213 (2009).
- ²²K. Takemura, U. Schwarz, K. Syassen, M. Hanfland, N. E. Christensen, D. L. Novikov, and I. Loa, *Physical Review B* **62**, R10603 (2000).
- ²³X.-J. Chen, C. Zhang, Y. Meng, R.-Q. Zhang, H.-Q. Lin, V. V. Struzhkin, and H.-k. Mao, *Physical Review Letters* **106**, 135502 (2011).
- ²⁴J.-i. Jang, M. J. Lance, S. Wen, and G. M. Pharr, *Applied Physics Letters* **86**, 131907 (2005).
- ²⁵Z. Zhao, H. Zhang, D. Y. Kim, W. Hu, E. S. Bullock, and T. A. Strobel, *Nature Communications* **8**, 13909 (2017).
- ²⁶L. Q. Huston, B. C. Johnson, B. Haberl, S. Wong, J. S. Williams, and J. E. Bradby, *Journal of Applied Physics* **122**, 175108 (2017).
- ²⁷B. C. Johnson, B. Haberl, S. Deshmukh, B. D. Malone, M. L. Cohen, J. C. McCallum, J. S. Williams, and J. E. Bradby, *Physical Review Letters* **110**, 085502 (2013).
- ²⁸K. Kosai, H. Huang, and J. Yan, *Crystals* **7**, 333 (2017).
- ²⁹B. D. Malone and M. L. Cohen, *Physical Review B* **86**, 054101 (2012).
- ³⁰Q. Yuan, S. Li, L. Zhou, and D. He, *Solid State Communications* **348-349**, 114742 (2022).
- ³¹B. Haberl, M. Guthrie, B. D. Malone, J. S. Smith, S. V. Sinogeikin, M. L. Cohen, J. S. Williams, G. Shen, and J. E. Bradby, *Physical Review B* **89**, 144111 (2014).
- ³²S. Tarkhorani and M. Sasani Ghamsari, *Materials Science and Engineering: B* **261**, 114665 (2020).

- ³³A. Garcia-Gil, S. Biswas, A. Roy, D. Saladukh, S. Raha, T. Blon, M. Conroy, V. Nicolosi, A. Singha, L.-M. Lacroix, and J. D. Holmes, *Nanoscale* **14**, 2030 (2022).
- ³⁴J. S. Williams, B. Haber, S. Deshmukh, B. C. Johnson, B. D. Malone, M. L. Cohen, and J. E. Bradby, *physica status solidi (RRL) – Rapid Research Letters* **7**, 355 (2013).
- ³⁵G. Dushaq, A. Nayfeh, and M. Rasras, *Scientific Reports* **9**, 1593 (2019).
- ³⁶A. G. Lyapin, V. V. Brazhkin, S. V. Popova, and A. V. Sapelkin, *physica status solidi (b)* **198**, 481 (1996).
- ³⁷B. Haberl, J. E. Bradby, M. V. Swain, J. S. Williams, and P. Munroe, *Applied Physics Letters* **85**, 5559 (2004).
- ³⁸A. Di Cicco, A. Congeduti, F. Coppari, J. C. Chervin, F. Baudelet, and A. Polian, *Physical Review B* **78**, 033309 (2008).
- ³⁹G. Kresse and J. Furthmüller, *Computational Materials Science* **6**, 15 (1996).
- ⁴⁰G. Kresse and J. Furthmüller, *Physical Review B* **54**, 11169 (1996).
- ⁴¹G. Kresse and D. Joubert, *Physical Review B* **59**, 1758 (1999).
- ⁴²J. P. Perdew, K. Burke, and M. Ernzerhof, *Physical Review Letters* **77**, 3865 (1996).
- ⁴³A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. In 'T Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, *Computer Physics Communications* **271**, 108171 (2022).
- ⁴⁴H. Wang, L. Zhang, J. Han, and W. E, *Computer Physics Communications* **228**, 178 (2018).
- ⁴⁵G. J. Martyna, D. J. Tobias, and M. L. Klein, *The Journal of Chemical Physics* **101**, 4177 (1994).
- ⁴⁶E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, *Physical Review Letters* **97**, 170201 (2006).
- ⁴⁷<https://wiki.fysik.dtu.dk/ase/>, for information about the ASE project.
- ⁴⁸D. G. Truhlar, B. C. Garrett, and S. J. Klippenstein, *The Journal of Physical Chemistry* **100**, 12771 (1996).
- ⁴⁹G. Henkelman and H. Jónsson, *The Journal of Chemical Physics* **111**, 7010 (1999).
- ⁵⁰L. J. Munro and D. J. Wales, *Physical Review B* **59**, 3969 (1999).
- ⁵¹H. B. Schlegel, *WIREs Computational Molecular Science* **1**, 790 (2011).
- ⁵²X.-J. Zhang, C. Shang, and Z.-P. Liu, *Journal of Chemical Theory and Computation* **9**, 5745 (2013).
- ⁵³S. T. Chill, J. Stevenson, V. Rühle, C. Shang, P. Xiao, J. D. Farrell, D. J. Wales, and G. Henkelman, *Journal of Chemical Theory and Computation* **10**, 5476 (2014).

- ⁵⁴X.-J. Zhang and Z.-P. Liu, *Journal of Chemical Theory and Computation* **11**, 4885 (2015).
- ⁵⁵A. Jay, C. Huet, N. Salles, M. Gunde, L. Martin-Samos, N. Richard, G. Landa, V. Goiffon, S. De Gironcoli, A. Hémerlyck, and N. Mousseau, *Journal of Chemical Theory and Computation* **16**, 6726 (2020).
- ⁵⁶D. Sheppard, P. Xiao, W. Chemelewski, D. D. Johnson, and G. Henkelman, *The Journal of Chemical Physics* **136**, 074103 (2012).
- ⁵⁷G. Henkelman, B. P. Uberuaga, and H. Jónsson, *The Journal of Chemical Physics* **113**, 9901 (2000).
- ⁵⁸G. Henkelman and H. Jónsson, *The Journal of Chemical Physics* **113**, 9978 (2000).
- ⁵⁹<http://theory.cm.utexas.edu/henkelman/code/>, to obtain the TSASE code.
- ⁶⁰E. V. Podryabinkin and A. V. Shapeev, *Computational Materials Science* **140**, 171 (2017).
- ⁶¹J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, *The Journal of Chemical Physics* **148**, 241733 (2018).
- ⁶²K. Gubaev, E. V. Podryabinkin, G. L. Hart, and A. V. Shapeev, *Computational Materials Science* **156**, 148 (2019).
- ⁶³C. Schran, K. Brezina, and O. Marsalek, *The Journal of Chemical Physics* **153**, 104105 (2020).
- ⁶⁴A. Jain, S. P. Ong, Geoffroy Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, *APL Materials* **1**, 011002 (2013).
- ⁶⁵C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, *The Journal of Chemical Physics* **108**, 1964 (1998).
- ⁶⁶G. Bussi and A. Laio, *Nature Reviews Physics* **2**, 200 (2020).
- ⁶⁷J. Kästner, *WIREs Computational Molecular Science* **1**, 932 (2011).
- ⁶⁸H. Zhou and P. Tao, *Journal of Chemical Theory and Computation* **14**, 14 (2018).
- ⁶⁹M. I. Zimmerman, J. R. Porter, X. Sun, R. R. Silva, and G. R. Bowman, *Journal of Chemical Theory and Computation* **14**, 5459 (2018).
- ⁷⁰J. Debnath, M. Invernizzi, and M. Parrinello, *Journal of Chemical Theory and Computation* **15**, 2454 (2019).
- ⁷¹J. Nam and Y. Jung, *Journal of Chemical Theory and Computation* **19**, 2735 (2023).
- ⁷²A. Hulm and C. Ochsenfeld, *Journal of Chemical Theory and Computation* **19**, 9202 (2023).
- ⁷³L. Zhang, J. Han, H. Wang, R. Car, and W. E, *Physical Review Letters* **120**, 143001 (2018).
- ⁷⁴L. Zhang, J. Han, H. Wang, W. Saidi, R. Car, and W. E, in *Advances in Neural Information Processing Systems*, Vol. 31, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman,

- N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018).
- ⁷⁵D. P. Kingma and J. Ba, (2014), 10.48550/ARXIV.1412.6980, publisher: arXiv Version Number: 9.
- ⁷⁶A. Mujica, C. J. Pickard, and R. J. Needs, *Physical Review B* **91**, 214104 (2015).
- ⁷⁷Z. Zhao, F. Tian, X. Dong, Q. Li, Q. Wang, H. Wang, X. Zhong, B. Xu, D. Yu, J. He, H.-T. Wang, Y. Ma, and Y. Tian, *Journal of the American Chemical Society* **134**, 12362 (2012).
- ⁷⁸A. Ghasemi, P. Xiao, and W. Gao, *The Journal of Chemical Physics* **151**, 054110 (2019).
- ⁷⁹A. Stukowski, *Modelling and Simulation in Materials Science and Engineering* **18**, 015012 (2010).
- ⁸⁰J.-T. Wang, C. Chen, H. Mizuseki, and Y. Kawazoe, *Physical Review Letters* **110**, 165503 (2013).
- ⁸¹A. Fantasia et al., “A NN potential for phase transformations in Ge,” *Materials Cloud Archive* 2024.55; <https://doi.org/10.24435/materialscloud:r2-qc>.