

PAPER

Machine learning interatomic potentials for aluminium: application to solidification phenomena

To cite this article: Noel Jakse *et al* 2023 *J. Phys.: Condens. Matter* **51** 035402

View the [article online](#) for updates and enhancements.

You may also like

- [Surface-enhanced Raman scattering](#)
A Otto, I Mrozek, H Grabhorn et al.
- [Proceedings of the Colloidal Dispersions in External Fields II Conference \(Bonn-Bad Godesberg, 31 March–2 April 2008\)](#)
H Löwen
- [Effect of transverse displacement of charged particle beams on quantum electrodynamic processes during their collision](#)
M. Filipovic, C. Baumann, A.M. Pukhov et al.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Machine learning interatomic potentials for aluminium: application to solidification phenomena

Noel Jakse¹ , Johannes Sandberg^{1,2,3}, Leon F Granz^{2,3} , Anthony Saliou¹, Philippe Jarry⁴, Emilie Devijver⁵, Thomas Voigtmann^{2,3} , Jürgen Horbach⁶ and Andreas Meyer^{2,7}

¹ Université Grenoble Alpes, CNRS, Grenoble INP, SIMaP, F-38000 Grenoble, France

² Institut für Materialphysik im Weltraum, Deutsches Zentrum für Luft- und Raumfahrt (DLR), 51170 Köln, Germany

³ Department of Physics, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

⁴ C-TEC, Parc Economique Centr' alp, 725 rue Aristide Bergès, CS10027, Voreppe 38341 CEDEX, France

⁵ Université Grenoble Alpes, CNRS, Grenoble INP, LIG, F-38000 Grenoble, France

⁶ Institut für Theoretische Physik II, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

⁷ Institut Laue-Langevin (ILL), 38042 Grenoble, France

E-mail: noel.jakse@grenoble-inp.fr

Received 6 August 2022, revised 10 October 2022

Accepted for publication 25 October 2022

Published 15 November 2022



Abstract

In studying solidification process by simulations on the atomic scale, the modeling of crystal nucleation or amorphization requires the construction of interatomic interactions that are able to reproduce the properties of both the solid and the liquid states. Taking into account rare nucleation events or structural relaxation under deep undercooling conditions requires much larger length scales and longer time scales than those achievable by *ab initio* molecular dynamics (AIMD). This problem is addressed by means of classical molecular dynamics simulations using a well established high dimensional neural network potential trained on a set of configurations generated by AIMD relevant for solidification phenomena. Our dataset contains various crystalline structures and liquid states at different pressures, including their time fluctuations in a wide range of temperatures. Applied to elemental aluminium, the resulting potential is shown to be efficient to reproduce the basic structural, dynamics and thermodynamic quantities in the liquid and undercooled states. Early stages of crystallization are further investigated on a much larger scale with one million atoms, allowing us to unravel features of the homogeneous nucleation mechanisms in the fcc phase at ambient pressure as well as in the bcc phase at high pressure with unprecedented accuracy close to the *ab initio* one. In both cases, a single step nucleation process is observed.

Supplementary material for this article is available [online](#)

Keywords: potentials, aluminiums, machine learning, molecular dynamics, homogeneous nucleation

(Some figures may appear in colour only in the online journal)

1. Introduction

Apart from steels, aluminium and its alloys represent the most used and attractive structural metallic materials due to their specific properties such as low weight, low energy cost of remelting, and the possibility of almost complete recycling. Therefore, these materials represent a major axis of the energy transition [1]. An intimate understanding of their condensed phase properties is founded upon a description of the atomic level structure and dynamics, and requires an accurate representation of chemical bonding [2]. This is of utmost importance in order to tackle phenomena such as phase changes and solidification process during which a liquid morphs into a solid either by crystallization or amorphization [3, 4], showing eventually a change in electronic structure as a metal-to-semiconductor transition [5, 6]. First-principles approaches, essentially through the density functional theory (DFT) [7, 8], represent the dedicated framework especially with the breakthrough provided by *ab initio* molecular dynamics (AIMD) simulations [9] in combining atomic dynamics with DFT. Despite its enormous success in many complex chemical bonding environments [2], DFT implementations are limited to a few hundred atoms over time scales less than 1 ns [10, 11] on current large-scale supercomputing facilities, impeding its use for phenomena at length and time scales typical of solidification [12].

The desire to bridge typical scales of the electronic structure to those of the properties under investigation has led to deriving interatomic potentials with semi-empirical functional forms that average out or otherwise model electronic degrees of freedom. For metallic materials of interest here, various approaches were proposed, starting in the second half of the 20th Century with pair-potentials based on a nearly-free electron gas description of the electronic structure [13] using simple models within the pseudopotential theory (PT) [14–17]. It was later acknowledged that it was impossible for pair potentials to describe on the same footing the structure, dynamic, and thermodynamic properties in the liquid and solid state [18], having in addition inherent mechanical instability under shear for crystals. Many-body approaches such as the embedded-atom model (EAM) [19, 20], modified EAM (MEAM) [21] of current widespread use as well as the Reactive Force Field (ReaxFF) [22], just to name a few among many others [23], can be considered as successful in this respect. Fitting the parameters of these potentials is most often oriented towards describing crystalline phases and transitions between some of them [24, 25], more rarely taking a full account of the liquid state [26]. This leads to a lack of transferability [18] and a limited ability to tackle phenomena involving several phases such as crystal nucleation [12].

Over the last decade, impressive progress was made in designing potentials from electronic structure calculations using supervised machine learning (ML) methods [27–33]. There are now standard libraries for the ML training [34] that can be used in combination with molecular dynamics (MD) simulation packages [35] such as LAMMPS [36] or in combination with workflow environments such as ASE [37]. On-the-fly ML force field methods have been also proposed

[38] and implemented directly into *ab initio* codes in order to bypass most of electronic-structure calculation steps [39]. Different ML techniques have been used, ranging from simple linear regression methods such as the spectral neighbor analysis potential method [31, 40] to highly non-linear regression methods using high dimensional neural networks [34, 41] (HDNNs) or kernel regression [42–44]. The designed potentials reach in general an accuracy close to the *ab initio* calculations from which the database was formed, with exceptional results for the description of relative stability between crystalline phases [45] and defects [25]. However, approaches taking full account of liquid and crystalline states remain scarce [6, 45–47] and are often limited to the objective of a good description of the melting point. The main reason for this stems from the fact that the chosen *ab initio* configurations should cover all situations, as ML techniques may become less reliable outside the training domain [27]. It becomes even more crucial for crystal nucleation occurring under deep undercooling conditions with a strong evolution of the liquid structure with respect to that above melting, showing an increasing icosahedral [10] ordering and structural heterogeneity [48] triggering homogeneous nucleation [49].

Machine-learning potentials for aluminium were designed very recently to describe essentially the properties of the solid states [50–52] and the melting temperature [50, 51]. Different approaches were put forward respectively with a Gaussian kernel regression [51] and a deep NN [52] with a dataset built from configurations extracted from AIMD simulations at various temperatures, and an active-learning approach with periodical retraining of the NN [50]. ML potentials were initially trained using the DFT energies starting from the work of Behler and Parrinello on bulk Si [41]. It was subsequently pointed out that the learning process could benefit from a wealth of additional information if the three components of the force and six components of the stress per atom are taken into account [27, 33], while one has only a single energy value per simulated configuration. Still in some works, only the forces have been used for the training showing that properties like the vibrational properties in the solid states can be reproduced, but they remain insufficient to get full account of thermodynamic quantities [44, 51]. Thus whether additional information enhances the training or not can still be questioned, also given the fact that the relative importance of the energy, forces and stresses for estimating the mean-square error (MSE) or the root-MSE (RMSE) introduce two additional free training parameters [27]. Moreover, the question of the transferability of a ML potential taking into account both the liquid and solid phases as mentioned above, remains essentially unexplored for aluminium. This aspect is also of importance when dealing with solidification phenomena.

The aim of the present work is to develop a ML potential for pure aluminium dedicated to the description of condensed phases, namely liquid and solid states for temperatures up to 8000 K and pressures up to 300 GPa being therefore able to deal with solidification phenomena. For this purpose, a HDNN was developed on the basis of well-known and robust Behler and Parrinello's approach [27, 41]. The latter was trained on a data set generated by DFT-based simulations for the main

crystalline structures and liquid states covering the targeted pressure and temperature domain, including their time fluctuations by an appropriate sampling of phase space trajectories [53]. It is shown that training the HDNN on sampled AIMD trajectories using solely energy labels leads to an accurate description of the structure, dynamics and thermodynamics in the investigated domain. More specifically, the single-particle as well as the collective dynamics are well reproduced, which is an essential ingredient in describing solidification. The resulting potential is then applied here to solidification processes, namely amorphization and early stages of crystal nucleation, allowing us to unravel features of the homogeneous nucleation mechanisms at ambient as well as high pressure.

The remaining part of the paper is organized as follows. In section 2, the specific features of the HDNN, the training procedure as well as the basic assessment of the potential on independent DFT and experimental thermodynamic data are outlined. Section 3 is devoted to the test of the potential's accuracy in describing some structural properties, the dynamics of the liquid state in the investigated pressure-temperature domain, as well as the homogeneous nucleation. Finally, in section 4, the main outcomes of the work are given.

2. Computational background

2.1. Constructing a machine learning potential

In the last three decades, many potentials of pure Al and its alloys for the use in atomistic simulations been developed [54–67], using the Morse potential ansatz [54], the PT method [62], EAM [56–58, 61, 63, 65], the modified MEAM [21, 60, 67], and many-body approaches such as COMB3 [66]. However, only very few studies have employed the ML approach [44, 51, 52], none of them taking systematically into account the properties of the liquid state and checking the dynamics which is very important for the solidification aspects. This is precisely one of the aims in building a ML-based potential here.

Among the various approaches put forward to design ML potentials [28, 31], the choice was made to set up a HDNN built in a similar way to the one proposed by Behler and Parrinello [41] and Zhang *et al* [68]. This well established approach has been proven successful for pure silicon [6] as well as water [45] to model reliably both their properties in the liquid and solid states. As a detailed description can be found for instance in the tutorial review by Behler [27] among others, the focus is made mainly on the specifics of our scheme. The main part consists in a supervised learning task from a relevant sample of atomic configurations with known energy generated by AIMD in various crystalline and liquids structures in the desired temperature and pressure domain. In the corresponding portion of potential energy landscape (PEL), each configurational energy is considered as a sum of individual atomic energies determined from their local atomic environment within a cut-off radius r_s often extending beyond the first-neighbor atomic shell, and taken here to be 6.4 Å for Al, corresponding to at least the second neighbor shell for all the considered thermodynamic states. This decomposition

allows us to train N neural networks (NN), each of them being assigned to an individual local atomic environment. The NN are then combined to recover the energy of the whole configuration of N atoms.

The individual NN is defined by the same network topology for given atomic species. It specifies the number of neurons formally named here y_i^l and their connectivity through the weights, $w_{i,j}^l$, and called a multi-layer perceptron (see figure 1). The weights associated with each node pair are optimized during the learning process by a back-propagation technique [69]. Thus, each of the M layers within the NN consists of sets of nodes receiving multiple inputs from the previous layer and passing outputs to the next layer. Here a fully connected network is used, in which every output of a layer is an input for every neuron in the next layer. The corresponding mathematical description is as follows: the input signals are linearly combined before being activated by function f to give each output y_i^l of a given fully connected layer l as

$$y_i^l = f \left(\sum_{j=1}^{M_{l-1}} w_{i,j}^l y_j^{l-1} + b_i^l \right), \quad (1)$$

where M_l refers to the size of the l th layer, i.e. the number of its neurons. Note that positive weights enhance connections while negative weights tend to inhibit them. Most of the activation functions are chosen to have a range in either $[0, 1]$ or $[-1, 1]$ and modulate the amplitude of the output. The activation function f is applied element-wise and is taken as the soft-plus function $f(x) = \log(1 + e^x)$. Back-propagation is used to update the network weights and their gradients.

The input layer of a NN takes values representative of one local atomic environment in the form of a feature vector whose dimension is then equal to the number of its nodes. The feature vector is built on the basis of Behler-Parrinello (BP) descriptors [41] to represent the radial and angular arrangements of atoms in the local structure using Gaussian symmetry functions having the translational and rotational invariance. For aluminium, the number of components of the BP feature was chosen to be 22, comprising of 12 radial and ten angular components, as described in more detail in the Supplementary Information (SI) file. Then, the NN architecture for aluminum is $22 \times 10 \times 10 \times 1$ with two hidden layers of ten nodes each.

The NN was coded using KERAS module from the TENSORFLOW PYTHON package [70] in the regression mode. NNs of all N atoms of a configuration are then associated using the ADD module to form the HDNN which is obviously invariant to permutation of atoms. The HDNN is then trained on the single DFT energy of the whole configuration.

2.2. Building the dataset

Designing an appropriate dataset is the crucial and demanding step for the construction of the NN interatomic potential. Various strategies can be put forward to construct it, which were reviewed very recently [71]. Here, it was built from AIMD simulations that were partly taken from our previous works [10, 72–74] and extended here to have a better representation of the undercooled liquid region, the crystalline fcc

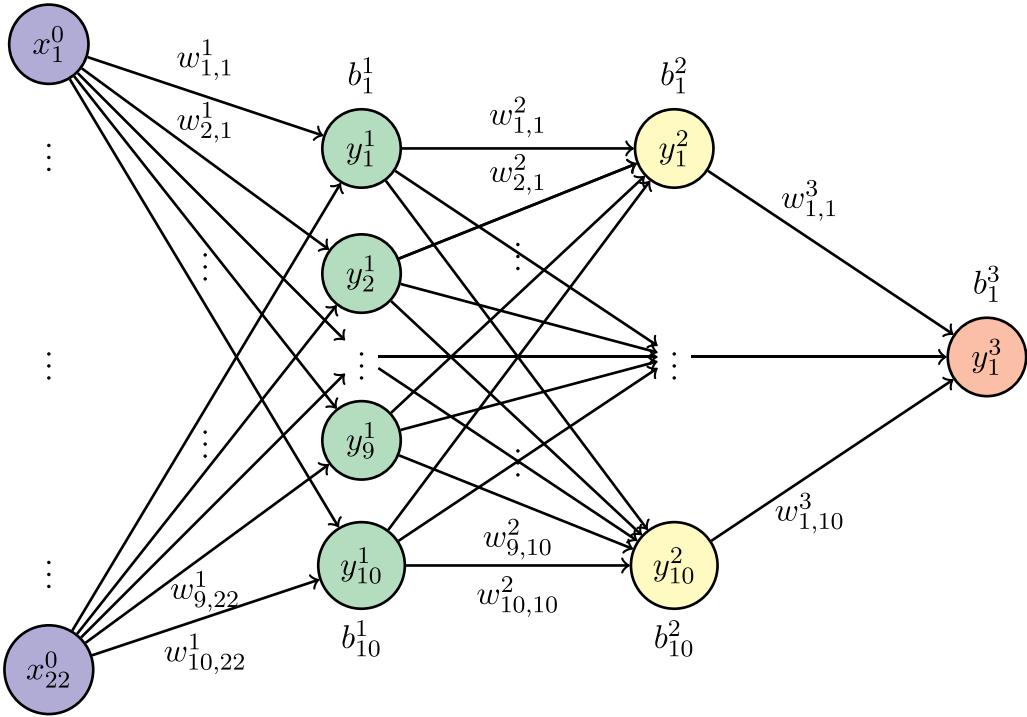


Figure 1. Schematic representation of the feed-forward neural network built as a densely connected multi-layer perceptron. The input layer x_i^0 will be fed with the atomic BP feature of the training set, and the output is the atomic energy. The neural network contains two hidden layers with superscript 1 and 2, respectively. The two layers are composed of ten neurons each. The weights w_{ij}^k and bias b_i^k are optimized during the training (see text).

configurations up to the melting point at zero pressure, and crystalline fcc, bcc, and hcp up to 300 GPa with and without defects. The different thermodynamic states and structures as well as the number of configurations sampled in each case are given in tables SI and SII in the SI file. In practice, for each temperature and pressure a number of 100 to 1000 configurations are sampled from the corresponding AIMD trajectory. In total, 24 300 configurations of $N = 256$ atoms were gathered in the database that enabled us to cover solid and liquid states at ambient pressure as well as liquid samples at temperatures up to 8000 K and pressures up to 300 GPa. Non-equilibrium trajectories in the undercooled region were also generated to take into account crystal nucleation and solidification processes in the ML fitting procedure.

For the sake of self-consistency, the main technical details of the AIMD simulations are recalled here. They were performed by means of the Vienna *Ab initio* Simulation Package [75]. The local density approximation (LDA) [76, 77] within projected augmented plane-waves [78] was applied to all simulations with a plane-wave cutoff of 241 eV. For the liquid states, only the Γ -point is used while for crystalline states, the Γ -centered grid of k -points in the irreducible part of the Brillouin zone was set to $2 \times 2 \times 2$ following the Monkhorst–Pack scheme [79, 80]. All the simulations were performed with $N = 256$ atoms placed in a cubic simulation box (except for the hcp crystal where an orthorhombic box was used) with standard periodic boundary conditions (PBCs). Newton’s Equations of motion were solved numerically with Verlet’s algorithm in the velocity form with a time step of 1.5 fs, and

phase-space trajectories were constructed within the canonical ensemble (NVT, i.e. constant temperature, volume, and number of atoms), by means of a Nosé thermostat to control the temperature T . The temperature evolution in the undercooled states was obtained by quenching the system stepwise down to 600 K with a temperature step of 50 K. For each temperature, the simulation cell was resized according to the experimental density [81] and the run was continued for 30 ps before performing the next quench, resulting in an average cooling rate of 3.3×10^{12} K s⁻¹. The calculated pressures for all the temperatures studied here were in the range ± 1 GPa generally, so that on average a quasi constant pressure during the quenching is observed. For temperatures ranging from $T = 1000$ K to 600 K, the run was continued for equilibration during a time up to 200 ps. A similar procedure was applied for heating the fcc crystal from 10 K to 900 K.

Several aspects deserve attention in the perspective of building the ML potential. First of all, the choice of the exchange and correlation (XC) functional for the electronic structure calculations is crucial. As the ML potential may reach an accuracy similar to the DFT calculations, it will mirror the ability of the XC functional used in predicting the properties, at least in the thermodynamic domain inside which it was trained. This has guided our choice of the LDA functional given the fact that the generalized-gradient-approximation (GGA) overestimates the atomic volume [82]. Moreover, it was shown in our previous contributions that the LDA gives a good description of the liquid structure [10]. More importantly, atomic transport properties such as the self-diffusion

coefficient, which are very sensitive to the details of the potentials, are well reproduced within the LDA compared to state-of-the-art experimental data [83, 84]. Such a good agreement with experiments was very recently confirmed on the dynamic structure factors as well as the structural relaxation times extracted from the intermediate scattering function [74]. For high pressures, it was shown [85] that the difference between LDA and GGA [86] is negligible in describing the pressure-density phase diagram of aluminium up to pressures as high as 10 TPa.

Secondly, in the perspective of performing MD simulations, care has to be taken in describing not only average thermodynamic properties but also the fluctuations around the mean value, especially in order to capture the features of local basins of the PEL [53]. This requires the sampling of a large number of configurations along AIMD phase space trajectories. Therefore, in the present work, for each of the considered thermodynamic states (see tables SII and SIII of the SI file), 1000 configurations were generated on AIMD production runs over 40 ps.

Finally, as mentioned in the introduction, the question whether including the additional information of the forces or even the stresses in addition to the energies improves the learning process and the accuracy of the potential deserves further attention. It was shown very recently for molecular systems that forces and energies contribute equally to the convergence of the prediction errors [87]. The choice of considering energies, forces or both of them in the training may depend on factors such as the application domain, the properties of interest, the complexity of the ML tool, and the strategy in building the data from *ab initio* calculations. When making static DFT calculations on chosen configurations, including forces and/or stresses labels make more sense, especially when augmentation of information is performed by generating configurations from it by random atomic displacements. Here a strategy solely based on energy labels for the training is chosen since the data consists of sampled AIMD trajectories for each thermodynamic state whose accessible microstates explore, through the thermal fluctuations, their local basin of the PEL, taking implicitly their gradients into account [53].

2.3. Training the neural network

The supervised training is carried out using as input the BP feature vectors describing local atomic environments in each configuration. AIMD energies of these configurations are used to find the optimal set of weights and biases. The complete dataset of configurations is firstly randomized and scaled using the standard scaler of SCIKIT-LEARN, i.e. centering the feature components about their mean and normalize them according to their standard deviation. It is then split into a training set of 80% of the data and a test set containing the 20% remaining part. In the training set 20% of the data are retained further to create validation sets. They are used (a) for a cross-validation procedure to estimate the performance of various NN architectures through the MSE, and (b) to monitor the MSE on the validation data during the learning process to detect overfitting. For a given architecture, the optimization is performed

using the training data without the validation set, and terminating when the validation error starts to increase. Reduction of the noise of the MSE during training is obtained by including a callback with a stepwise reduction of the learning rate. Simultaneously, a L_2 norm regularization with strength 10^{-5} is performed to reduce the model complexity, and thus to prevent overfitting. Once trained, the weights and biases are stored in a format compatible with the LAMMPS HDNNP pair-style [35].

This training stage is repeated with various NN architectures to find the optimal one capturing at best the functional dependence of the data. Evaluation of the MSE is carried out through a stochastic gradient descent minimization using the Adam optimization algorithm [69] giving a measure of the loss with a learning rate starting at 0.01 and reducing most of the time to 0.0001 during the training, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The early-stopping was performed with maximum loss variation of 10^{-6} and a patience of 45 epochs. The typical duration of the training period was about 10 000 to 15 000 epochs. The least MSE loss is obtained for an architecture of ten neurons in the first and second hidden layers. A typical evolution of loss and the validation loss is shown in figure 2(a). A cross-validation performed over five independent trainings gives a RMSE of (1.2) meV on the per atom energy. Figure 2(b) displays the predictive ability of the model on the unseen data of the test set, with a high quality over the whole range of the energies.

The predictive ability of the HDNN is illustrated in figure 3 on the three force components extracted from AIMD configurations of a simulation at 1500 K over 2 ps, with a RMSE of $0.074 \text{ eV } \text{\AA}^{-1}$ at this high temperature. At $T = 10 \text{ K}$ in similar sampling conditions the RMSE reduces to $0.030 \text{ eV } \text{\AA}^{-1}$. These RMSE values are consistent with those obtained for previous trained ML potentials for Al for which forces were included explicitly in the training [51, 52]. Our results lead to similar conclusions for molecular systems [87] saying that the forces can be predicted with a good accuracy without being explicitly part of the learning process, and thus is in favor of the supervised learning strategy based only on the energy of the configurations, thus avoiding additional parameters in the loss function.

The HDNN is further tested on the prediction of the energy for configurations along a given AIMD trajectory. Reproducing fluctuations is important as they describe the derivatives of thermodynamic quantities, and for instance the specific heat as the derivative of the energy that will be considered in section 3.2. Consecutive configurations from AIMD simulation of liquids were considered at $T = 600 \text{ K}$ in the undercooled region, $T = 950 \text{ K}$ in the vicinity of the melting point, at 1500 K far above the melting point at zero pressure, and $T = 8000 \text{ K}$ just above the melting line for a pressure of 322 GPa, as shown in figure 4. Energy fluctuations are very well reproduced for all the temperatures, even for the extreme values for which the sampling is scarce, as they correspond to the tail of the Gaussian distribution of energy fluctuations. Probably the most impressive agreement is that of the simulation at $T = 8000 \text{ K}$ where the energy range of the fluctuation is as large as $0.25 \text{ eV atom}^{-1}$ and still very well predicted. It is

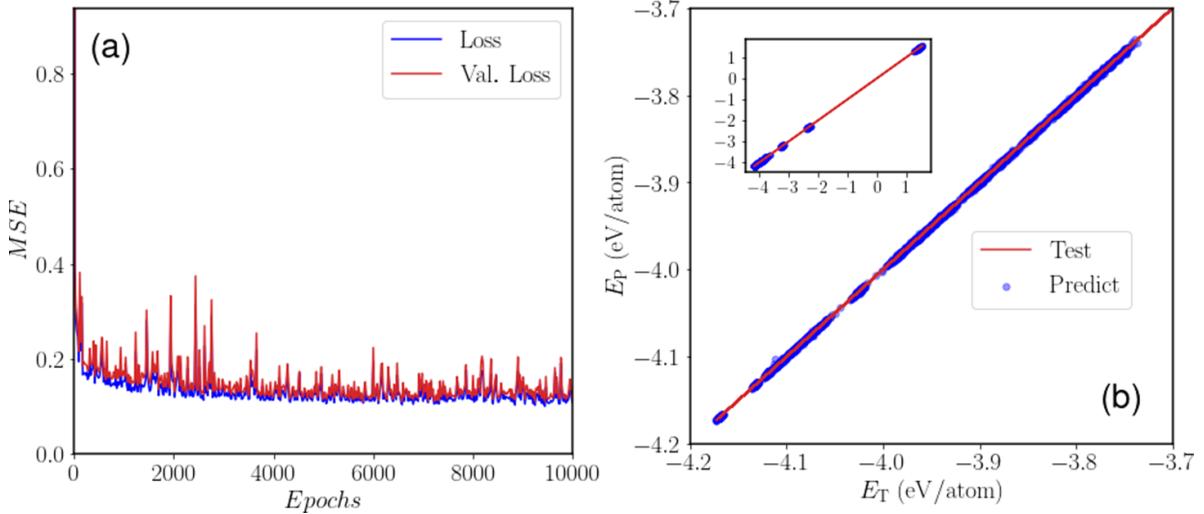


Figure 2. (a) Evolution of the MSE losses as a function of number of epochs for the training and validation sets. Loss and Val. Loss correspond to the evaluation of the MSE expressed in eV^2 on the training set and validation set, respectively. (b) Test-predict curve showing the quality of the prediction on the test set for the optimized architecture $22 \times 10 \times 10 \times 1$. The red solid line represents the known output of the energies representing the know $E_T = E_P$ line that would correspond to a perfect prediction, and the blue dots the values predicted against the known ones. The main panel corresponds to thermodynamic states at low pressures ($< 5 \text{ GPa}$) and temperatures between 10 K and 1500 K containing either crystalline and liquid configurations. The inset corresponds to energies in the full range of temperatures and pressures (see text).

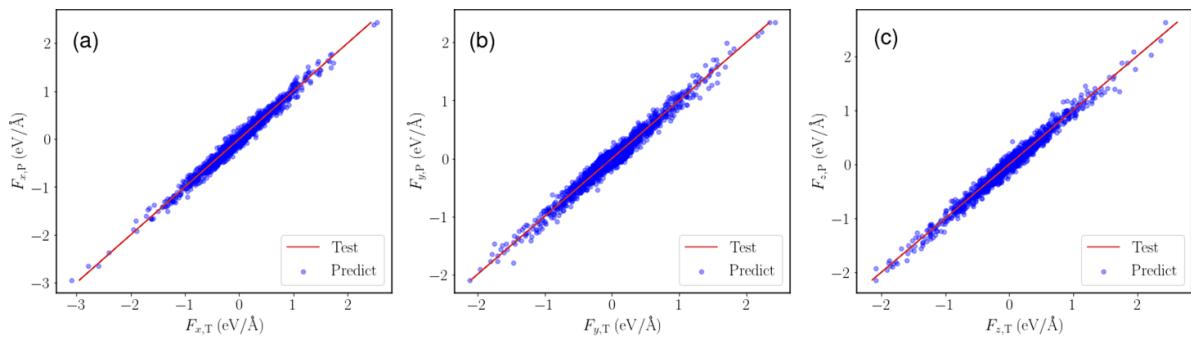


Figure 3. Test-predict curves for the three components of the predicted forces against those extracted from the AIMD configurations at $T = 1500 \text{ K}$ over 2 ps. The red solid line and blue dots have the same meaning as in figure 2.

worth mentioning that none of the configurations considered in figure 4 are included in the training set so that the energies for each temperature are purely predictive, and demonstrate the quality of the ML potential.

2.4. Molecular dynamics simulation

Classical MD simulations were carried out using the LAMMPS package [36]. These simulations were performed in various ensembles, namely the canonical ensemble (*NVT*), the isobaric-isothermal ensemble (*NPT*, i.e. constant temperature, pressure, and number of atoms), and the isobaric-isoenthalpic ensemble (*NPH*, i.e. constant pressure, enthalpy, and number of atoms). Temperature and pressure were kept constant via the Nose-Hoover thermostat and barostat [88, 89], respectively. In all simulations, PBC were employed in the three spatial directions. The integration of the equations of motion was done via Verlet's algorithm in the velocity form, choosing a time step of 1 fs. The use of our HDNN potential in LAMMPS was

possible through the library-based implementation of HDNN potentials by Singraber *et al* [35]. To assess the quality of our HDNN potential, some of the MD simulations were also repeated with the previously published ANI-AI potential of Smith *et al* [50].

Structural analysis is performed using the common-neighbor analysis (CNA) [90] with the indexing of Faken and Jonsson [91] and a bond-based algorithm as implemented in the OVITO software [92] where a uniform cut-off radius corresponding to the first minimum of the pair-correlation function of the liquid is applied to create bonds between pairs of particles. The CNA classifies pairs around each atom by sets of three indices: the first index represents the number of nearest-neighbors common to this pair, the second index corresponds to the number of nearest-neighbor bonds among the shared neighbors, and the third index indicates the longest chain of bonded atoms among them. For instance, 421 and 422 bonded pairs are characteristic of close packed structures fcc and hcp, respectively. The occurrence of 444 and 666 pairs, with

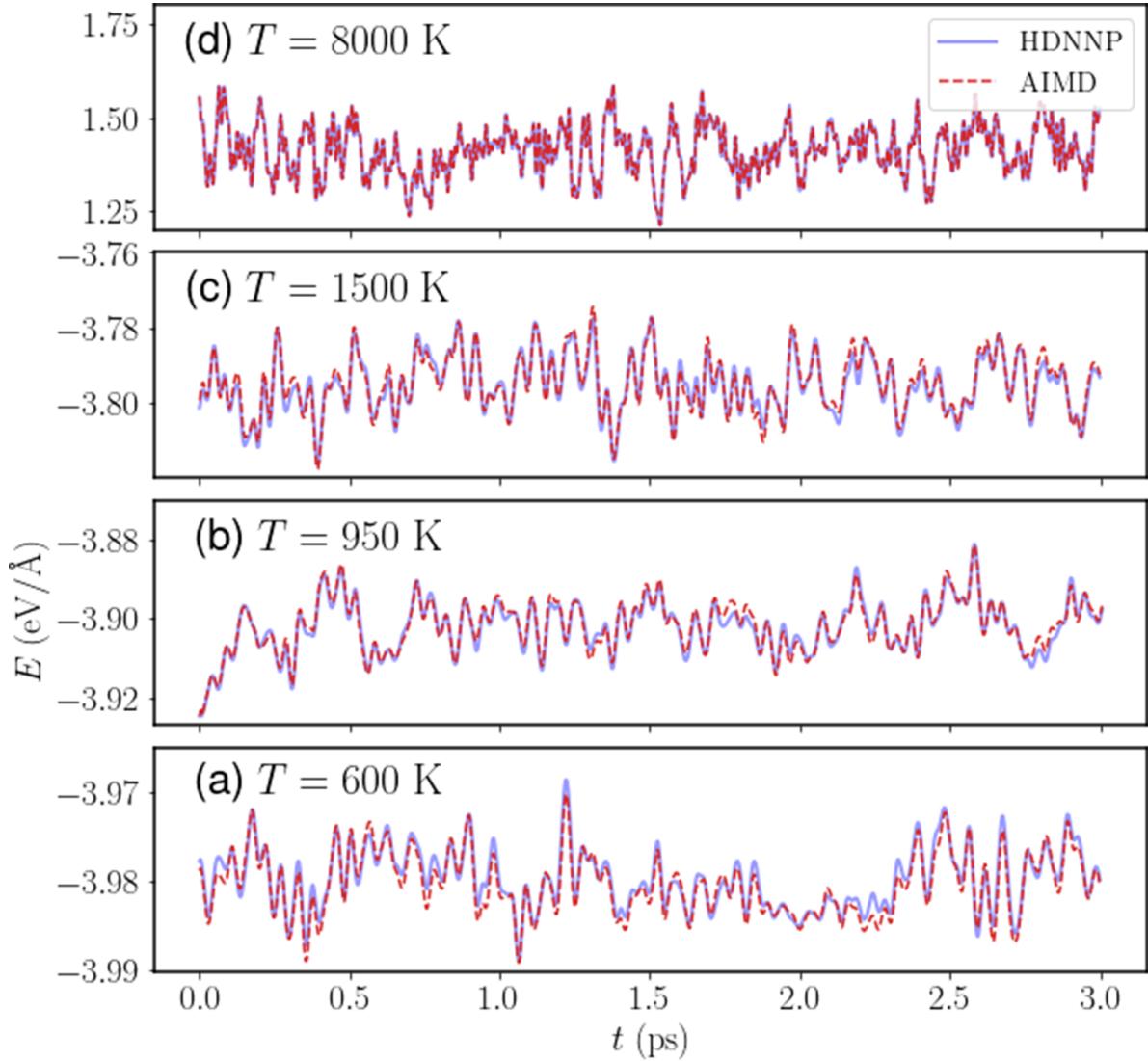


Figure 4. Energy per atom as a function of time from AIMD simulation of liquids and predicted by the HDNN potential with $N = 256$ atoms at $T = 600$ K in the undercooled region, $T = 950$ K in the vicinity of the melting point, at 1500 K far above the melting point at zero pressure, and $T = 8000$ K just above the melting line for a pressure of 322 GPa.

specific proportions, signals the presence of bcc ordering. The degree of five-fold symmetry is obtained from the proportion of 555, 554 and 433 pairs, which represent perfect (555) and distorted FFS based motifs.

An alternate way of studying the local ordering before and during nucleation is to make use of the Steinhardt bond-ordering parameters [93]. More specifically, the averaged form [94] as implemented in the PYSCAL code [95] is considered here. First, for each atom i , the following vector is defined

$$q_{lm}(i) = \frac{1}{N(i)} \sum_{j=1}^{N(i)} Y_{lm}(\mathbf{r}_{ij}) \quad (2)$$

where $N(i)$ is the number of nearest neighbors of atom i , \mathbf{r}_{ij} is the displacement of nearest-neighbor atom j from i , and Y_{lm} are

the spherical harmonics. From these, the averaged bond-order parameters can be defined as

$$\bar{q}_l(i) = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^l \left| \frac{1}{N(i)+1} \sum_{k=0}^{N(i)} q_{lm}(k) \right|^2} \quad (3)$$

where the sum from $k=0$ to $N(i)$ includes both the atom i and its nearest neighbors. Due to being averaged over nearest neighbors, these parameters take into account not just the first coordination shell, but also the second one. To perform structural analysis using these parameters one typically selects specific values for l , with $l=4$ and $l=6$ being a common choice. It is then possible to compare the resulting values of \bar{q}_l to those of ideal crystals in order to identify crystal structure.

3. Results and discussion

3.1. Local structure and dynamics

In a first step, the optimized HDNN potential is assessed on the local structure and dynamics. The simulations are performed at a constant volume, given in top row of table 1, with $N = 10976$ atoms at selected temperatures. For the undercooled states the system is first prepared at a temperature of 1500 K, before being cooled down to the desired temperature. The other states are prepared directly at the target temperature from a fully equilibrated liquid. Following an equilibration time of 10 ps relevant quantities are calculated over a production time ranging from 100 ps to 1 ns depending on the thermodynamic state under consideration.

The pair-correlation function $g(r)$ gives the probability of finding a particle j at distances r_{ij} relative to a particle i located at the origin, and reads:

$$g(r_{ij}) = \frac{N}{V} \frac{n(r_{ij})}{4\pi r_{ij}^2 \Delta r}, \quad (4)$$

where $n(r)$ represents the mean number of particles j in a spherical shell of radius r and thickness Δr centered on particle i . Finally, an average of $g(r)$ over all N particle i of the simulation box is performed. Integrating $4\pi r^2 \rho g(r)$, with $\rho = N/V$ up to the first minimum $g(r)$ gives access to the mean coordination number. Figure 5 displays the curves of $g(r)$ from NVT simulations. An excellent match with AIMD simulations is seen for all the thermodynamic states. A quantitative estimation of the deviation was obtained by calculating the MSE between the classical MD and AIMD curves for each temperature. The MSE ranges from 6×10^{-4} typically in the case of the liquid states to 1.2×10^{-2} in the case of fcc solid states. The larger MSE for the solid might come from the fact that even a very small position shift of $g(r)$ can induce a significant deviation as peaks are sharp and narrow. The curves of $g(r)$ obtained from the ANI-Al ML potential of Smith *et al* [50] are slightly shifted to larger distances for all liquid states considered, thus overestimating the bond lengths, and their peaks in the crystalline states are more pronounced. The average coordination numbers with the HDNN potential display a deviation from AIMD that does not exceed 0.2 (see table 1). Nevertheless, a comparison of $g(r)$ of the present HDNN potential to those obtained with published ML potentials is considered. Figure 5(c) shows that our potential leads to overall better results than the ML potential of Kruglov *et al* [51] as compared to their *ab initio* simulations and experimental data at $T = 1000$ K. In figure 5(d) our potential leads to results very close to the recent ANI-Al potential of [50], and in good agreement with experiments [96] at $T = 1123$ K, $T = 1183$ K and $T = 1273$ K. It is worth mentioning that both mentioned ML potentials have been trained using only the forces [51] or using the energies and forces [50], contrary to the present potential. Additional comparison with the most widely used EAM [63] and MEAM [60] potentials is shown in figure S1 in the SI file for the same thermodynamic states. They are shown to perform

less well than the HDNN potential, as assessed by a t -statistics, and especially for the high pressures.

Beside the local structural properties, dynamic properties represent a stringent test as they are even more sensitive to the details of the potentials. Among these, diffusion plays an important role in the solidification process [12, 97] and was evaluated here through the mean-square displacement (MSD)

$$R^2(t) = \frac{1}{N} \sum_{l=1}^N \left\langle [\mathbf{r}_l(t + t_0) - \mathbf{r}_l(t_0)]^2 \right\rangle_{t_0}, \quad (5)$$

where $\mathbf{r}_l(t)$ denotes the position of atom l at time t and N is the number of atoms. In addition to the mean over all atoms, an averaging over time origins t_0 as indicated by the angular brackets is performed. The self-diffusion coefficient D is determined from the slope of the linear behavior at long times of the MSD. In figure 6, the MSD is shown for temperatures in the stable and undercooled liquid states at ambient pressure as well as for temperatures along the melting line for pressures up to 340 GPa. The overall trends of AIMD curves are well reproduced by the HDNN potential for all temperatures and pressures as can be seen in table 1. It is worth mentioning that our potential gives a better prediction of the diffusion coefficients than the ANI-Al potential at ambient pressure and low temperatures where solidification phenomena occur. This is mainly due to the fact that the ANI-Al potential was trained with the PBE functional which is known to underestimate the diffusion coefficient of aluminium as shown in [10] and this also consistent with the overestimation of the bond length at a given density mentioned above. The MSD curves show a ballistic regime at very short times ($t < 0.05$ ps), followed by a diffusive regime at long times. For the lower temperatures at ambient pressures and at high pressures a well-known caging effect [97] takes place after the ballistic motion and delays the diffusive regime, which is well predicted by the HDNN potential with respect to the AIMD.

The collective dynamics is examined by means of the Intermediate Scattering Function (ISF) $F(Q, t)$ and its time Fourier transform $S(Q, \omega)$, the dynamic structure factor that can be measured by means of Neutron Diffraction. Q represents the wave-number and ω the frequency. Figure 6(c) shows the good agreement of the ISF between the HDNN potential and the AIMD results in the liquid state at $T = 950$ K and $T = 1300$ K for $Q_0 = 2.65 \text{ \AA}^{-1}$ corresponding to the position of the first maximum of the static structure factor $S(Q)$. The good match is confirmed for $S(Q, \omega)$ in figure 6(d) for both temperatures. Further, the comparison with Neutron diffraction data [74] demonstrates that the *ab initio* calculations as well as the HDNN potential predict the dynamic properties of liquid aluminium quite accurately.

The real predictive character of the HDNN potential is assessed for $T = 950$ K, $T = 2000$ K and $T = 8000$ K at the lowest pressure: These thermodynamic states were not included in the training, but the accuracy is similar to the other states for the pair-correlation function and the diffusion. This is all the more true for results at $T = 1123$ K, $T = 1183$ K,

Table 1. Atomic volume V , pressure P , coordination number N_C , first minimum r_1 of $g(r)$, and diffusion coefficient D for selected temperatures. Values in second rows are from the AIMD, and those in parenthesis are the diffusion coefficient calculated from the ANI-Al potential [50]. Classical MD simulations for the diffusion were performed with $N = 256$ atoms as for the AIMD one in order to have similar finite size effects. (H), (L), and (s) specifies high and low pressure, and solid state simulations respectively.

T (K)	300 (s)	600	800 (s)	950	1500	2000	4500	8000 (H)	8000 (L)
V (\AA^3)	16.48	18.06	16.68	18.89	20.37	14.00	10.67	7.629	8.616
P (GPa)	0	0	0	0	0	30	110	240	340
	0	0	0	0	0	27	107	227	320
N_C	—	11.93	—	11.57	11.10	12.22	12.70	12.70	12.65
	—	12.05	—	11.65	11.17	12.32	12.80	12.80	12.75
r_1	—	3.727	—	3.775	3.850	3.448	3.225	3.875	3.000
	—	3.752	—	3.800	3.825	3.548	3.225	2.950	3.075
D ($\text{\AA}^2 \text{ps}^{-1}$)	—	0.15	—	0.63	1.69	0.35	0.78	1.01	1.35
	—	0.14	—	0.62	1.61	0.37	0.68	0.93	1.34
	—	(0.10)	—	(0.46)	(1.54)	(0.33)	(0.66)	(0.96)	(1.38)

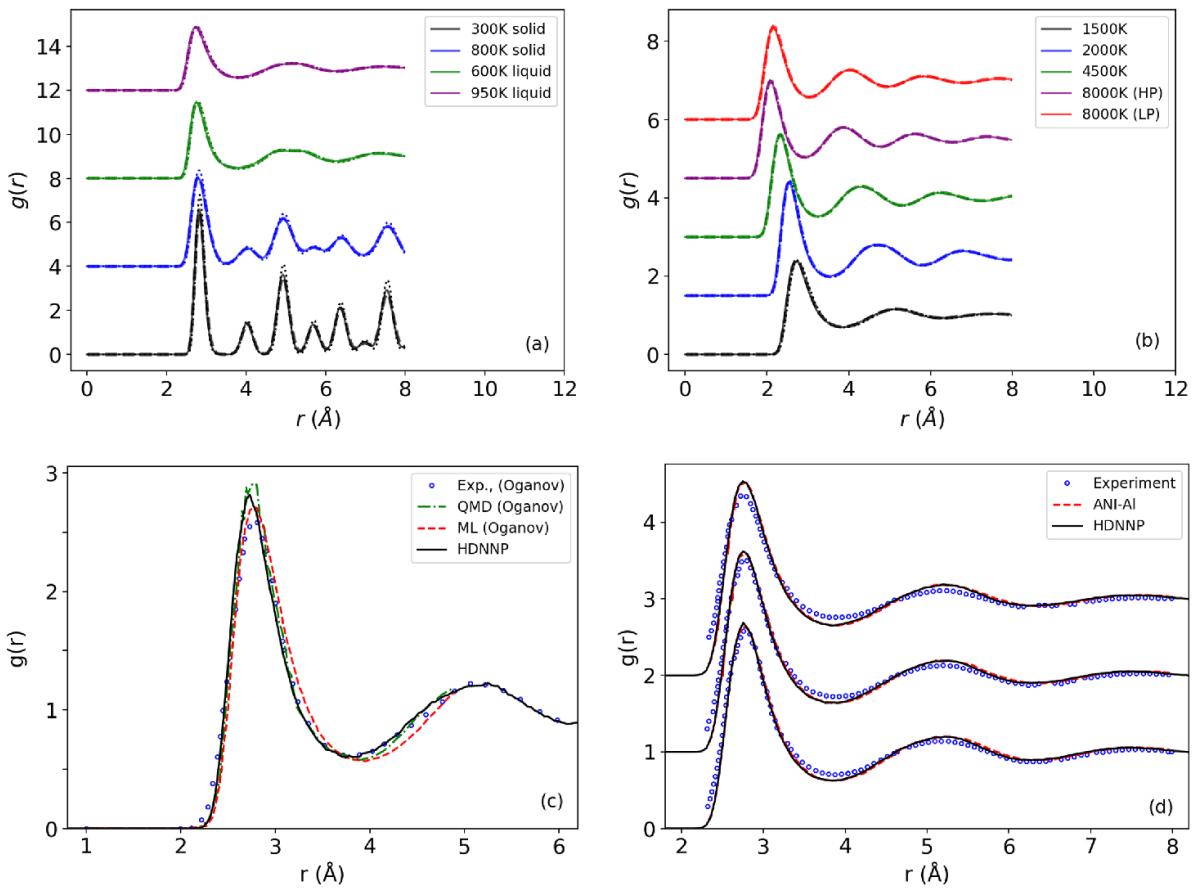


Figure 5. Pair-correlation function for various temperatures and pressures (a) for low temperature liquid and solid states at room pressure and (b) for high temperature and pressure liquid states. Curves for 800 K, 600 K, 950 K, 2000 K, 4500 K, 8000 K (340 GPa) and 8000 K (240 GPa) are shifted upwards by an amount of 4, 8, 12, 1.5, 3, 4.5 and 6, respectively. The solid lines are results with the HDNNP potential, the dotted lines to the ANI-Al potential, and the dashed lines with corresponding colors are those of the AIMD simulations. High-temperature simulations ran for 500 ps (4500 K, 8000 K (HP)) and 250 ps (8000 K (LP)), with others running for 1 ns. (c) Comparison in the liquid at $T = 1023$ K to simulations with the ML potential of [51] as well as their *ab initio* calculations and experimental data. (d) Comparison in the liquid at $T = 1123$ K, $T = 1183$ K and $T = 1273$ K to simulations with the ANI-Al potential of [50] as well as experimental data of [96].

$T = 1273$ K and $T = 1300$ K for the comparison with other ML potentials and experiments in figures 5 and 6, as well as for the pressure for all states shown in table 1 with a deviation less than 7% at the highest ones, which is remarkable since the forces were not included in the training.

3.2. Thermodynamic properties

Important quantities for solidification phenomena are the latent heat of fusion [98, 99] as well as the densities of the solid and liquid phases at the melting temperature, T_M .

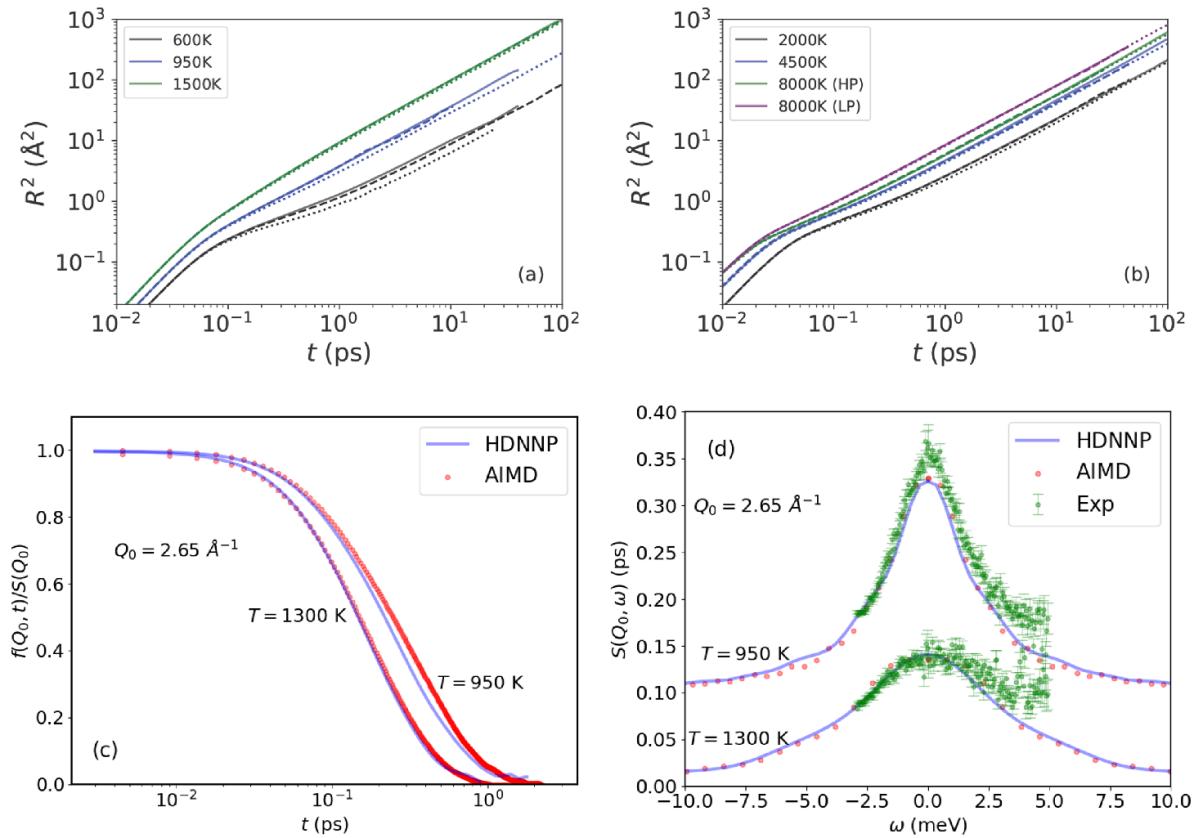


Figure 6. Mean-square displacement for various temperatures and pressures: (a) for high temperature and pressure liquid states and (b) for low temperature liquid and solid states at room pressure. The solid lines correspond to the HDNN potential, the dotted lines to the ANI-Al potential [50], and the dashed lines are the AIMD simulations. Classical MD simulations for the MSD were performed with $N = 256$ atoms as for the AIMD one in order to have similar finite size effects. (c) Intermediate scattering function in the liquid state at $T = 950$ K and $T = 1300$ K for wave vector $Q_0 = 2.65 \text{ \AA}^{-1}$. (d) Corresponding dynamic structure factor at the same temperatures and wavevector that are compared to the neutron diffraction experiments at $T = 943$ K and $T = 1293$ K from [74].

Its determination requires the calculation of the enthalpy difference between the liquid and solid branches at T_M . The temperature evolution of the enthalpy at ambient pressure for the solid and liquid branches of the HDNN potential are shown in figure 7(a), obtained via simulation of $N = 2048$ atoms in the *NPH* ensemble. The simulation is started with a perfect fcc crystal at $T = 300$ K and heated stepwise with a temperature step of 50 K with an average heating rate of 10^{12} K s^{-1} . At each temperature, a simulation is performed over 50 ps (25 ps equilibration and 25 ps production) during which an average value of the enthalpy is calculated. The increase of temperature is repeated until a dynamic melting is observed at $T = 1250$ K. The latter value is noticeably higher than the thermodynamic melting temperature $T_M = 970$ K obtained from liquid–solid interface (LSI) simulations (see section 3.3) due to overheating effects. For the liquid branch, the simulations are started at $T = 1600$ K with an equilibrated configuration after the heating process. The same procedure as for the solid branch is followed but with a step-wise cooling down to 300 K that is called here the slow cooling. Above $T = 1250$ K, the difference in the enthalpy from the heating and cooling processes is negligible, indicating that there is no reminiscence of the crystalline state. Below 600 K the liquid undergoes a partial crystallization during cooling. Then, the cooling procedure

for the liquid branch is repeated with a higher cooling rate of 10^{13} K s^{-1} to avoid crystallization, and a glass transition is seen at $T_G = 378$ K inferred from a crossover between the liquid and glassy branches as shown in figure 7(a).

For such a high cooling rate of 10^{13} K s^{-1} , the time to reach each temperature is shorter than its corresponding nucleation time. This is illustrated from the time temperature transition (TTT) curve plotted in figure 7(b). It is determined for a system of approximately one million atoms (see section 3.4) by measuring the time it takes until 40% of the atoms are identified as in a solid state by CNA. At 600 K, the measurement is repeated from the initial configuration, but with new velocities picked from a proper Maxwell distribution, to estimate the variability between measurements.

From the liquid and solid branches an enthalpy of melting of $11.67 \text{ kJ mol}^{-1}$ is found, which compares reasonably well to the experimental value of $13.34 \text{ kJ mol}^{-1}$ [100]. Taking the numerical derivative of the solid branch yields a value of the specific heat at constant pressure, C_P , of $0.99 \text{ J g}^{-1} \text{ K}^{-1}$ which is also in good agreement with the experimental value of $0.91 \text{ J g}^{-1} \text{ K}^{-1}$. For the liquid a value of $1.158 \text{ J g}^{-1} \text{ K}^{-1}$ is obtained, which is in the range of experimental data between $1.03 \text{ J g}^{-1} \text{ K}^{-1}$ and $1.18 \text{ J g}^{-1} \text{ K}^{-1}$ close to the latest assessed values of $1.127 \text{ J g}^{-1} \text{ K}^{-1}$ [100]. The specific heat is a typical

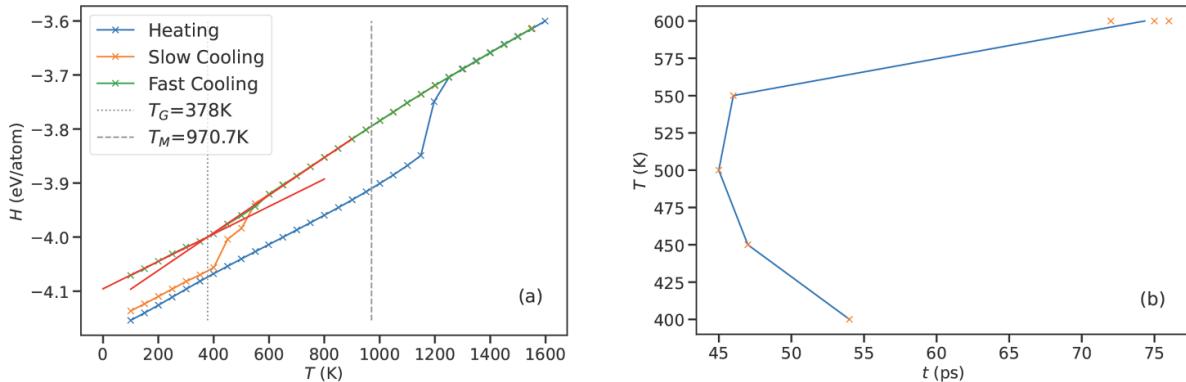


Figure 7. (a) Enthalpy as a function of temperature for the solid (from heating) and liquid branches (from cooling) at ambient pressure as described in the text. The red lines indicate the slope of the amorphous and the supercooled states as a guide for the eyes for the crossover between the liquid and amorphous regimes, and marked by the vertical dotted line. The vertical dashed line marks the melting temperature obtained from the LSI simulations. (b) The TTT curve, as described in the text. The orange crosses are individual measurements of the time it takes to reach 40% solidification, with the blue line connecting the averages for each temperature.

derivative quantity that depends on the fluctuations of the enthalpy [88, 89]. The very good agreement is a strong indication that including the time fluctuations from AIMD is a fruitful strategy to describe at least the thermodynamics.

Regarding densities, the HDNN potential gives a value of 0.0587 \AA^{-3} and 0.0547 \AA^{-3} , respectively in the solid and liquid at its melting point $T_M = 970\text{ K}$, giving rise to a density change of 0.004 \AA^{-3} . These values compare well to the respective experimental values [81, 100, 101] of 0.0573 \AA^{-3} and 0.05306 \AA^{-3} with a density change of 0.0042 \AA^{-3} . At the experimental melting temperature, the calculated density change remains essentially unchanged, and the densities in both phases deviate only by 2% with respect to the measurements.

3.3. Liquid-solid interfaces

Liquid-solid interface simulations are performed for the purpose of determining the melting line by the two-phase coexistence. The procedure follows the approaches proposed in [102–110] and is similar to the protocol used in [26]. A simulation cell containing $N = 5488$ atoms is set up with an initial crystalline configuration with a shape corresponding to $28 \times 7 \times 7$ primitive cells on which PBC are applied to the three directions of space. Starting at zero pressure, this system is heated and equilibrated at constant pressure to a temperature of 50 K below a guess of the melting temperature. Half of the simulation cell in the x direction is further heated and maintained at a much higher temperature until a complete melting is observed. The liquid part is then cooled down and equilibrated at a temperature 50 K above the guess, thus creating a solid-liquid coexistence containing two crystal-melt interfaces due to the PBC. The simulation of the entire system is pursued in the isobaric-isoenthalpic ensemble so that the temperature of the LSI is an internal parameter free to evolve toward a steady state corresponding to the thermodynamic melting temperature if both phases survive. The simulation is continued for 1 ns,

and the average melting temperature is determined on the last 100 ps when a steady position of the two interfaces is observed. If a complete melting or solidification occurs, the procedure is started over again with a refined guess of the melting temperature. This procedure is repeated with subsequent higher pressures by first shrinking the volume of the whole simulation cell from the coexistence configuration at the preceding pressure and then increasing the temperature at constant pressure to a new guess of the melting line.

The melting curve of aluminium was measured [111, 112] up to 80 GPa using diamond anvil cells and even higher at 125 GPa by means of shock experiments [113]. In figure 8, the results obtained from the HDNN are compared to these experimental data [111–113], the *ab initio* based equation of states (EOSs) [85] as well as the AIMD two-phase approach [114] for which the GGA for the XC functional and 512 atoms were used. At ambient pressure, the HDNN potential yields a value of $T_M = 970\text{ K}$ which overestimates the experimental one of 933 K by 5%. This is also the case for the AIMD [114] to a lesser extent, recalling that the GGA was used and overestimates the atomic volume [82, 114]. With increasing pressure, the melting curve from the HDNN potential slightly underestimates the experiments as well as the EOS. By using two different sizes and shapes, negligible influence on the determination of the melting curve was found, confirming earlier results on pure Zr [26]. Noticeably, the reliability of the present potential on the melting line up to 200 GPa is then assessed, even if high pressure thermodynamic states included in the training set are really scarce. Interestingly, the HDNN curve is similar to the one obtained with the ANI-AI ML potential by Smith *et al* [50].

3.4. Homogeneous nucleation

Finally, the homogeneous nucleation is investigated and depicted in figure 9. As pointed out in our preceding work [26], a more accurate investigation require the use of large enough

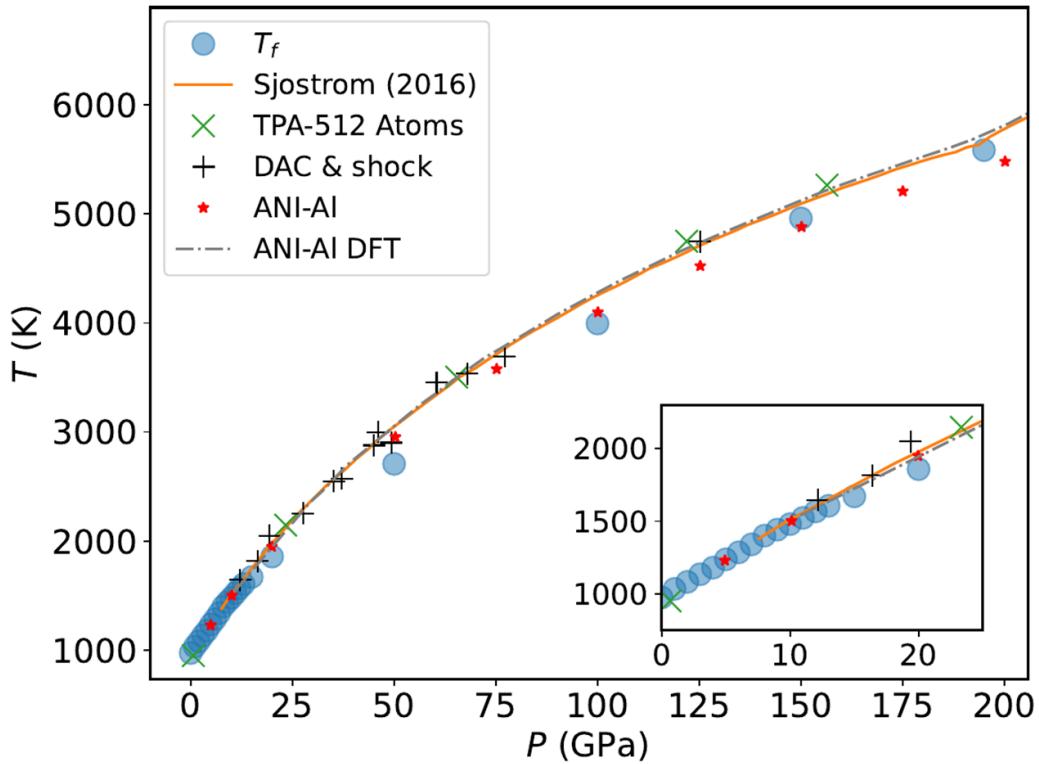


Figure 8. Melting curve for aluminium as a function of pressure. The blue circles obtained from our LSI simulations with the HDNN are compared to the experimental data [111–113], the AIMD two-phase approach [114], the ANI-AI ML potential [50], and the equation of state of Sjostrom *et al* [85]. The Inset highlights the lower pressure range up to 25 GPa.

simulation boxes, with typically one million atoms or more. This allows the occurrence of multiple nuclei during the nucleation process. Therefore, the system of 1000 188 atoms at ambient pressure is analyzed further along the 600 K isotherm used for the determination of the TTT curve in figure 7(b). At such a high degree of undercooling $\Delta T = (T_M - T)/T_M = 0.38$, an extremely fast nucleation process is observed [98, 99]. Similarly, such a fast homogeneous nucleation is seen at the high pressure of 200 GPa along the 4000 K with $\Delta T = 0.24$. Inherent structure configurations shown in figures 9(c)–(j) were first analyzed using the CNA [91] and only atoms with a crystalline environment (fcc, hcp, and bcc) are shown. As expected, nucleation occurs showing growing nuclei in the fcc ordering with hcp stacking faults at ambient pressure. At 200 GPa, nucleation starts with nuclei having a bcc order with sometimes some fcc ordering at their boundary that transform back to the bcc structure during the growth. These nucleation pathways are pretty much consistent with the (P, T) phase diagram [85] showing the reliability of the HDNN potential designed here.

The onset of nucleation occurs at about 30 ps at ambient pressure and 4 ps at 200 GPa, earlier than the nucleation time defined for the construction of the TTT curve above. Using the averaged Steinhardt Order Parameters q_4 and q_6 [115], embryos with atoms having a crystalline ordering showed that they dissolve back to the liquid with a size less than 90 atoms in both cases. The latter value does not represent *per se* the

size of the critical nucleus but rather a lower limit. As expected at ambient pressure, the main crystalline phase during the growth was identified as fcc, as can be seen in figure 10(a), but a significant hcp ordering also appears during the nucleation and remains as stacking fault after complete solidification of the simulation box as shown in figure 9. At high pressure, the onset of nucleation occurs in the bcc ordering, with fcc and hcp ordering at their surface at later stages.

It is an open question in general, whether the homogeneous nucleation process follows the Landau Theory in which the bcc precursor is favored in the early stages of crystal nucleation [116] or the Ostwald step rule [117] according to which a primary crystal phase could be different from the fcc one. From the distributions of the q_6 and q_4 shown in figures 10(b) and (c) only fcc ordering emerges at the onset of nucleation. Our findings show that aluminium follows a single step process with an onset of homogeneous nucleation showing emerging embryos with a fcc ordering. The resulting nuclei grow in a rather patchy shape with a small amount of hcp stacking fault defects. This nucleation scenario is different from the Lennard-Jones case [118, 119] which follows the Landau theory and the Ostwald step rule. The present large-scale molecular dynamics results with close *ab initio* accuracy allows us to assess very recent molecular dynamics simulations [120, 121] with EAM empirical potentials. Our findings further show that such a single step nucleation pathway also occurs at high pressure with bcc ordering in the emerging nuclei.

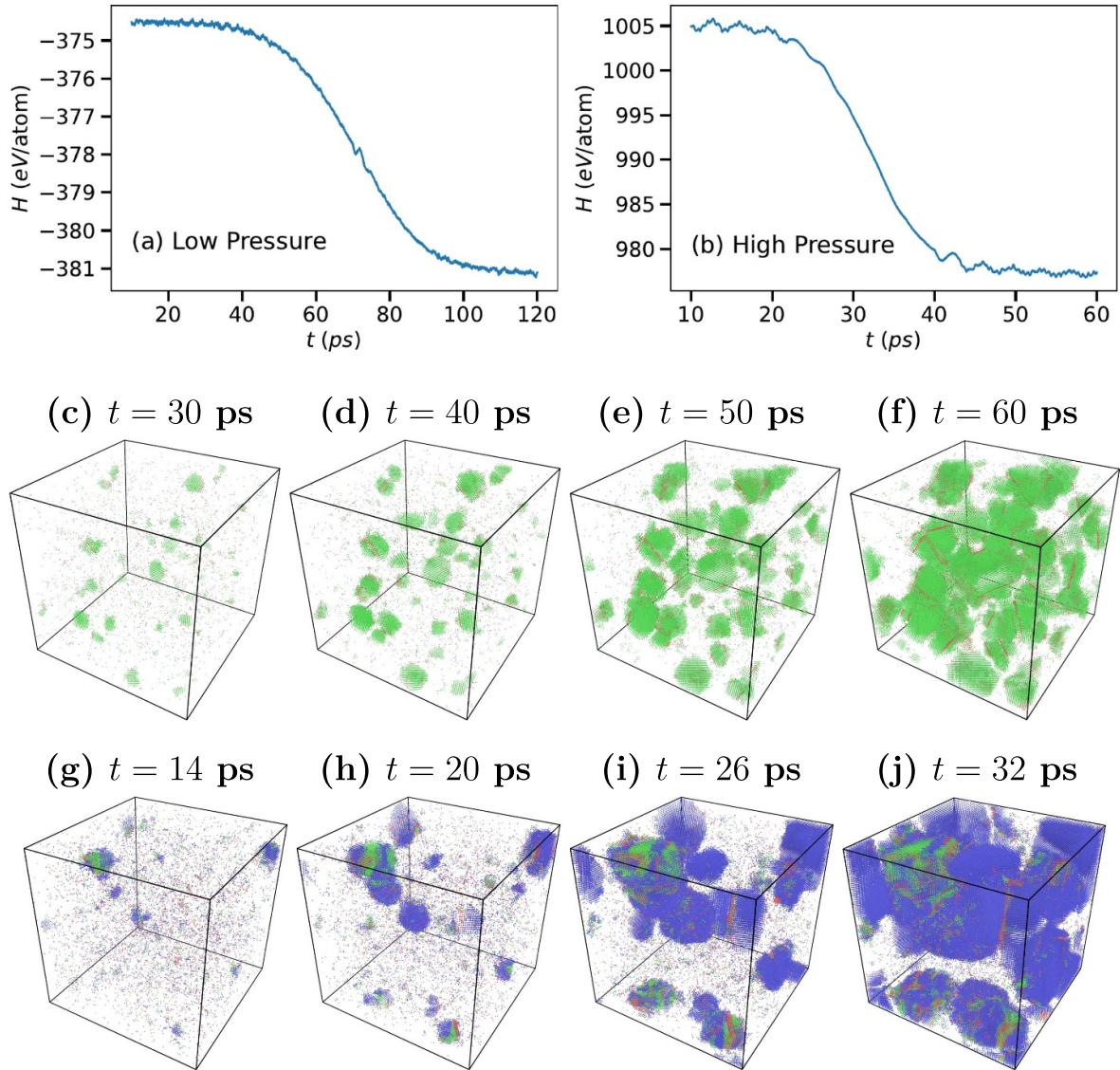


Figure 9. Homogeneous nucleation of deeply undercooled aluminium along the $T = 600$ K and 4000 K isotherms, respectively for ambient pressure and 200 GPa. Time evolution of the enthalpy (a) at ambient pressure, and (b) at 200 GPa. Snapshots of the simulation as various times during the nucleation (c) to (f) at ambient pressure, and (g) to (j) at 200 GPa. Only the atoms with crystalline ordering in the sense of the common-neighbor analysis are drawn: fcc (green); hcp (red); bcc (blue).

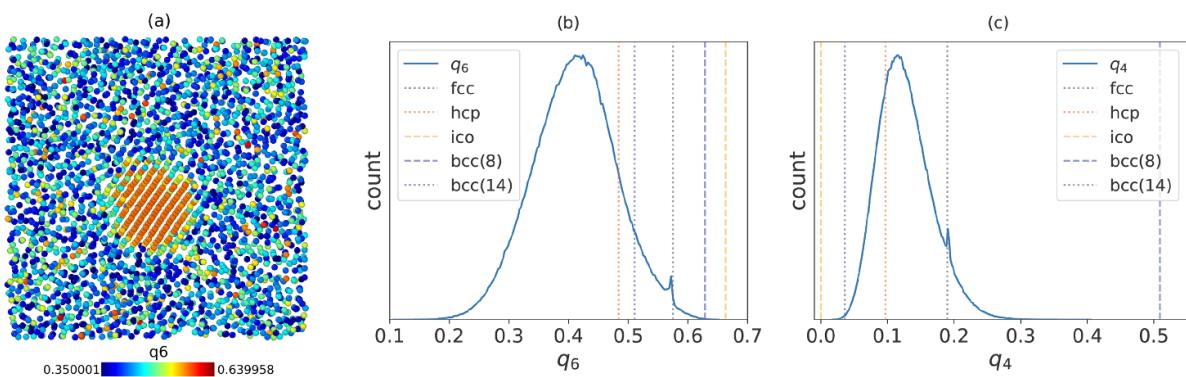


Figure 10. (a) First identified nucleus at the onset of nucleation in a small box extracted from the snapshot at 30 ps (figure 9) for ambient pressure. Atoms are colored according to the value of the q_6 averaged Steinhardt order parameter. (b)–(c) Corresponding histograms for all atoms in the simulation box of the q_6 and q_4 parameters showing clearly that nucleation starts with only a fcc ordering.

4. Conclusion

In the present work, a machine learning potential for pure aluminium by means of a HDNN on the basis of the well-known Behler-Parrinello approach [35, 41] was developed. This ML potential is devoted to the description of condensed phases, namely liquid and solid states at ambient pressure as well as those at pressures up to 300 GPa with resulting temperatures as high as 8000 K. A crucial point was the training of the potential with a data set generated by DFT-based simulations not only to cover the targeted domain of thermodynamic states for a question of transferability but also to consider for each of them in a physical meaningful manner the time fluctuations by an appropriate sampling of phase space trajectories obtained by *ab initio* molecular dynamics. This allows to include in the training the relevant accessible microstates of the considered thermodynamic states. Another approach based on metadynamics was shown to be efficient in selecting the relevant configurations to train the NN [6].

The HDNN potential thus obtained was shown to be efficient in reproducing the structural, dynamics as well as thermodynamic quantities in the liquid, undercooled and crystalline states at ambient pressures as well as in the liquid state at high pressure up to 300 GPa, including the melting line. One important outcome is that a reliable ML potential could be obtained without including explicitly the forces in the training by using an appropriate sampling of AIMD trajectories. The procedure was shown for Al and HDNN to perform well, giving a RMSE on forces similar to what is currently obtained. The early stages of the homogeneous crystal nucleation was further investigated on a scale much larger than what is possible from the *ab initio* molecular dynamics but with a similar accuracy. Results show that aluminium follows a single step nucleation process with an emerging fcc ordering and hcp stacking fault defects, confirming recent works using large scale molecular dynamics [120, 121], and also consistent with very recent simulations on nucleation during cooling [122]. A single step nucleation pathway with bcc nuclei is also observed at high pressure.

Finally, the fact that the HDNN potential keeps a good accuracy even in domains where the thermodynamic states in the training set are scarce opens up a research line based upon active learning for regression approaches to reduce efficiently the training set. Dynamical properties such as the diffusion coefficients considered here are sensitive to the details of the potential and should be introduced in the training procedure in a more direct way than through the choice of the XC functional in the DFT calculations. This would represent a real step forward in designing ML potentials. Extending the present methodology to other systems is under progress.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

We acknowledge the CINES and IDRIS under Project No. INP2227/72914, as well as CIMENT/GRICAD for computational resources. This work was performed within the framework of the Centre of Excellence of Multifunctional Architected Materials ‘CEMAM’ ANR-10-LABX-44-01 funded by the ‘Investments for the Future’ Program. This work has been partially supported by MIAI@Grenoble Alpes (ANR-19-P3IA-0003). Fruitful discussions within the French collaborative networks in high-temperature thermodynamics GDR CNRS 3584 (TherMatHT) and in artificial intelligence in materials science GDR CNRS 2123 (IAMAT) are also acknowledged. We thank J Smith and K Barros for their kind help in setting up the simulations with their ANI-Al potential from [50]. J S acknowledges funding from the German Academic Exchange Service (DAAD) through the DLR-DAAD programme, Grant No. 509.

ORCID iDs

Noel Jakse  <https://orcid.org/0000-0002-4031-0965>
 Leon F Granz  <https://orcid.org/0000-0001-7096-0419>
 Thomas Voigtmann  <https://orcid.org/0000-0002-1261-9295>

References

- [1] Davis J R 2001 *Aluminum and Aluminum Alloys* (Novelty, OH: ASM International) p 351
- [2] Hafner J 2008 *J. Comput. Chem.* **29** 2044
- [3] Royall C P and Williams S R 2015 *Phys. Rep.* **560** 1
- [4] Kelton K F and Greer A L 2010 *Nucleation in Condensed Matter: Applications in Materials and Biology* (Oxford: Pergamon)
- [5] Jakse N and Pasturel A 2007 *Phys. Rev. Lett.* **99** 2
- [6] Bonati L and Parrinello M 2018 *Phys. Rev. Lett.* **121** 265701
- [7] Payne M C, Teter M P, Allan D C, Arias T and Joannopoulos J 1992 *Rev. Mod. Phys.* **64** 1045
- [8] Burke K 2012 *J. Chem. Phys.* **136** 150901
- [9] Car R and Parrinello M 1985 *Phys. Rev. Lett.* **55** 2471
- [10] Jakse N and Pasturel A 2013 *Sci. Rep.* **3** 3135
- [11] Pasturel A and Jakse N 2018 *J. Appl. Phys.* **123** 145103
- [12] Sosso G C, Chen J, Cox S J, Fitzner M, Pedevilla P, Zen A and Michaelides A 2016 *Chem. Rev.* **116** 7078
- [13] Hafner J 1987 *From Hamiltonians to Phase Diagrams* (Berlin: Springer)
- [14] Ashcroft N W 1966 *Phys. Lett.* **23** 48
- [15] Wills J M and Harrison W A 1983 *Phys. Rev. B* **28** 4363
- [16] Moriarty J A 1990 *Phys. Rev. B* **42** 1609
- [17] Jakse N and Bretonnet J-L 1995 *J. Phys.: Condens. Matter* **7** 3803
- [18] Belashchenko D K 2013 *Phys.-Usp.* **56** 1176
- [19] Daw M S and Baskes M I 1984 *Phys. Rev. B* **29** 6443
- [20] Daw M S, Foiles S M and Baskes M I 1993 *Mater. Sci. Rep.* **9** 251
- [21] Baskes M I 1992 *Phys. Rev. B* **46** 2727
- [22] Huang H S, Ai L Q, van Duin A C T, Chen M and Lü Y J 2019 *J. Chem. Phys.* **151** 094503
- [23] Pettifor D G 1996 *Bonding and Structure of Molecules and Solids* (Oxford: Oxford University Press)

- [24] Zong H, Pilania G, Ding X, Ackland G J and Lookman T 2018 *npj Comput. Mater.* **4** 48
- [25] Goryaeva A M, Lapointe C, Dai C, Dérès J, Maillet J-B and Marinica M-C 2020 *Nat. Commun.* **11** 4691
- [26] Becker S, Devijver E, Molinier R and Jakse N 2020 *Phys. Rev. B* **102** 104205
- [27] Behler J 2015 *Int. J. Quantum Chem.* **115** 1032
- [28] Behler J 2016 *J. Chem. Phys.* **145** 170901
- [29] Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A and Kim C 2017 *npj Comput. Mater.* **3** 54
- [30] Schmidt J, Marques M R G, Botti S and Marques M A L 2019 *npj Comput. Mater.* **5** 83
- [31] Goryaeva A M, Maillet J B and Marinica M C 2019 *Comput. Mater. Sci.* **166** 200
- [32] Mueller T, Hernandez A and Wang C 2020 *J. Chem. Phys.* **152** 050902
- [33] Marques M R G, Wolff J, Steigemann C and Marques M A L 2019 *Phys. Chem. Chem. Phys.* **21** 6506
- [34] Singraber A, Morawietz T, Behler J and Dellago C 2019 *J. Chem. Theory Comput.* **15** 3075
- [35] Singraber A, Behler J and Dellago C 2019 *J. Chem. Theory Comput.* **15** 1827
- [36] Thompson A P et al 2022 *Comput. Phys. Commun.* **271** 108171
- [37] Larsen A H et al 2017 *J. Phys. Condens. Matter* **29** 273002
- [38] Li Z, Kermode J R and De Vita A 2015 *Phys. Rev. Lett.* **114** 096405
- [39] Jinnouchi R, Karsai F and Kresse G 2019 *Phys. Rev. B* **100** 014105
- [40] Thompson A, Swiler L, Trott C, Foiles S and Tucker G 2015 *J. Comput. Phys.* **285** 316
- [41] Behler J and Parrinello M 2007 *Phys. Rev. Lett.* **98** 146401
- [42] Bartók A P and Csányi G 2015 *Int. J. Quantum Chem.* **115** 1051
- [43] Bartók A P, Payne M C, Kondor R and Csányi G 2010 *Phys. Rev. Lett.* **104** 136403
- [44] Botu V, Batra R, Chapman J and Ramprasad R 2017 *J. Phys. Chem. C* **121** 511
- [45] Morawietz T, Singraber A, Dellago C and Behler J 2016 *Proc. Natl Acad. Sci. USA* **113** 8368
- [46] Sosso G C, Miceli G, Caravati S, Giberti F, Behler J and Bernasconi M 2013 *J. Phys. Chem. Lett.* **4** 4241
- [47] Zhang C, Sun Y, Wang H-D, Zhang F, Wen T-Q, Ho K-M and Wang C-Z 2021 *J. Phys. Chem. C* **125** 3127
- [48] Pasturel A and Jakse N 2017 *npj Comput. Mater.* **3** 33
- [49] Russo J and Tanaka H 2016 *J. Chem. Phys.* **145** 211801
- [50] Smith J S et al 2021 *Nat. Commun.* **12** 1257
- [51] Kruglov I, Sergeev O, Yanilkin A and Oganov A R 2017 *Sci. Rep.* **7** 8512
- [52] Bochkarev A S, van Roekeghem A, Mossa S and Mingo N 2019 *Phys. Rev. Mater.* **3** 093803
- [53] Wales D J 2003 *Energy Landscapes* (Cambridge: Cambridge University Press)
- [54] Girifalco L A and Weizer V G 1959 *Phys. Rev.* **114** 687
- [55] Jacobsen K W, Norskov J K and Puska M J 1987 *Phys. Rev. B* **35** 7423
- [56] Mei J and Davenport J W 1992 *Phys. Rev. B* **46** 21
- [57] Ercolelli F and Adams J B 1994 *Europhys. Lett.* **26** 583
- [58] Mishin Y, Farkas D, Mehl M J and Papaconstantopoulos D A 1999 *Phys. Rev. B* **59** 3393
- [59] Sturgeon J B and Laird B B 2000 *Phys. Rev. B* **62** 14720
- [60] Lee B-J, Shim J-H and Baskes M I 2003 *Phys. Rev. B* **68** 144112
- [61] Liu X-Y, Ercolelli F and Adams J B 2004 *Modelling Simul. Mater. Sci. Eng.* **12** 665
- [62] El Mendoub E B, Albaki R, Charpentier I, Bretonnet J-L, Wax J-F and Jakse N 2007 *J. Non-Cryst. Solids* **353** 3475
- [63] Mendelev M I, Kramer M J, Becker C A and Asta M 2008 *Phil. Mag.* **88** 1723
- [64] Winey J M, Kubota A and Gupta Y M 2009 *Modelling Simul. Mater. Sci. Eng.* **17** 055004
- [65] Zhakhovskii V V, Inogamov N A, Petrov Y V, Ashitkov S I and Nishihara K 2009 *Appl. Surf. Sci.* **255** 9592
- [66] Choudhary K, Liang T, Chernatynskiy A, Lu Z, Goyal A, Phillpot S R and Sinnott S B 2015 *J. Phys.: Condens. Matter* **27** 015003
- [67] Pascuet M I and Fernández J R 2015 *J. Nucl. Mater.* **467** 229
- [68] Zhang L, Han J, Wang H, Car R and E W 2018 *Phys. Rev. Lett.* **120** 143001
- [69] Hastie T, Tibshirani R and Friedman J 2008 *The Elements of Statistical Learning Data Mining, Inference and Prediction* (Springer Series in Statistics) 2nd edn (Berlin: Springer)
- [70] The Python package 2019 (available at: www.tensorflow.org)
- [71] Unke O T, Chmiela S, Sauceda H E, Gastegger M, Poltavsky I, Schütt K T, Tkatchenko A and Müller K-R 2021 *Chem. Rev.* **121** 10142
- [72] Jakse N and Pasturel A 2013 *J. Phys.: Condens. Matter* **25** 28510
- [73] Jakse N and Bryk T 2019 *J. Chem. Phys.* **151** 034506
- [74] Demmel F, Hennet L and Jakse N 2021 *Sci. Rep.* **11** 11815
- [75] Kresse G and Furthmüller J 1996 *Comput. Mater. Sci.* **6** 15
- [76] Ceperley D M and Alder B J 1980 *Phys. Rev. Lett.* **45** 566
- [77] Perdew J P and Zunger A 1981 *Phys. Rev. B* **23** 5048
- [78] Kresse G and Joubert D 1999 *Phys. Rev. B* **59** 1758
- [79] Monkhorst H J and Pack J D 1976 *Phys. Rev. B* **13** 5188
- [80] Blöchl P E, Jepsen O and Andersen O K 1994 *Phys. Rev. B* **49** 16223
- [81] Assael M J et al 2006 *J. Phys. Chem. Ref. Data* **35** 285
- [82] Alfè D 2003 *Phys. Rev. B* **68** 064423
- [83] Demmel F, Szubrin D, Pilgrim W C and Morkel C 2011 *Phys. Rev. B* **84** 014307
- [84] Kargl F, Weis H, Unruh T and Meyer A 2012 *J. Phys.: Conf. Ser.* **340** 012077
- [85] Sjöstrom T, Crockett S and Rudin S 2016 *Phys. Rev. B* **94** 144101
- [86] Perdew J P, Chevary J A, Vosko S H, Jackson K A, Pederson M R, Singh D J and Fiolhais C 1992 *Phys. Rev. B* **46** 6671
- [87] Christensen A S and von Lilienfeld O A 2020 *Mach. Learn. Sci. Technol.* **1** 045018
- [88] Allen M P and Tildesley D J 1989 *Computer Simulation of Liquids* (Oxford: Oxford Science Publication)
- [89] Smit B and Frenkel D 2002 *Understanding Molecular Simulations* 2nd edn (San Diego, CA: Academic)
- [90] Honeycutt J D and Andersen H C 1987 *J. Phys. Chem.* **91** 4950
- [91] Faken D and Jonsson H 1994 *Comput. Mat. Sci.* **2** 279
- [92] Stukowski A 2010 Visualization and analysis of atomistic simulation data with OVITO - the open visualization tool *Modelling Simul. Mater. Sci. Eng.* **18** 015012
- [93] Steinhardt P, Nelson D R and Ronchetti M 1983 *Phys. Rev. B* **28** 784
- [94] Lechner W and Dellago C 2008 *J. Chem. Phys.* **129** 114707
- [95] Menon S, Leines G D and Rogal J 2019 *J. Open Source Softw.* **4** 1824
- [96] Mauro N A, Bendert J C, Vogt A J, Gewin J M and Kelton K F 2011 *J. Chem. Phys.* **135** 044502
- [97] Binder K and Kob W 2005 *Glassy Materials and Disordered Solids* (Singapore: World Scientific Publishing)
- [98] Herlach D M, Binder S, Galenko P, Gegner J, Holland-Moritz D, Klein S, Kolbe M and Volkmann T 2015 *Metall. Mater. Trans. A* **46** 4921
- [99] Orava J and Greer A L 2014 *J. Chem. Phys.* **140** 214504
- [100] Leitner M, Leitner T, Schmon A, Aziz K and Pottlacher G 2017 *Metall. Mater. Trans. A* **48** 3036

- [101] Simmons G and Wang H 2009 *Single Crystal Elastic Constants and Calculated Aggregate Properties (A Handbook)* 2nd edn (Cambridge: The MIT Press)
- [102] Sun D Y, Asta M and Hoyt J J 2004 *Phys. Rev. B* **69** 024108
- [103] Morris J R, Wang C Z, Ho K M and Chan C T 1994 *Phys. Rev. B* **49** 3109
- [104] Morris J R and Song X Y 2002 *J. Chem. Phys.* **116** 9352
- [105] Zykova-Timan T, Rozas R E, Horbach J and Binder K 2009 *J. Phys.: Condens. Matter* **21** 464102
- [106] Zykova-Timan T, Horbach J and Binder K 2010 *J. Chem. Phys.* **133** 014705
- [107] Kuhn P and Horbach J 2013 *Phys. Rev. B* **87** 014105
- [108] Benjamin R and Horbach J 2015 *J. Chem. Phys.* **143** 014702
- [109] Rozas R, Demirag A, Toledo P G and Horbach J 2016 *J. Chem. Phys.* **145** 064515
- [110] Rozas R E, MacDowell L G, Toledo P G and Horbach J 2021 *J. Chem. Phys.* **154** 184704
- [111] Boehler R and Ross M 1997 *Earth Planet. Sci. Lett.* **153** 223
- [112] Hänsström A and Lazor P 2000 *J. Alloys Compd.* **305** 209
- [113] Shaner J W, Brown J M and McQueen R G 1984 *High Pressure in Science and Technology* (Amsterdam: North Holland)
- [114] Bouchet J, Bottin F, Jomard G and Zérah G 2009 *Phys. Rev. B* **80** 1
- [115] Lechner W and Dellago C 2008 *J. Chem. Phys.* **129** 114707
- [116] Alexander S and McTague J P 1978 Should all crystals be bcc? Landau theory of solidification and crystal nucleation *Phys. Rev. Lett.* **41** 702–5
- [117] Ostwald W 1897 The formation and changes of solids (Translated from German) *Z. Phys. Chem.* **22** 289–330
- [118] ten Wolde P R, Ruiz-Montero M J and Frenkel D 1995 *Phys. Rev. Lett.* **75** 2714–17
- [119] ten Wolde P R, Ruiz-Montero M J and Frenkel D 1996 Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling *J. Chem. Phys.* **104** 9932–47
- [120] Mahata A, Zaeem M A and Baskes M I 2018 Understanding homogeneous nucleation in solidification of aluminum by molecular dynamics simulations *Modelling Simul. Mater. Sci. Eng.* **26** 025007
- [121] Becker S, Devijver E, Molinier R and Jakse N 2022 *Sci. Rep.* **12** 3195
- [122] Zhou L, Pan J-M, Lang L, Tian Z-A, Mo Y-F and Dong K-J 2021 *RSC Adv.* **11** 39829