# Searching for iron nanoparticles with a general-purpose Gaussian approximation potential

Richard Jana [*] and Miguel A. Caro

*Department of Chemistry and Materials Science, Aalto University, 02150 Espoo, Finland*

We present a general-purpose machine learning Gaussian approximation potential (GAP) for iron that is applicable to all bulk crystal structures found experimentally under diverse thermodynamic conditions, as well as surfaces and nanoparticles (NPs). By studying its phase diagram, we show that our GAP remains stable at extreme conditions, including those found in the Earth's core. The new GAP is particularly accurate for the description of NPs. We use it to identify new low-energy NPs, whose stability is verified by performing density functional theory calculations on the GAP structures. Many of these NPs are lower in energy than those previously available in the literature up to $N_{atoms} = 100$. We further extend the convex hull of available stable structures to $N_{atoms} = 200$. For these NPs, we study characteristic surface atomic motifs using data clustering and low-dimensional embedding techniques. With a few exceptions, e.g., at magic numbers $N_{atoms} = 59, 65, 76$, and 78, we find that iron tends to form irregularly shaped NPs without a dominant surface character or characteristic atomic motif, and no reminiscence of crystalline features. We hypothesize that the observed disorder stems from an intricate balance and competition between the stable bulk motif formation, with bcc structure, and the stable surface motif formation, with fcc structure. We expect these results to improve our understanding of the fundamental properties and structure of low-dimensional forms of iron and to facilitate future work in the field of iron-based catalysis.

## I. INTRODUCTION

Iron nanoparticles (NPs) are widely used for catalytic purposes, e.g., for the hydrogen evolution reaction (HER) [1,2], the oxygen reduction reaction (ORR) [2], or light olefin synthesis [3,4]. To enable an economy not relying on crude oil for energy or base chemicals, the development of cost-effective and scalable catalysts for these reactions is crucial. Nonprecious catalysts such as iron NPs are of special interest because of their high reactivity and low cost compared to Pt- and Pd-based catalysts.

To study the atomic-level processes of these reactions in detail, while considering a wide variety of NPs and active sites, an accurate but also computationally cheap model is necessary. However, among the existing models, density functional theory (DFT) is too expensive for such comprehensive studies and the different classical interatomic potentials lack in accuracy and general applicability. Early on, many embedded atom method (EAM) potentials were developed for general application to body-centered cubic (bcc) $\alpha$-Fe [5–8]. These were followed by a bond-order potential (BOP) [9], even able to describe magnetic interactions [10] and, to some extent, face-centered cubic (fcc) $\gamma$-Fe. Later, potentials for special purposes were created, for instance to study radiation defects [11] or the conditions in the Earth's core [12]. Recently, different flavors of machine learning (ML) potentials [13] have been trained for iron, including neural network potentials [14,15] and Gaussian approximation potentials (GAPs) [16,17].

Some of the potentials mentioned above claimed general applicability, but are not truly general as they are not applicable to all crystal phases, surfaces, NPs and disordered structures (including the liquid). Rather, they are typically designed to describe a wide range of properties of $\alpha$-Fe, with no guarantee of transferability outside of this range. In this work, we present a new GAP ML potential (ML) trained on a much wider range of structures for true general applicability and transferability across a wide range of problems in atomistic modeling of iron. The results are compared to the GAP potential by Dragoni *et al.* [16], as the current state of the art, and the EAM potential by Mendelev *et al.* [5], as one of the most used classical alternatives with a lower computational cost. These were chosen as representative of two different philosophies for deriving interatomic potentials: empirical potentials fitted to material properties and MLPs trained on energies and forces computed from DFT. While both types of potentials are ultimately designed to reproduce the material properties, which are in turn linked to the energies and forces of atomic configurations, the two approaches should be distinguished as they lead to different tradeoffs between accuracy and computational cost.

We showcase the ability of our GAP to accurately describe the potential energy surface (PES) of crystalline and nanostructured iron, including a reasonable description of different surfaces and phase transformations at extreme thermodynamic conditions. The highlight application of this paper is the search for stable iron NPs of different sizes, a task for which our GAP achieves accuracy remarkably close to that of

_____
*richard.jana@aalto.fi

DFT. We hope that this work will speed up the discovery of efficient iron-based nanocatalysts.

## II. DATABASE GENERATION

There are three critical steps in training a GAP: (1) training database generation, (2) selection of model architecture and hyperparameters, including the choice of atomic descriptors and data regularization, and (3) the computation of the fitting coefficients. These have been covered in detail in the literature, and we refer the reader to Refs. [18,19] for an in-depth discussion. Here, we will only give a brief account of the technical ingredients of our GAP and focus mostly on accuracy benchmarks and applications.

Ensuring the accuracy and transferability of a GAP, or any other MLP, for that matter, relies on the availability of a database of atomic structures covering the relevant regions of configuration space. For a general-purpose MLP, this means that comprehensive sampling of the PES needs to be done. Our iron database contains dimers and trimers, crystalline structures [bcc, fcc, hexagonal close-packed (hcp), simple cubic (sc) and diamond] over a wide range of cell parameters and with "rattled" atomic positions (i.e., atoms slightly displaced about their equilibrium positions), transitional structures between bcc-fcc and bcc-hcp, surface slabs cleaved from the relaxed bulk structures, NPs and liquid configurations. For each structure, the magnetic configuration with the lowest energy was chosen for inclusion in the database. In this way, our GAP is fitted to the DFT ground state with regard to the magnetic degrees of freedom, which are otherwise not explicitly taken into account in our PES description. Detailed types and numbers of structures in our training database are given in Table S1 of Ref. [20].

The energy, forces, and virials for the atomic structures in our training database were computed at the DFT level of theory using VASP [21–23]. We used the PBE functional [24] with standard PAW pseudopotentials [25,26] for Fe (with eight valence electrons, $4s^2 3d^6$). The kinetic energy cutoff for plane waves was set to 400 eV and the energy threshold for convergence was $10^{-7}$ eV. All the DFT calculations were carried out with spin polarization, which can describe collinear magnetism. While noncollinear magnetic effects can in principle be described in VASP, the gain in accuracy in the context of MLP simulation is only modest compared to the increased CPU cost and difficulty to systematically converge thousands of individual calculations in a high-throughput setting.

On this database, we trained our GAP with two-body, three-body, and many-body smooth overlap of atomic positions (SOAP) [27,28] atomic descriptors using a cutoff of 5, 3, and 5 Å, respectively. A "core" potential, a tabulated pairwise interaction at very short interatomic distances, was added to model the strongly repulsive regime down to 0.1 Å. The number of sparse configurations and the regularization parameter were both chosen per configuration type and are listed in Table S1 of Ref. [20]. The training was carried out with the QUIP/GAP codes [29,30]. The full command passed to the `gap_fit` binary is given in Ref. [20]. We again refer the reader to the literature for further details on GAP training [18,19].

## III. GAP VALIDATION

In this section, we validate our GAP against a wide range of simulation problems and compare it to existing potentials. We first motivate the need and usefulness of a general-purpose MLP for iron. We then benchmark the GAP for the description of bulk iron, phase transitions, elastic properties, and surface calculations.

### A. General-purpose versus bcc-specific iron potential

While ferromagnetic bcc is the ground-state structure of bulk iron at room temperature and pressure, iron transitions to other stable structures as the thermodynamic conditions change. In addition, surfaces cleaved from the bulk look different depending on the bulk crystal structure. Nanostructured iron, in particular NPs, will not necessarily have a bcc or, for that matter, ordered structure, even at room temperature and pressure. Finally, liquid iron is simply disordered and thus its structure differs significantly from bcc or any other crystal structure. Interatomic potentials trained from bcc data can be very useful to accurately describe the properties of $\alpha$-Fe, but their accuracy deteriorates rapidly as they extrapolate in regions of configuration space away from the training data. Here we show how our GAP overcomes these issues and provides a consistent prediction of the PES of iron for widely different problems, enabling an accurate description of NPs of varying sizes. We will also show that, in the absence of an explicit inclusion of the magnetic degrees of freedom, this transferability is achieved at the cost of sacrificing accuracy in the description of some of the properties, e.g., of the surface energetics of the different crystal phases.

Figure 1 shows the energy and force errors of (a) the GAP potential developed in this work, (b) the GAP potential by Dragoni *et al.* [16] and (c) the EAM potential by Mendelev *et al.* [5] against the corresponding DFT values in two *training* databases: the training database of our GAP on the left and that of the Dragoni GAP on the right. The energy values have been referenced to the bulk energy of bcc iron for each potential, to make the results comparable between the potentials. Each panel shows the RMSE for the respective data. Importantly, Fig. 1 is not intended as an accuracy test of our GAP, for which reporting the training-set RMSE is meaningless. We rather use it to showcase the difficulty encountered by bcc-specific iron potentials to reproduce the PES of other structures, on the one hand, and to quantify the ability of the GAP framework to learn the PES within a significantly more comprehensive region of configuration space than that corresponding to bcc, on the other. In this regard, our GAP is able to learn our general-purpose training database to an accuracy of 20 meV/atom which, while satisfactory for many purposes, is significantly higher than the 2 meV/atom with which the Dragoni GAP can learn its own bcc-specific training database. At the same time, our GAP only significantly deviates in the predictions of high-energy structures in the Dragoni database, with an overall RMSE of 60 meV/atom, mostly arising from outliers in the high-energy regions of the bcc PES, whereas the Dragoni GAP struggles to capture the energetics of many low-energy structures in our database (as well as the high-energy ones), with an overall RMSE of
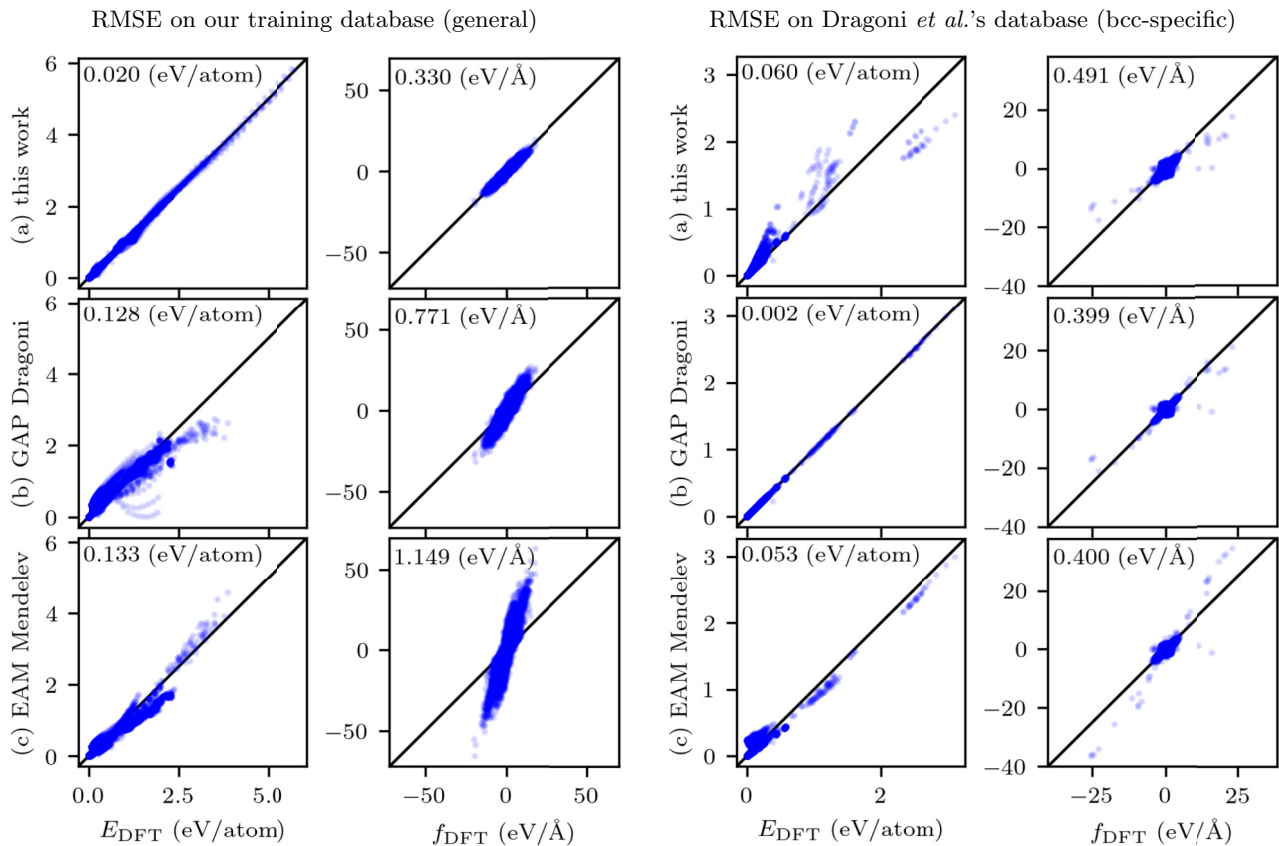
FIG. 1. Energy and force RMSE of (a) the GAP potential developed in this work, (b) the GAP potential by Dragoni *et al.* [16] and (c) the EAM potential by Mendelev *et al.* [5] compared to DFT. On the left are the RMSEs computed on our training database and on the right are the RMSEs on the training database of Dragoni *et al.* [16]. The energy values have been referenced to the bulk energy of bcc iron for each potential. Note that on the left, only the panel for our GAP includes the diamond and sc structures.

128 meV/atom. The EAM's performance is more predictable, with reasonably good RMSE's for bcc iron.

We note that the test in Fig. 1 (left) was done only for physically meaningful structures. The diamond and simple cubic structures used in the training of our GAP are not included in the plots for the reference potentials, as these structures are high in energy and not physically meaningful and would not make for a fair, nor instructive, comparison. They are included in the plot for our GAP, though, increasing the RMSE from 12 to 20 meV/atom there. Dimer and trimer structures are excluded from the plots as well, as they reach very high energies and would obscure the more important data ranges. All other configurations used in the training are shown here, including the different bulk crystal structures, surfaces, melt, vacancies and NPs.

Unsurprisingly, the energies in our training database are very well reproduced by the GAP potential developed in this work. The GAP potential by Dragoni *et al.* [16] reproduces the energies almost to the same RMSE as the EAM potential by Mendelev *et al.* [5]. For both potentials, nucleation clusters, NPs and structures derived from hcp pose the greatest problems, with RMSEs of 312, 171, and 151 meV/atom for the Dragoni GAP, and 460, 343, and 119 meV/atom for the Mendelev EAM. The energies of structures derived from bcc on the other hand are predicted best, with RMSEs of 8 and 11 meV/atom, respectively.

Again, unsurprisingly, the forces in our training database are reproduced very well with the GAP potential developed in this work. Both the GAP potential by Dragoni *et al.* [16] and the EAM potential by Mendelev *et al.* [5] predict forces that are systematically too large by factors of approximately 1.1 and 1.6, respectively.

The energy and force errors shown in Fig. 1 are significantly lower with our GAP than with the reference potentials, which had to be expected as the data contains many types of structures that the reference potentials were never intended for. It is still not a given that our GAP would reproduce the energies and forces so well in the different regions of configuration space, as we have observed that the addition of training data in one region usually leads to a very slight degradation in the other regions, a phenomenon that we ascribe to the absence of explicit magnetic degrees of freedom in our GAP. Also note the slopes of the force error data for the two reference potentials, especially the Mendelev EAM. This shows a typical behavior of empirical potentials to have too strong a driving force towards stable configurations, by design, in order to avoid unstable trajectories in molecular dynamics (MD) simulations.

### B. Description of bulk iron

Figure 2 shows the relative stability of the bulk crystal phases versus the atomic volume for all three potentials.
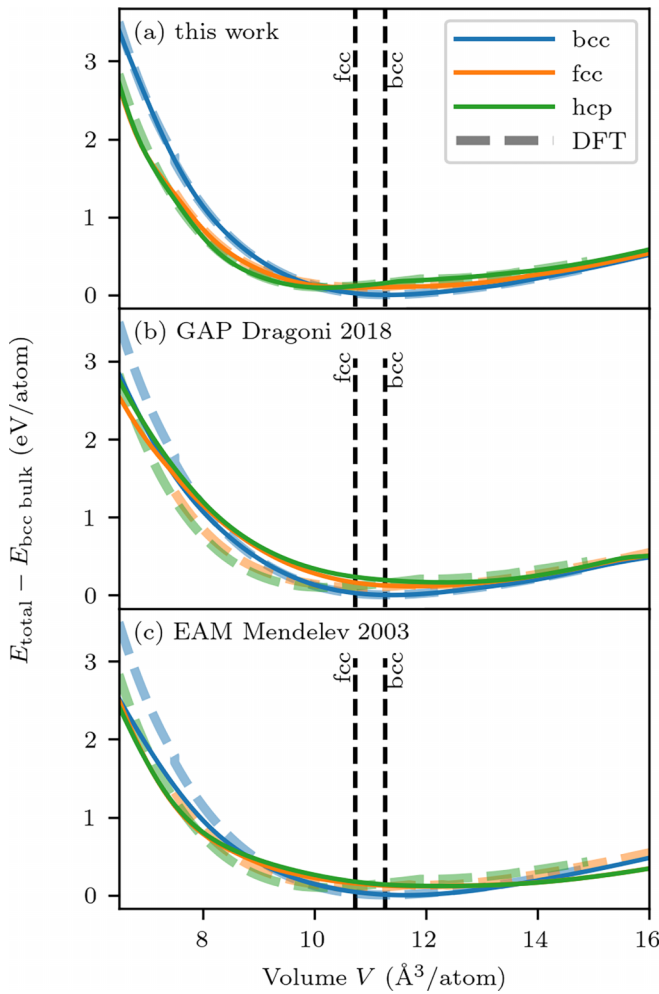
FIG. 2. Energy of different Fe crystal phases over a wide range of atomic volumes: (a) the GAP of this work, (b) the GAP potential by Dragoni *et al.* [16], and (c) the EAM potential by Mendelev *et al.* [5]. Vertical black lines mark the equilibrium atomic volumes of the bcc and fcc phases. DFT data for each crystal phase are underlayed as thick dashed lines in each panel.

Curves over a larger range of atomic volumes and additional diamond and sc structures are shown in Ref. [20], Fig. S1. The shaded dashed curves represent DFT reference values. The atomic volumes of the DFT minimum structures for bcc and fcc are marked in each panel.

The GAP developed in this work and the GAP by Dragoni *et al.* capture the bcc curve very well, including the minimum. The EAM by Mendelev *et al.* has the minimum at a slightly too large volume and diverges from the DFT reference data at lower and higher atomic volumes. The fcc energies are only reproduced well throughout the whole range considered by our GAP. The GAP by Dragoni *et al.* only gives the correct energies from approximately 11.5 to 16 $Å^3$/atom. Towards lower atomic volumes the slope is much too steep, erroneously predicting fcc iron to be less stable there than bcc iron. The EAM by Mendelev *et al.* does not capture fcc well at all, with an exception around 12 $Å^3$/atom where it gives approximately the correct energies (but wrong trends). However, there the hcp energy is much too low, with both phases predicted as

having similar energy. From 8 to 10 $Å^3$/atom, where hcp should be the stable crystal phase, either bcc appears as more stable or all three crystal phases are practically identical in energy. With the Dragoni GAP, hcp is never the most stable structure in the atomic volume range shown here. At even larger volumes (see Fig. S1 in Ref. [20]), hcp becomes more stable, but spuriously so, with a predicted energy almost as low as for the bulk bcc minimum. The GAP developed in this work predicts the correct energies also for the hcp structure over the whole volume range studied.

Therefore both the Dragoni GAP and the Mendelev EAM show strong deviations from the DFT stabilities and predict the bcc phase as the stable one over too wide a volume range, with the high-pressure hcp phase missing. This can be easily attributed to the fact that the reference potentials were developed for the bcc phase, neglecting the other crystal structures. Of the problems at very high volumes (see Fig. S1 of Ref. [20]), only one seems important: the unphysical behavior of hcp beyond 16 $Å^3$/atom could lead to configurations blowing up during MD. (This has been fixed in the fracture GAP [17], based on the Dragoni GAP, see Fig. S2.)

Figure 3 shows the energies for strained bulk cells at different cell parameters for all three crystal structures and the three different potentials, each as the difference to the corresponding DFT value. For bcc, the energy landscape is reproduced equally well by the GAP potential by Dragoni *et al.* [16] and the GAP developed in this work. The EAM potential by Mendelev *et al.* [5] shows larger disagreement with the DFT reference, especially at low atomic volumes. Both for fcc and hcp, the energy error is significantly lower for the GAP developed in this work compared to the other two potentials. Noticeably, the Mendelev EAM performs slightly better than the Dragoni GAP. Both potentials overestimate the energy at low and underestimate it at high atomic volumes.

The lowest-energy structure is marked in black in each panel of Fig. 3 for the respective potential [ferromagnetic (FM) bcc, antiferromagnetic (AFM) fcc, and nonmagnetic (NM) hcp]. The lowest DFT energy structure is marked in green in all panels for reference. Local minima with higher energy are marked in grey and purple for the interatomic potentials and DFT, respectively. The minima for the GAP developed in this work coincide with the DFT minima for bcc and hcp. For the Dragoni GAP the bcc minimum differs only marginally from the DFT reference, for the Mendelev EAM slightly more. The hcp cell parameters predicted by the Dragoni GAP and the Mendelev EAM deviate substantially from the DFT reference.

For the fcc cell, multiple minima exist for DFT [9,31]. The lowest in energy is the AFM magnetic configuration with a tetragonal cell ($c$ longer than $a$), but two FM minima with cubic cells exist as well, usually called ferromagnetic low spin (FMLS) and ferromagnetic high spin (FMHS). Of the three potentials, only the GAP developed in this work reproduces more than one minimum structure correctly: the AFM and the FMLS. The Dragoni GAP does have a cubic and a tetragonal minimum structure as well, but with $a = 3.416$ Å and $c = 4.042$ Å, the tetragonal minimum is outside of the plotting range of Fig. 3. The Mendelev EAM does not have a tetragonal cell minimum at all, but just a cubic cell minimum corresponding to the FMHS at approximately the correct
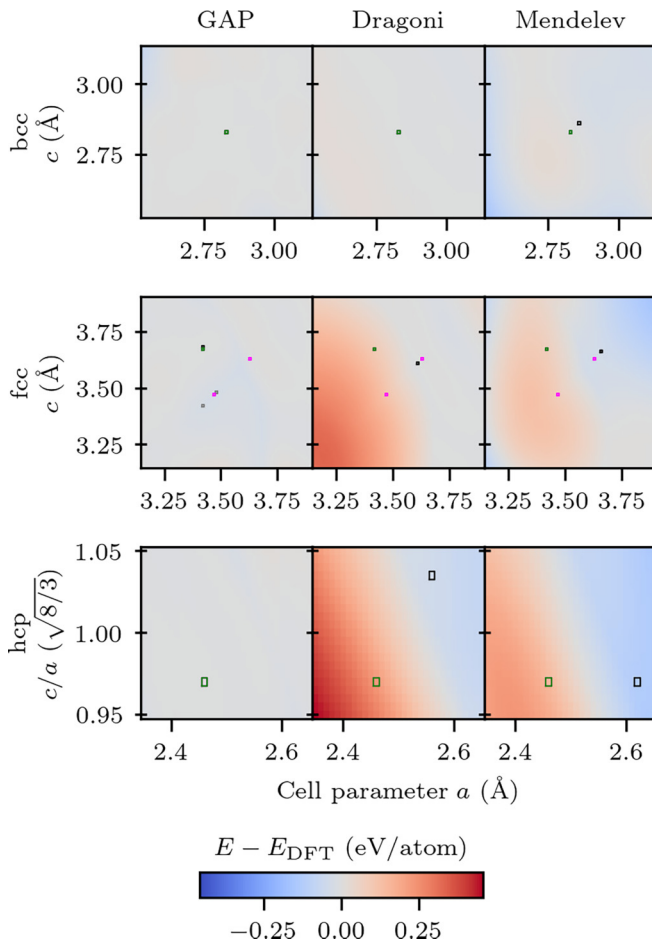
FIG. 3. Energy error (difference between interatomic potential and DFT energies) over a space of structural parameters. Rows contain bcc, fcc and hcp crystal structures and columns contain the three different potentials: the GAP potential developed in this work, the GAP potential by Dragoni *et al.* [16] and the EAM potential by Mendelev *et al.* [5]. The first structural parameter is always the cell parameter $a$ and the second is the cell parameter $c$ for bcc and fcc, but the aspect ratio $c/a$ for hcp. Marked in black are the lowest-energy structures in each panel for the corresponding potential, marked in green the lowest DFT energy structures. Interatomic potential and DFT data agree when green and black rectangles overlap on the graph. Additional local minima are marked in grey and purple for the interatomic potentials and DFT, respectively.

cell parameter, with cell parameter slightly too large at $a = 3.658$ Å, compared $a = 3.634$ Å from DFT. All cell parameters are given in Table S2 in Ref. [20].

The reproduction of the cell parameters of the crystalline phases shown in Fig. 3 works well for the bcc phase with all three potentials. Still, for the Mendelev EAM the deviation from the DFT cell parameters and the errors in the surrounding energy landscape are larger than with the Dragoni GAP and the GAP developed in this work. However, for the fcc and hcp phases, both reference potentials yield large errors in the energy landscape and cell parameters far from the DFT ones, while our GAP gives very low errors and the correct cell parameters. While multiple local minima exist for the cell parameters of the fcc cell, one for the tetragonal AFM configuration and two for the cubic FMLS and FMHS states,
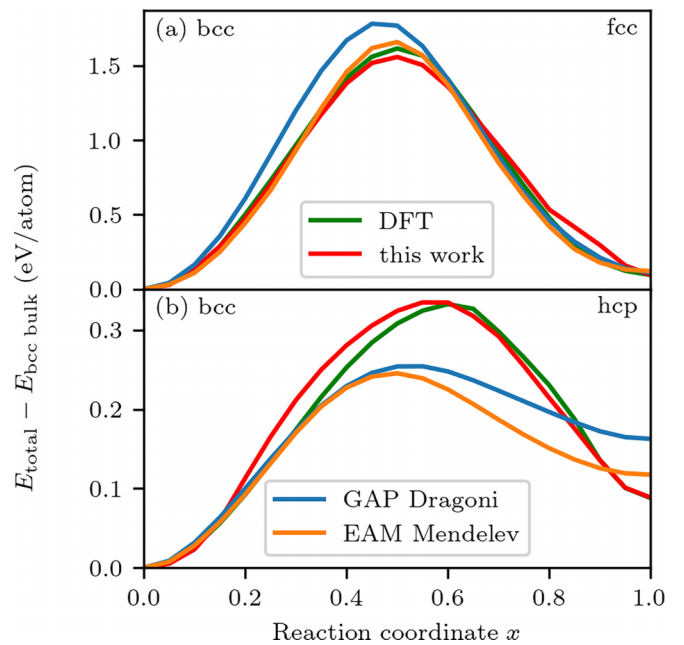


FIG. 4. Transition between (a) bcc and fcc and (b) bcc and hcp crystal structures calculated with DFT, the GAP potential developed in this work, the GAP potential by Dragoni *et al.* [16] and the EAM potential by Mendelev *et al.* [5]. Energies are referenced to the bcc bulk energy for each potential.

no potential has a minimum for all three of these. Thus, for our GAP, the only shortcoming here is that it does not have a minimum for the FMLS structure. The reason for this is the way that GAP fits the underlying data smoothly, potentially removing shallow minima in some instances. Note that, even if the FMHS minimum is missing, the error is still low.

We remark here that the inability of the Dragoni GAP and Mendelev EAM potentials to accurately describe crystal phases other than bcc is to be expected, and not an artifact, since they were designed to correctly describe bcc iron only. An accurate description of the low-pressure bcc structures can still be obtained with these potentials, especially with the Dragoni GAP, which should be able to indeed outperform our GAP for simulation of single-phase bcc iron.

### C. Phase transitions

So far we have discussed the accuracy of our GAP to describe (meta) stable structures. However, a general-purpose potential to be used in dynamic structure generation, e.g., involving MD simulation, also needs to accurately describe the PES along important transition paths between crystal phases. Since the initial and final states, as well as the minimum energy path for a transition, depend on the specific force field used, we choose the following approach to be able to compare the potentials to DFT and among them. For the transformations from bcc to fcc and from bcc to hcp, transition structures were created by *linear interpolation* between the cell parameters and atomic positions of the endpoint structures at 19 points along the path as shown in Fig. 4. The minimum DFT energy structures were used for all potentials, i.e., to the two reference potentials these endpoints are not the minimum energy structures, but for our GAP they are, as there the cell
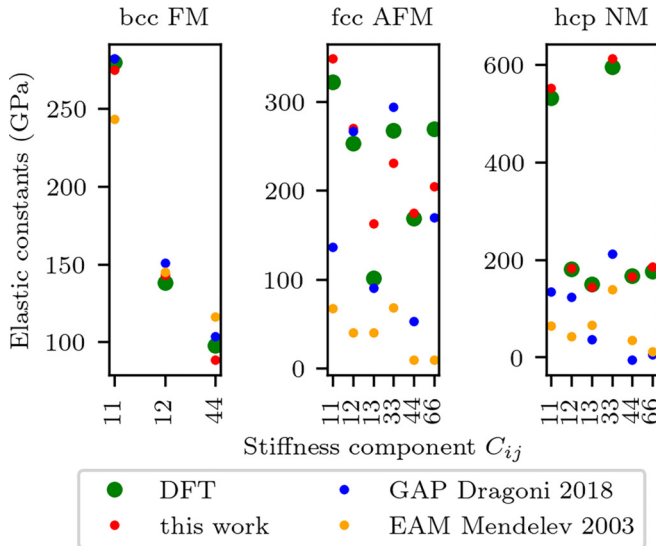
FIG. 5. Elastic constants for all crystal structures and potentials, compared with DFT results. Note that the Mendelev EAM predicts a cubic symmetry for the fcc structure, that is predicted tetragonal by DFT.

parameters are identical to the DFT ones. Thus, at the fcc and hcp endpoints, the Dragoni and Mendelev energies differ from the DFT reference values. All curves are referenced to the bcc bulk energy.

Along the bcc to fcc path shown here, our GAP and the Mendelev EAM reproduce the energy barrier reasonably well, our GAP a little too low and the Mendelev EAM a little too high. The Dragoni GAP energies along the path are considerably higher than with the other potentials. Along the bcc to hcp path, the Dragoni GAP and the Mendelev EAM trace the DFT curve up to $x \sim 0.3$ (where $x$ is the reaction coordinate), but then quickly diverge and reach an endpoint far from the DFT one (more so for the Dragoni GAP than the Mendelev EAM). Our GAP's curve has a slightly different shape, with a steeper incline at low $x$ and a higher maximum, but still fits the DFT curve much better than the other two potentials.

We note again that these curves were not obtained through the minimum energy path, e.g., by performing a nudged elastic band (NEB) calculation [32], but rather by linear interpolation between the endpoint structures. Hence, the maxima in the transition curves cannot be interpreted as "barriers" (i.e., the energy calculated at the saddle point along the minimum energy path), as lower-barrier paths might exist.

### D. Elastic properties

The elastic constants of all three DFT minimum-energy structures have been computed with all potentials by straining the structures in all relevant directions (depending on the symmetry) in 5 steps with strain increments in the range of $10^{-5}$ to $10^{-2}$ to check for consistency, as implemented in the Atomic Simulation Environment (ASE) [33]. The results are shown in Fig. 5.

The elastic constants of the bcc structure are generally reproduced well by both the GAP by Dragoni *et al.* and our GAP. With the Mendelev EAM, $C_{11}$ is noticeably too low by approximately 10% and for $C_{44}$ the difference to the DFT

reference value is about twice as large as for the two GAP potentials. The error for $C_{12}$ is only half as large as with the Dragoni GAP potential, though. For the hcp elastic constants, our GAP also yields low errors with respect to the DFT data. Both the reference potentials significantly underestimate the elastic constants. For the fcc elastic constants, the results are much more mixed: while most of the Mendelev EAM values are significantly too low, our GAP and the Dragoni GAP give values that are a mix of too low or high ones with some that are spot on.

For the elastic constants shown in Fig. 5, again only the bcc phase is represented well by all three potentials, with the Mendelev EAM showing the largest errors compared to the DFT reference values. However, the elastic constants of the hcp phase are only reproduced well by our GAP and the fcc phase is not reproduced well by any of the potentials. While the reference potentials were not developed for these crystal structures, our GAP does have the necessary structures in the training database and still fails in the prediction for the fcc phase. Although we tried extensively to train a GAP that could correctly reproduce the elastic constants of all three crystal phases by fine tuning the regularization parameters and sparse set configurations of the strained+rattled structures, we did not manage to obtain a fit that predicted all of them accurately at the same time. We attribute this to the fact that each crystal structure belongs to a different magnetic branch with possibly significantly different energetics, including the energy derivatives (i.e., forces and the stress tensor, used to compute the elastic constants). Different magnetic configurations exist for the fcc phase, depending on the stained state. Without the explicit treatment of the magnetic moments, the underlying energy landscape has discontinuities where the lowest energy magnetic state changes. Our GAP can resolve these branches, but only implicitly, whenever the structures are sufficiently different in terms of atomic arrangements. While it should be possible to train a dedicated potential to reproduce the elastic constants of any one crystal phase and magnetic configuration, predicting all of them accurately with a general potential seems impossible within our current methodological framework, especially when many other types of configurations are also considered. We speculate that only an iron MLP which explicitly accounts for the magnetic structure of iron will be able to accurately capture all of these features simultaneously. Augmenting the GAP framework to incorporate magnetism is far beyond the scope of this work, but we expect advances in this area within the next few years.

### E. Surfaces

Another stringent test for an interatomic potential is the prediction of surface energies cleaved and reconstructed along various crystallographic planes, since these structures look significantly different from the bulk. To calculate the energies of surfaces with various Miller indices cleaved from the bcc, fcc and hcp bulk structures, slabs with a number of layers between 4 and 16 were set up using ASE [33]. These were then relaxed using DFT calculations with a fixed box size. We ensured that the amount of added vacuum perpendicular to the slabs was sufficient to allow for relaxation in this direction and rule out any interaction between the periodic copies of

the slabs. We observed these slabs to have nontrivial magnetic structure, e.g., showing a strong dependence on the number of atomic monolayers. Thus converging these DFT relaxations was not possible in every case and, even when the calculations converged, local energy minima were found that were not necessarily also the global minima.

The result of the DFT relaxations is primarily dependent on the setup of the initial magnetic moments and the final magnetic configuration resulting from those. Large differences up to ∼100 meV/atom for the same slab with different magnetic configurations were found. The fcc surface slabs proved more problematic than those of the other crystal structures in this regard. Typically, the fcc magnetic configurations consisted of layers with opposite local magnetic moments (not necessarily all with the same magnitude), one or more atomic layers thick.

To extract meaningful surface energies from the energies of the slabs, convergence of the surface energy with respect to the slab thickness would be expected. This was found for many surface indices, but in some cases no convergence could be observed, indicating that only local minima were found for at least some of the slabs. To improve the performance of our GAP for surfaces, surface slabs relaxed with DFT were further relaxed with earlier versions of our GAP and single-point DFT calculations of the resulting structures fed back into the training set in an iterative manner (known as "iterative training" [34]).

The surface structures in our database were split into two categories: thin slabs that are too thin to have a bulklike region in the center and thicker slabs that do contain such a bulklike center. For the thin slabs, the regularization parameter during training was chosen 25 times higher than the default (the higher the regularization parameter the less stringently the GAP is required to follow the data). This was done to keep the structures in the database, but focus on the more realistic surfaces from the thicker slabs.

The surface energies for all three crystal structures and various surface indices are shown in Fig. 6, for our GAP and the GAP by Dragoni *et al.*, compared against the DFT values. Data points for a given surface index are connected by lines between the different potentials and the indices are encoded in color for easier tracking. An equivalent plot comparing our GAP to the EAM by Mendelev *et al.* is shown in Fig. S3 of Ref. [20].

For the bcc surface energies, both Dragoni's and our GAP give results that are comparable and match the DFT references reasonably well. While the overall value range for the Dragoni GAP better matches the one found using DFT (especially for high-energy surfaces, which our GAP overestimates), our GAP performs better in terms of reproducing the correct order of the low-energy surface indices, i.e., there are fewer crossings of the connecting lines. For the fcc surfaces, both GAPs predict the DFT reference values poorly. While the energies come out too high with our GAP, the one by Dragoni shows energies that are too low. Recall the issues already mentioned with convergence of fcc slab calculations with DFT as a function of the number of atomic layers; this is likely a problem that necessitates the explicit inclusion of magnetic structure and cannot be fully solved within the standard GAP framework.
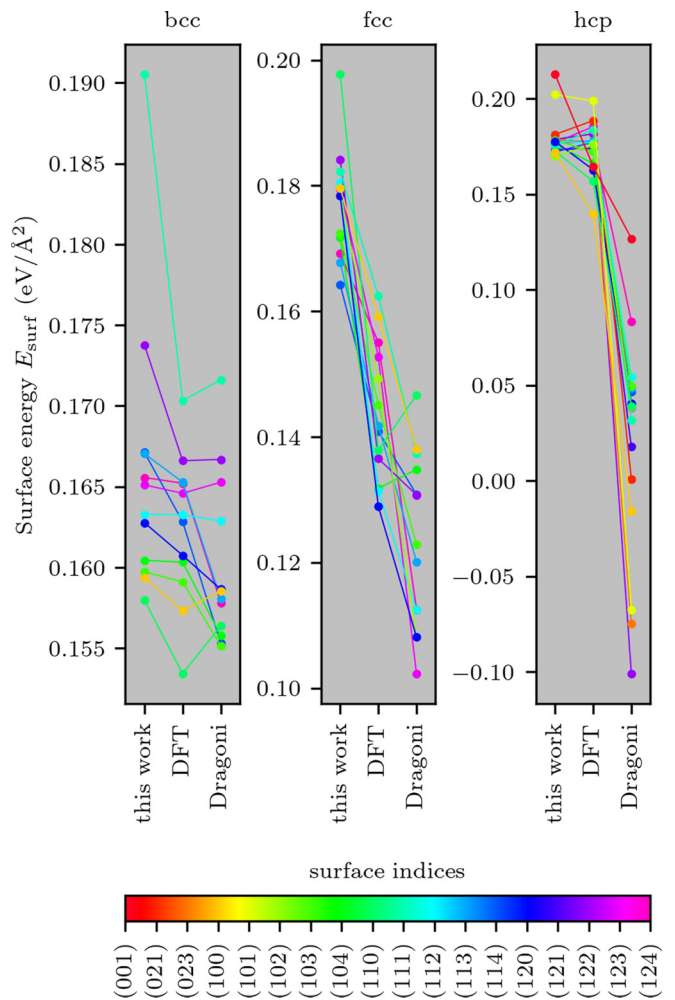


FIG. 6. Surface energies of a wide range of surface indices for the GAP by Dragoni *et al.* and our GAP, compared with the DFT values.

The hcp surface energies are predicted much too low with the Dragoni GAP, for some indices even negative, a clear sign of extrapolation outside of the training set. Our GAP, on the other hand, predicts all surface energies in a range similar to the one predicted by DFT, although with sizable errors. The EAM by Mendelev *et al.* gives bad predictions for the surface energies of all three crystal structures (Fig. S3 of Ref. [20]).

The calculation of the surface energies shown in Fig. 6 was challenging for some of the surface indices, as finding the magnetic ground state for these systems is much harder than for small bulk unit cells. In a few instances, it was not even possible to converge the DFT calculation at all. In other cases the magnetic layering changed erratically between slabs with a different number of layers, so that no convergence of the surface energy with increased slab thickness could be found. There, even trying to use the magnetic configuration of one slab to inform the initial configuration for a similar slab failed. This and the interpolation between different magnetic states in the fitting of our GAP led to mixed results. Consequently, the surface energies are more accurate for bcc where only the FM state exists and fewer convergence problems occurred in the creation of the training database. Still, our

GAP performs noticeably better than the other two potentials, giving a reasonable range of values for all surface energies and the correct order of the lowest-energy surfaces for bcc. The Dragoni GAP performs only slightly worse than our GAP for bcc (where it was trained), but gives even negative surface energies for hcp, which could lead to instabilities in dynamics simulations. Lastly, the Mendelev EAM yields high errors for the surface energies of all crystal structures, as it was fitted only to properties of the crystalline bulk and liquid.

## IV. APPLICATION TO SELECTED PROBLEMS

In this section, we benchmark our Fe GAP with representative use cases. We go beyond simple numerical scores, like RMSE and mean absolute error (MAE), and focus on how the potential performs when trying to reproduce experimental trends in (1) thermal expansion, (2) the solid-liquid phase transition, and (3) the temperature-pressure phase diagram.

### A. Thermal expansion

The thermal expansion was studied using MD calculations with ASE [33,35]. Systems with 1024 atoms were set up in the minimum DFT bcc structure and kept initially at 200 K (the lowest temperature in the series) for 1 ps to equilibrate and for another 2 ps for averaging with a time step of 1 fs. This procedure was repeated, step by step, increasing the temperature in intervals of 200 K up to 1600 K. Detailed simulation parameters are specified in Ref. [20] (see also Refs. [36,37] therein).

The data for the Dragoni GAP lines up almost perfectly with their DFT data [38], which is itself at lower volume than the experimental data by, e.g., Basinski *et al.* [39] and Ridley *et al.* [40]. The Mendelev EAM comes closest to the experimental data, but only really agrees around 200 K and between 1190 and 1660 K (where Basinski *et al.* found fcc Fe, as opposed to the bcc Fe predicted by EAM). Our GAP predicts the lowest atomic volumes at all temperatures, lower than the DFT data reported by Dragoni *et al.*, consistent with our DFT data. We note the two discontinuities in experimental data by Basinski *et al.* that take place first at the $\alpha$ to $\gamma$ and then at the $\gamma$ to $\delta$ phase boundaries. The only one of these happening where simulation data is available ($\alpha$ to $\gamma$) is not captured by any of the potentials.

The coefficients of thermal expansion were fitted around 400 to 600 K, where data points were available. The fitting ranges are shown in Fig. 7 as solid lines and the coefficients are noted next to the curves. From the three potentials studied, the Dragoni GAP comes closest to the experimental values and our GAP differs the most. All three potentials underestimate the experimental values.

The thermal expansion curve for our GAP starts at lower atomic volume than the reference potentials, in accordance with the cell parameters shown in Fig. 3. Also note the much larger sampled range of temperatures and volumes around the 1600 K data point for our GAP. We attribute this to the fact that the potential is exploring the energy landscape above the melting point (which is underestimated by our GAP, compared to experiment), but is missing a nucleation center for the liquid phase. We deal with the solid-liquid phase transition in Sec. IV B.
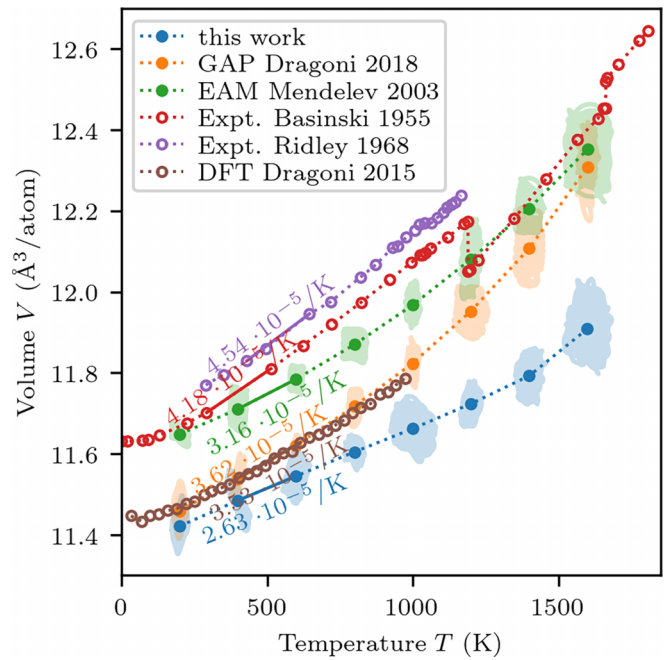


FIG. 7. Atomic volume over a wide range of temperatures to show the thermal expansion. Experimental results [39,40] and DFT calculations [38] shown as empty circles. For the calculations done in the scope of this work, volume and temperature along the trajectories are shown below the average values (full circles). For all curves, the coefficients of thermal expansion were fitted over the temperature ranges marked with solid lines. The coefficients are noted next to the curves.

### B. Solid-liquid phase transition

The melting temperature of bcc iron was calculated using the two-phase method [41,42] with all three potentials. Systems with $\approx$11000 atoms were set up as bcc crystals and heated to 4000 K for 15 ps to melt half of the system, while the positions of the other half were kept fixed. These half crystalline and half molten systems were then run at different target temperatures for 100 ps of MD with a time step of 1 fs to find the temperature where both phases coexisted. MD simulations were carried out using LAMMPS [43] for the Mendelev EAM [5] and Dragoni GAP [16] reference potentials and TurboGAP [28] for the GAP developed in this work. In the LAMMPS calculations, temperature and pressure were controlled with the Nosé-Hoover [44,45] thermostat and barostat with damping constants of 1 ps and 2 ps, respectively. In TurboGAP, a Berendsen [46] thermostat and barostat with the same damping constants and a `gamma_p`[1] of 55 were used.

Crystalline and molten states in the systems were identified with the Steinhardt parameter $Q_8$ [47]. Figure 8 shows the

---

[1]In TurboGAP, the bulk modulus for the barostat is expressed in units of the inverse compressibility of liquid water. For example, `gamma_p` = 55 means that the material is assumed to be 55 times as incompressible as liquid water for the purpose of barostating. This allows the user to provide an intuitive value for this parameter whenever the compressibility factor of the system is not known a priori (as is usually the case).
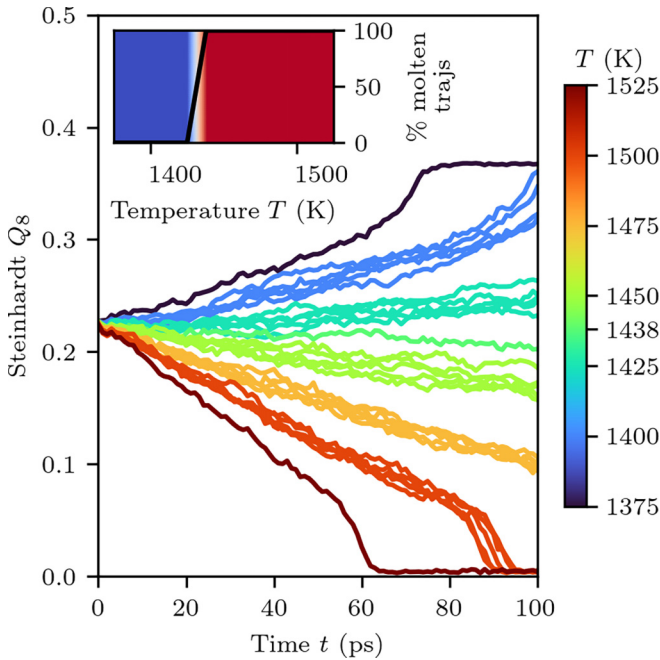
FIG. 8. Steinhardt parameter $Q_8$ [47] over time for MD simulations at various temperatures in the range of 1400–1550 K, around the melting temperature $T_m$. The inset shows the transition from the crystallizing to the melting state as a function of temperature.
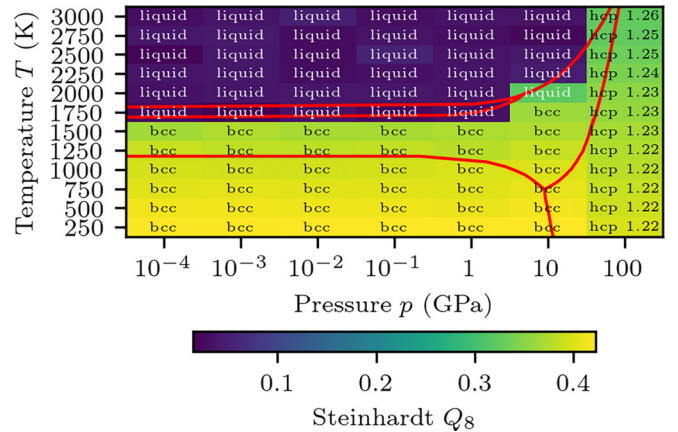


FIG. 9. Phase diagram for our GAP up to high temperatures and pressures. Crystalline structures were identified using XRD spectra and by comparing SOAP descriptors. For the hcp phase, the $c/a$-ratio is given. Encoded in the color is the Steinhardt parameter $Q_8$ [47], to further highlight the solid-liquid transition. Red lines indicate the phase boundaries in the experimental phase diagram [8,54]: $\alpha$, $\gamma$, $\delta$ (in increasing $T$, at low $p$) and $\epsilon$ (at high $p$).

results for our GAP, each line indicating a separate MD run. (For the results using the Mendelev EAM potential, see Fig. S5 in Ref. [20].) Decreasing values of $Q_8$ indicate a melting system, while increasing values indicate a crystallizing one. This gives approximate melting temperatures of 1760 and 1438 K for the Mendelev EAM and our GAP, respectively.

For our GAP, this is well below the experimental value of 1811 K. (The experimental value is for the $\delta$ phase, which is also bcc as in our simulations.) We attribute this deviation to the properties of the DFT functional used for the training, which has in previous works been shown to predict too low melting points. Partáy computed the phase diagram for the Mendelev EAM using the nested sampling method [8], believed to be the most comprehensive and accurate method for this purpose. She found a melting temperature higher than the experimental value, at approximately 1810–1940 K. The value of approximately 1760 K we find for the Mendelev EAM is right in the range of 1750–1775 K they give in their original paper, somewhat in disagreement with the nested sampling result. This disagreement can be attributed to a finite-size effect in the nested sampling calculations, overestimating the temperature of the melting transition compared to coexistence simulations.

For the Dragoni GAP, we found that the trajectories expanded to very high atomic volumes (about three times the volumes found with the other potentials for two-phase systems at the same temperature) upon releasing the crystalline atoms, immediately melting the crystalline half of the box. Thus we were not able to stabilize the two-phase state at any temperature. We attribute this to a spurious local minimum in the PES for a low-pressure melt, lower than the pressurized crystal.

We note that our GAP was not trained for thermal properties specifically, while the Dragoni GAP was trained on data for thermomechanical properties and the Mendelev EAM was fitted to the pair correlation function at 1820 K. It is therefore not surprising that the Mendelev EAM predicts the melting temperature better than the other two potentials.

### C. Phase diagram

To compute the phase diagram we carried out free-energy calculations within the 2PT framework [48] as implemented in the DoSPT code [49–51]. 2PT computes the free energy of an ensemble of atoms from the integral of the density of states (calculated from MD), which is partitioned between solidlike and gaslike degrees of freedom. This method is particularly suited to estimate the thermodynamic properties of liquids. In this work, we use it both for the liquid and the solid to be able to directly compare the free energy of the two and draw the melting curve: at any given set of thermodynamic conditions, the phase with the lowest free energy is the stable phase.

We calculated the iron phase diagram with our GAP up to high pressures of 100 GPa ($10^6$ bar) and temperatures of 3000 K, shown in Fig. 9. At each pressure, three MD trajectories were initialized from 250 K as bcc, fcc and hcp and one from 3000 K as liquid. The temperature was then increased/decreased in steps of 250 K using the Bussi thermostat [52] while controlling the pressure with the Berendsen barostat [46]. The calculations were done in TurboGAP [28] with equilibration constants of 100 and 1000 fs, respectively, and a gamma_p of 100. At each point, the trajectories were equilibrated for 80 ps and subsequently sampled for 80 ps. Liquid structures were detected using the Steinhardt $Q_8$ parameter [47], with low values indicating the liquid. We detected the crystalline structures by both comparing SOAP descriptors and x-ray diffraction (XRD) spectra calculated using the Debyer software package [53]. Example spectra at $p = 10^{-4}$ GPa (1 bar) and for the reference structures are

shown in Fig. S6 of Ref. [20]. For the hcp structures at 10 GPa, the $c/a$ ratio is given in Fig. 9, determined by comparing to the SOAP descriptors of hcp with $c/a$ ratios in the range of 1.10 to 1.40. The plot in Fig. 9 shows the structure of the trajectory with the lowest free energy at each point.

The phase diagram for our GAP shown in Fig. 9 reproduces very well the trends in the melting curve found experimentally [54], including the raised melting temperature at 10 GPa and the missing liquid phase at 100 GPa up to 3000 K. There is some disagreement between the melting temperature estimated using this method, which is located in the range 1500–1750 K, and that estimated using the two-phase method in Sec. IV B, which is situated a bit below 1450 K. The fcc phase is missing completely from our phase diagram, although it is found experimentally within a narrow band of temperatures, above ∼1180 K and below ∼1670 K (depending on the pressure).

At the highest pressures studied in this work, all the trajectories led to spontaneous nucleation of hcp up to a temperature of at least 3000 K (we did not check higher temperatures than this), also in agreement with the experimental data. In our phase diagram we also show the $c/a$ ratio, which is much lower than at ambient pressure. With increased temperature the spacing between the close-packed planes grows, which seems sensible. This leads us to believe that our GAP could be suitable to study iron at the conditions of the Earth's core (exceeding 136 GPa and approximately 4000 K [55]). Under such conditions, the atomic volumes are in the range of the lowest volumes shown in Fig. 2, corresponding to pressures in excess of 375 GPa at a temperature of 6000 K [12].

To further elucidate the suitability of our GAP to model the behavior of iron under extreme conditions, we computed the (RDFs) of liquid iron at various $(p, T)$ and compared them to reference data from *ab initio* MD (AIMD) calculations [56], shown in Fig. 10. Our RDFs agree very well with the reference data, capturing the general shape of the curves as well as the shift of the first peak towards lower distances.

## V. NANOPARTICLES

NPs were created in four distinct ways: (1) by condensation from a random starting atomic distribution, (2) using the genetic algorithm (GA) implementation by Weal *et al.* [57], (3) with the Wulff method [58], and (4) by annealing NPs found with the other methods at elevated temperature. For the first method, the atoms were randomly placed in a periodic box with double the atomic volume of the bcc bulk. The atomic positions were then relaxed using the current version of the GAP developed in this work using the algorithm detailed in the Atomic Simulation Recipes [59]. The suitable selection of the box volume ensured that the atoms coalesced into a single NP. The GA was run with a population size of 100 particles for 2500 generations with 20 offsprings per generation for each NP size. The starting populations were created using the condensation method. For detailed GA settings, see Sec. VIII of Ref. [20]. From all NPs evaluated during a GA run, the 1st, 50th, 100th and 500th lowest in energy (as per the GAP developed in this work) were then calculated with DFT to be used in the training of the next iteration of the GAP potential. In total, four iterations of GA were done to improve
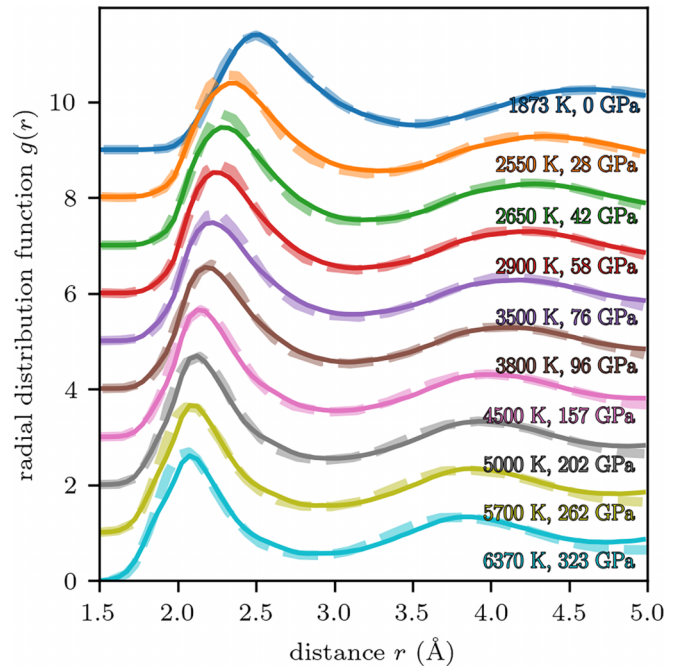


FIG. 10. Liquid iron RDFs at high $(p, T)$ along the AIMD melting line [56] (dashed lines). Data computed using our GAP overlaid (solid lines). Curves offset for clarity.

the performance of the potential for (increasingly) low-energy NPs. Crystalline NPs were generated using the Wulff method [58] as implemented in ASE [33], using the surface energies we calculated using DFT. To augment the search space, NPs found with the aforementioned methods were also annealed at 1200 K for 20 ps, quenched down to 300 K over another 20 ps and finally relaxed using gradient-descent minimization.

The region of configuration space corresponding to NPs displays a rather complex PES, due to the coexistence of diverse atomic motifs not encountered in the bulk: small surfaces, edges and vertices. To make this problem tractable, we used the iterative training approach [34] combined with the GA to incrementally improve the accuracy in this region of configuration space. Figure 11 shows the energies of the NPs created by and calculated with our new GAP and the two reference potentials, compared to the energies calculated using DFT and referenced to the energy of the bcc bulk. The energies calculated with the two reference potentials differ significantly from the DFT energy, while our GAP predicts the energies with good accuracy, regardless of which potential was used to generate the NPs. The RMSEs are given in the legend. Note how the Mendelev EAM predicts too low energies for all the low-energy NPs and the Dragoni GAP predicts a number of NPs to be lower in energy than the bcc bulk material, again a sign of extrapolation outside of the training set, as seen for hcp surfaces. Note that none of the NPs in Fig. 11 belong to the GAP training set, and thus this test gives a clear indication of the ability of our GAP to accurately generate and predict iron NPs, a particularly challenging modeling task.

In Fig. 11, we show the accuracy of the energy predictions of our GAP and the two reference potentials for a large number of NPs over a wide range of stabilities. This is meant to
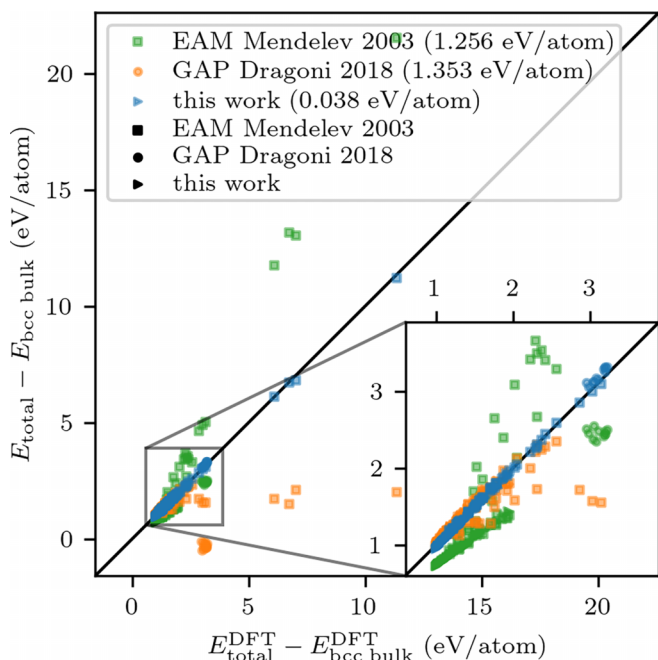
FIG. 11. Energy for NPs generated and computed with the GAP potential developed in this work, the GAP potential by Dragoni *et al.* [16] and the EAM potential by Mendelev *et al.* [5] compared to DFT. Shapes indicate the potential used to generate the NPs and colors indicate the potential used to calculate the energies, respectively. The NPs were generated using the random condensation method (for details, see text). None of the NPs were used in the training of our GAP. The inset shows the data at lower energies more clearly.
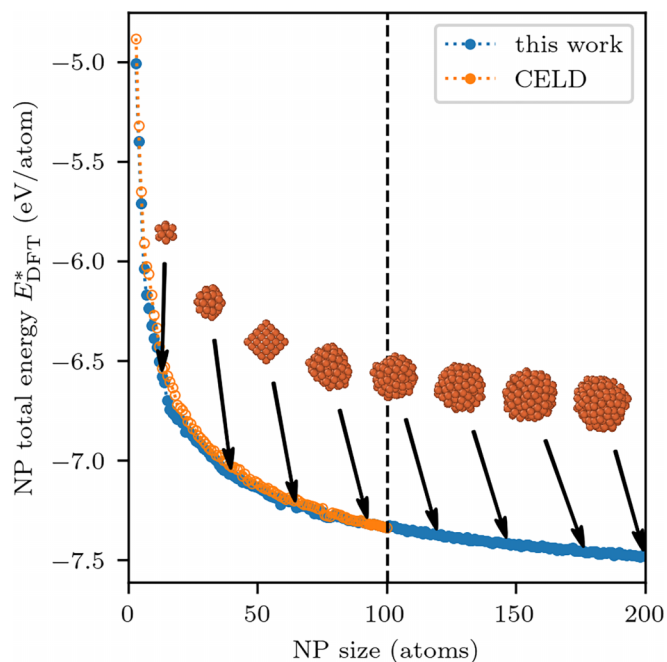


FIG. 12. DFT total energy convex hull of the NPs from the Cambridge Energy Landscape Database [67] and the NPs discovered in this work. Full circles indicate NPs in the convex hull that were improved by this work (90 out of 98 particles). Snapshots show a selection of NPs along the convex hull.

ensure that our GAP can model small NPs as well as ones that are far from the lowest energy for their size, as might be observed at the elevated temperatures of a catalytic process. The inability of the bcc reference potentials to correctly describe NPs can be seen for example in the isolated group of points for the Dragoni GAP at the bottom of the plot: all the NPs in this group were created with the Dragoni GAP and the performance on these is clearly different than on the NPs created with the Mendelev EAM. In contrast to the two reference potentials, our GAP performs very well on all the NPs shown here, regardless of which potential they were created with or their relative stability. This can be measured in the RMSE value which is about 30 times lower than for the reference potentials. More importantly though, there is no region in the plot where our GAP has substantial errors. This is particularly important for NP modeling, when we use the GAP to generate NPs with a few hundred or thousands of atoms, which cannot be directly validated with DFT due to CPU cost.

In the search for stable NPs, the common practice is to generate low-energy NPs within a range of sizes which, for small NPs, is measured in terms of the number of atoms. The energies are then used to construct a convex hull of NP stability. Because of the high computational cost of DFT calculations, the known convex hulls reach only up to a size of 30 atoms at most [60–63]. More comprehensive convex hulls, up to a size of 100 atoms, have been computed using the Finnis-Sinclair [6,64] EAM [65–67]. With our GAP, we have reconstructed the convex hull of the lowest-energy NPs

for each size up to a size of 200 atoms, using the search methods detailed above. The GA was only used up to a size of 100 atoms, due to its comparatively high computational cost. To validate the GAP results, the energies of the NPs from the Cambridge Energy Landscape Database (CELD) [67] as well as the NPs in the convex hull found in this work were recalculated using DFT up to a size of 200 atoms. The total energies for both are shown in Fig. 12. In the curve for the convex hull from this work, empty circles indicate NPs that are higher in energy than the CELD NP of the same size and full circles such particles that are lower in energy.

In the size range from 3 up to 100 atoms, where CELD data is available, 90 out of 98 NPs in our convex hull are lower in energy than the CELD ones. Most of these were found using the annealing method, but $Fe_{59}$ and $Fe_{65}$ were constructed with the Wulff method [58] (subsequently relaxed using our GAP). From these, the two particles created using the Wulff method stand out from the curve with particularly low energy (e.g., see snapshot at 65 atoms in Fig. 12). That we found more stable NPs despite relying on a less sophisticated search method than Liu *et al.* [66], in terms of the number of NPs that could be sampled, can be attributed to the much higher accuracy of our potential compared to the Finnis-Sinclair EAM. This is especially true for very small NPs, where the error for the Finnis-Sinclair EAM is the highest. (For example, see the energies for the CELDs convex hull shown in Fig. S7 [20].)

Among the DFT data reported in the literature, the structures of the NPs are unavailable. We were thus unable to compute the total energies with our GAP for comparison. The published binding energies, on the other hand, strongly depend on the exchange-correlation functional used (e.g., BLYP
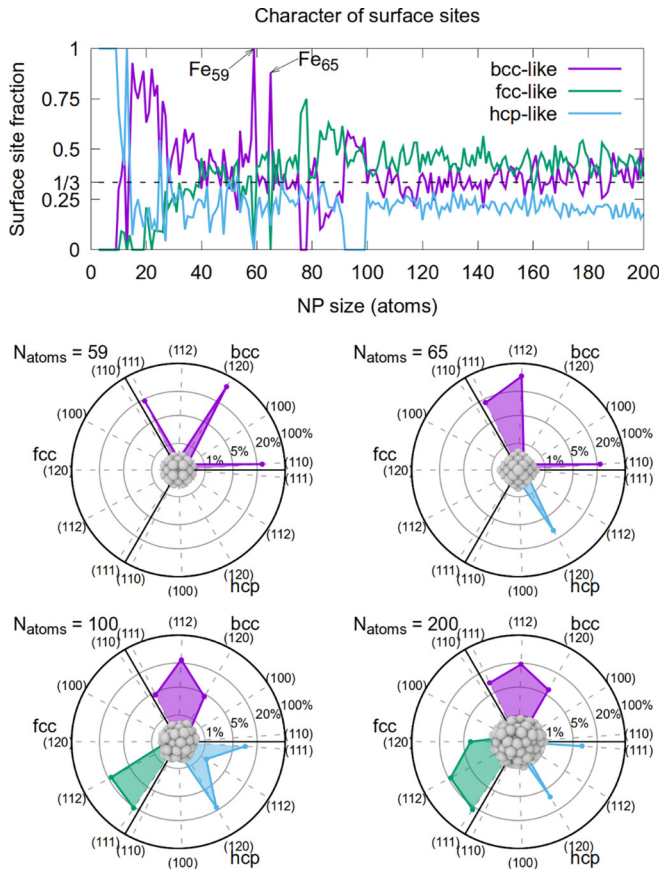
Character of surface sites



FIG. 13. (Top) Fraction of surface sites that resemble [(100), (110), (120), (112), and (111)] surface motifs from pristine bcc, fcc and hcp surfaces more than the others for all NPs in our convex hull. (Bottom) Log-scaled fraction of surface sites that resemble the reference surfaces most for four example NPs, two crystalline and two amorphous ones. A video with the panels corresponding to every NP is available on Zenodo [68].
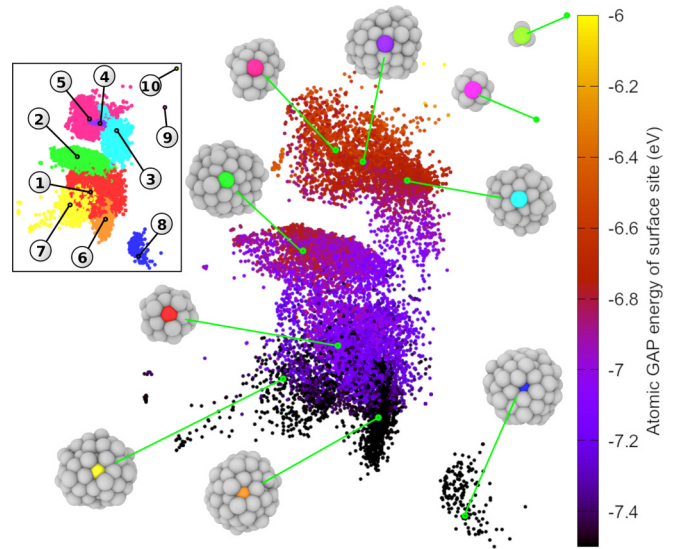


FIG. 14. `cl-MDS` representation (low-dimensional embedding) of the surface sites on the NPs in our convex hull clustered by $k$-medoids into ten characteristic motifs. The snapshots show the medoids representing the clusters. Encoded in the color is the atomic GAP energy of each surface site. The inset shows the same map color coded according to the $k$-medoids clusters.

by Ma *et al.* [62], BLYP/SDD by Aktürk *et al.* [63]) and are not easily comparable to our results. Hence, the convex hulls derived by Ma *et al.* and Aktürk *et al.* were omitted from Fig. 12.

In the remainder of this section, we try to uncover trends in the structure of these NPs, in particular regarding surface features. We first analyze the similarity between NP surface motifs and selected [(100), (110), (120), (112), and (111)] surface motifs present in pristine bcc, fcc and hcp surfaces. As a first step, we identify surface atoms in the NPs with a rolling-sphere algorithm as implemented in `ase_tools` [69]. Then, the SOAP descriptors characterizing the environment of these surface atoms within a 4 Å sphere are computed, as well as the SOAP descriptors of the atoms in the reference surfaces. Finally, we calculate the SOAP kernels between NP and surface descriptors, yielding a measure of similarity between 0 and 1. These kernels are used to rank the surface "character" of each NP as a histogram counting the number of motifs of each type divided by the total number of surface sites on that NP. The results of this analysis are given in Fig. 13 (top) for the overall bcc/fcc/hcp character, whereas the bottom panel of the figure shows four examples further resolving the surface character for $Fe_{59}$, $Fe_{65}$, $Fe_{100}$, and $Fe_{200}$.

From the figure we infer that, except for very small NPs with just a couple dozen atoms, the highly symmetric $Fe_{59}$ and $Fe_{65}$ corresponding to magic numbers, and the also highly symmetric $Fe_{76}$ and $Fe_{78}$, the motif distribution in these small NPs is close to random, oscillating around an equal distribution of bcc, fcc, and hcp sites up to $N_{atoms} = 100$. Beyond that, there are slightly more fcc sites and slightly less hcp sites. Given this degree of disorder, classifying surface sites in small Fe NPs in terms of the crystalline surface motifs is not very useful. Instead, we resort to a motif classification scheme that draws the classes directly from the database of structures. We do this using $k$-medoids, a data-clustering technique that separates data points into classes ("clusters") according to their similarity [70]. The most representative data point in each cluster is called a medoid. In our case, a collection of medoids provides a catalog of representative motifs found in our database [71]. We used the `fast-kmedoids` library for the $k$-medoids computation [72] and `cl-MDS` to graphically represent the clustering via low-dimensional embedding [73]. The results are shown in Fig. 14.

In the figure we classified the surface sites of all the lowest-energy NPs for a given number of atoms, from 3 to 200 atoms, into 10 data clusters. The size of the clusters decreases with their number, i.e., cluster 1 represents the most common surface motifs and 10 the least common. The color encodes the local GAP energy of the surface atoms in the big map, and the cluster index in the figure inset, for better reference. We observe "buried" (almost sub-surface) motifs to be the most stable (clusters 6, 7, and 8, ≈ −7.5 eV/atom). Then, motifs with a six-fold surface coordination (i.e., they appear to be at the center of a hexagon on the surface) but embedded within the NP facet are contained in clusters 1 (≈ −7.2 eV/atom) and 2 (≈ −6.9 eV/atom). Six-fold coordinated motifs that
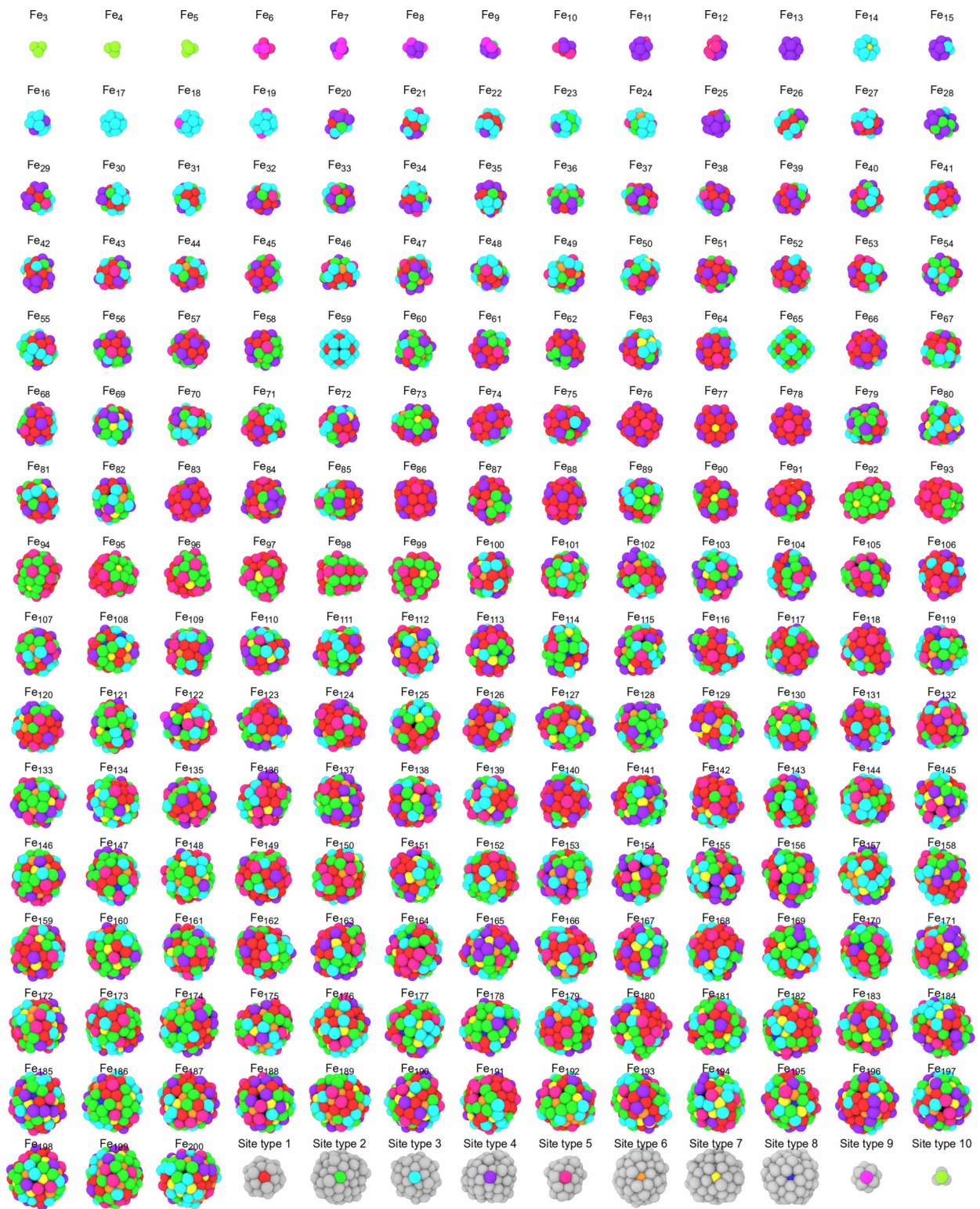
FIG. 15. Snapshots of all NPs from our convex hull, shown in Fig. 12. Atoms are colored according to the ten most characteristic surface site motifs according to the $k$-medoid clustering shown in Fig. 14. The medoids for the ten motifs are shown as well.

are raised further from the surrounding atoms are higher in energy, and contained within cluster 5 ($\approx -6.5$ eV/atom). Five-fold coordinated motifs are similar in energy to the latter, with the central atom similarly raised, and belong to clusters 3 and 4. Finally, clusters 9 and 10 contain just a handful of motifs found in the extremely small NPs.

The atomic GAP energy of bulk bcc iron is $\approx -8.25$ eV/ atom. The energy difference between this bulk value and the GAP energy of a less stable motif is directly related to the cohesive energy that could be gained by, e.g., increasing the coordination of the less stable motif. Therefore we expect the less stable motifs to readily

passivate while in contact with a surrounding medium, e.g., by forming strong bonds with and effectively capturing hydrogen atoms. Opposed to this, some of the motifs (especially those in clusters 6, 7, and 8) may be too stable to interact with adsorbants via covalent interactions. The sites with intermediate atomic GAP energies might be the most interesting from the catalytic point of view, e.g., because of their potential to adsorb or desorb reactants as a function of applied external bias. We will explore the precise relationship between adsorption energy of typical adsorbants and atomic GAP energies in subsequent work on Fe NP reactivity.

Finally, a gallery of all the NPs in our convex hull database is given in Fig. 15, with each surface site colored according to the data cluster to which it belongs (the reference motifs are also shown in the figure). We can easily observe that, except for the highly symmetric NPs at very small size and magic numbers 59 and 65, as well as the stability island between 76 and 78 atoms, the distribution of surface motifs is highly irregular. That is, there is no obvious facet formation in these NPs. Tests that we carried out for a significantly larger NP with a few thousand atoms, generated using the condensation method, also showed lack of significant facet formation. This contrasts with the very clear facet formation in other metal NPs, for instance (111) facets in Pt NPs as we have recently observed using very similar methodology [74]. A possible explanation for this is that structural disorder in iron NPs is driven by the interplay between the formation of the stable fcc surface facets versus the formation of the stable bcc bulk motifs. Since the bulk motif will nucleate facet formation with its same crystal structure and vice versa, this may lead to a nontrivial dynamics which in turn results in highly disordered NPs. Indeed, it has been shown experimentally that nanostructured Fe, e.g., Fe thin films on a substrate, can be grown in the fcc structure even at room temperature [75,76].

## VI. CODE AND DATA AVAILABILITY

The GAP is available for free on Zenodo [77] and can be used with QUIP/GAP, LAMMPS via the QUIP interface, ASE via QUIPPY, and TurboGAP. Incidentally, we note an improvement in computational efficiency of our GAP over the previous state-of-the-art Dragoni GAP by a factor of approximately 4. This speedup can be attributed mostly to the use of SOAP descriptor compression [78,79] in our GAP, as available from the `soap_turbo` descriptor [80]. When used

with the TurboGAP MD engine [28], better speedups can usually be achieved.

To facilitate further work in this area, we have made the structures of the NPs derived in this work available to the community. A full database is available for download on Zenodo [68], including the energies computed with the reference potentials and the Finnis-Sinclair EAM.

## VII. SUMMARY

In summary, we created a generally applicable GAP ML potential for the iron system which is stable in the whole configuration space and performs well for a wide range of applications, from bulk to nanostructured iron, from ambient conditions to those at the Earth's core. The accurate description of nanoparticles at elevated temperatures is particularly useful for the simulation of catalytic processes, which often occur at those temperatures. A straight-forward approach for such studies would be to combine our GAP with another GAP for the other involved species and extend the training databases with mixed structures, followed by some iterative training. While our GAP cannot beat previously existing specialized potentials in every case, it can be used reliably for most problems, including the study of systems where two or more Fe phases coexist. We found it to be the most accurate for NPs from among the potentials considered. We have derived a series of low-energy Fe NPs and made these structures available for further use. The GAP potential itself, which in addition to accuracy also achieves a sizable speedup over the previous state-of-the-art potential, has also been made freely available. We hope that this will enable and stimulate further work in this field, in particular with regard to catalytic applications on low-dimensional iron structures.

Some limitations remain from the implicit treatment of the magnetic states, especially regarding surface energies and elastic constants. These could be addressed by training a GAP including an explicit description of atomic magnetic moments. To this end, the necessary methodology and infrastructure to treat magnetism explicitly within the GAP framework needs to be developed.

## ACKNOWLEDGMENTS

[1] G. Cilpa-Karhu, O. J. Pakkanen, and K. Laasonen, Hydrogen evolution reaction on the single-shell carbon-encapsulated iron nanoparticle: A density functional theory insight, J. Phys. Chem. C **123**, 13569 (2019).

[2] M. A. Ahsan, A. R. Puente Santiago, A. Rodriguez, V. Maturano-Rojas, B. Alvarado-Tenorio, R. Bernal, and J. C. Noveron, Biomass-derived ultrathin carbon-shell coated iron nanoparticles as high-performance tri-functional HER, ORR and Fenton-like catalysts, J. Clean. Prod. **275**, 124141 (2020).

[3] H. M. Torres Galvis, J. H. Bitter, C. B. Khare, M. Ruitenbeek, A. I. Dugulan, and K. P. De Jong, Supported iron nanoparticles as catalysts for sustainable production of lower olefins, Science **335**, 835 (2012).

[4] B. Gu, C. Zhou, S. He, S. Moldovan, P. A. Chernavskii, V. V. Ordomsky, and A. Y. Khodakov, Size and promoter effects on iron nanoparticles confined in carbon nanotubes and their catalytic performance in light olefin synthesis from syngas, Catalysis Today **357**, 203 (2020).

[5] M. I. Mendelev, S. Han, D. J. Srolovitz, G. J. Ackland, D. Y. Sun, and M. Asta, Development of new interatomic potentials appropriate for crystalline and liquid iron, Philos. Mag. **83**, 3977 (2003).

[6] M. W. Finnis and J. E. Sinclair, A simple empirical N-body potential for transition metals, Philos. Mag. **50**, 45 (1984).

[7] D. Dragoni, D. Ceresoli, and N. Marzari, Vibrational and thermoelastic properties of bcc iron from selected EAM potentials, Comput. Mater. Sci. **152**, 99 (2018).

[8] L. B. Pártay, On the performance of interatomic potential models of iron: Comparison of the phase diagrams, Comput. Mater. Sci. **149**, 153 (2018).

[9] M. Müller, P. Erhart, and K. Albe, Analytic bond-order potential for bcc and fcc iron-comparison with established embedded-atom method potentials, J. Phys.: Condens. Matter **19**, 326220 (2007).

[10] M. Mrovec, D. Nguyen-Manh, C. Elsässer, and P. Gumbsch, Magnetic Bond-Order Potential for Iron, Phys. Rev. Lett. **106**, 246402 (2011).

[11] R. Alexander, L. Proville, C. S. Becquart, A. M. Goryeava, J. Dérès, C. Lapointe, and M. C. Marinica, Interatomic potentials for irradiation-induced defects in iron, J. Nucl. Mater. **535**, 152141 (2020).

[12] Y. Sun, M. I. Mendelev, F. Zhang, X. Liu, B. Da, C.-Z. Wang, R. M. Wentzcovitch, and K.-M. Ho, *Ab initio* melting temperatures of bcc and hcp iron under the Earth's inner core condition, Geophys. Res. Lett. **50**, e2022GL102447 (2023).

[13] V. L. Deringer, M. A. Caro, and G. Csányi, Machine learning interatomic potentials as emerging tools for materials science, Adv. Mater. **31**, 1902765 (2019).

[14] L. Cian, G. Lancioni, L. Zhang, M. Ianese, N. Novelli, G. Serra, and F. Maresca, Atomistic Graph Neural Networks for metals: Application to bcc iron, arXiv:2109.14012.

[15] F.-S. Meng, J.-P. Du, S. Shinzato, H. Mori, P. Yu, K. Matsubara, N. Ishikawa, and S. Ogata, General-purpose neural network interatomic potential for the α-iron and hydrogen binary system: Toward atomic-scale understanding of hydrogen embrittlement, Phys. Rev. Mater. **5**, 113606 (2021).

[16] D. Dragoni, T. D. Daff, G. Csányi, and N. Marzari, Achieving DFT accuracy with a machine-learning interatomic potential: Thermomechanics and defects in bcc ferromagnetic iron, Phys. Rev. Mater. **2**, 013808 (2018).

[17] L. Zhang, G. Csányi, E. V. D. Giessen, and F. Maresca, Atomistic fracture in bcc iron revealed by active learning of Gaussian approximation potential, arXiv:2208.05912v2.

[18] A. P. Bartók and G. Csányi, Gaussian approximation potentials: A brief tutorial introduction, Int. J. Quantum Chem. **115**, 1051 (2015).

[19] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, Gaussian process regression for materials and molecules, Chem. Rev. **121**, 10073 (2021).

[20] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.107.245421 for the gap_fit command used to train our GAP potential; an extension of the plot in Fig. 2 to larger atomic volumes; detailed cell parameters for all crystal structures mentioned in the main text; a plot like Fig. 6 showing the surface energies computed with our GAP compared to the Mendelev EAM; details on the calculations for the thermal expansion; two-phase calculations for the melting temperature with the Mendelev EAM; sample XRD spectra used to label

the phase diagram in Fig. 9; detailed settings for the Genetic Algorithm used to find low-energy NPs; a test on the accuracy of the Finnis-Sinclair EAM for NPs.

[21] G. Kresse and J. Hafner, Ab initio molecular dynamics for liquid metals, Phys. Rev. B **47**, 558 (1993).

[22] G. Kresse and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Comput. Mater. Sci. **6**, 15 (1996).

[23] G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).

[24] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. **77**, 3865 (1996).

[25] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B **50**, 17953 (1994).

[26] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B **59**, 1758 (1999).

[27] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, Phys. Rev. B **87**, 184115 (2013).

[28] M. A. Caro, Optimizing many-body atomic descriptors for enhanced computational performance of machine learning based interatomic potentials, Phys. Rev. B **100**, 024112 (2019), turbogap.fi.

[29] G. Csányi, S. Winfield, J. R. Kermode, A. De Vita, A. Comisso, N. Bernstein, and M. C. Payne, Expressive programming for computational physics in Fortran 950+, IoP Comput. Phys. Newsletter, Spring 2007 (2007).

[30] https://libatoms.github.io.

[31] H. C. Herper, E. Hoffmann, and P. Entel, Ab initio full-potential study of the structural and magnetic phase stability of iron, Phys. Rev. B **60**, 3839 (1999).

[32] H. Jónsson, G. Mills, and K. W. Jacobsen, Nudged elastic band method for finding minimum energy paths of transitions, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, Singapore, 1998), p. 385.

[33] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, The atomic simulation environment - a Python library for working with atoms, J. Phys.: Condens. Matter **29**, 30 (2017).

[34] V. L. Deringer and G. Csányi, Machine learning based interatomic potential for amorphous carbon, Phys. Rev. B **95**, 094203 (2017).

[35] J. R. Kermode, f90wrap: An automated tool for constructing deep Python interfaces to modern Fortran codes, J. Phys.: Condens. Matter **32**, 305901 (2020).

[36] S. Melchionna, G. Ciccotti, and B. L. Holian, Hoover NPT dynamics for systems varying in shape and size, Mol. Phys. **78**, 533 (1993).

[37] S. Melchionna, Constrained systems and statistical distribution, Phys. Rev. E **61**, 6165 (2000).

[38] D. Dragoni, D. Ceresoli, and N. Marzari, Thermoelastic properties of $\alpha$-iron from first-principles, Phys. Rev. B **91**, 104105 (2015).

[39] Z. S. Basinski, W. Hume-Rothery, and A. L. Sutton, The lattice expansion of iron, Proc. Roy. Soc. Lond. Ser. A **229**, 459 (1955).

[40] N. Ridley and H. Stuart, Lattice parameter anomalies at the Curie point of pure iron, J. Phys. D: Appl. Phys. **1**, 1291 (1968).

[41] J. R. Morris, C. Z. Wang, K. M. Ho, and C. T. Chan, Melting line of aluminum from simulations of coexisting phases, Phys. Rev. B **49**, 3109 (1994).

[42] J. R. Morris and X. Song, The melting lines of model systems calculated from coexistence simulations, J. Chem. Phys. **116**, 9352 (2002).

[43] S. Plimpton, Fast parallel algorithms for short–range molecular dynamics, J. Comput. Phys. **117**, 1 (1995).

[44] S. Nosé, A unified formulation of the constant temperature molecular dynamics methods, J. Chem. Phys. **81**, 511 (1984).

[45] W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, Phys. Rev. A **31**, 1695 (1985).

[46] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. **81**, 3684 (1984).

[47] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Bond-orientational order in liquids and glasses, Phys. Rev. B **28**, 784 (1983).

[48] S.-T. Lin, M. Blanco, and W. A. Goddard III, The two-phase model for calculating thermodynamic properties of liquids from molecular dynamics: Validation for the phase diagram of Lennard-Jones fluids, J. Chem. Phys. **119**, 11792 (2003).

[49] M. A. Caro, T. Laurila, and O. Lopez-Acevedo, Accurate schemes for calculation of thermodynamic properties of liquid mixtures from molecular dynamics simulations, J. Chem. Phys. **145**, 244504 (2016).

[50] M. A. Caro, O. Lopez-Acevedo, and T. Laurila, Redox potentials from ab initio molecular dynamics and explicit entropy calculations: application to transition metals in aqueous solution, J. Chem. Theory Comput. **13**, 3432 (2017).

[51] DoSPT code, http://dospt.org (last accessed: 2023-01-24).

[52] G. Bussi, D. Donadio, and M. Parrinello, Canonical sampling through velocity rescaling, J. Chem. Phys. **126**, 014101 (2007).

[53] The Debyer code, https://github.com/wojdyr/debyer.

[54] G. Morard, S. Boccato, A. D. Rosa, S. Anzellini, F. Miozzi, L. Henry, G. Garbarino, M. Mezouar, M. Harmand, F. Guyot, E. Boulard, I. Kantor, T. Irifune, and R. Torchio, Solving controversies on the iron phase diagram under high pressure, Geophys. Res. Lett. **45**, 11 (2018).

[55] K. Hirose, S. Labrosse, and J. Hernlund, Composition and state of the core, Annu. Rev. Earth Planet Sci. **41**, 657 (2013).

[56] L. E. González and D. J. González, Structure and dynamics in liquid iron at high pressure and temperature. A first principles study, J. Geophys. Res. Solid Earth. **128**, e2022JB025119 (2023).

[57] G. R. Weal, S. M. Mcintyre, and A. L. Garden, Development of a structural comparison method to promote exploration of the potential energy surface in the global optimization of nanoclusters, J. Chem. Inf. Model. **61**, 1732 (2021).

[58] G. Z. Wulff, Zur Frage der Geschwindigkeit des Wachstums und der Auflösung der Kristallflächen, Kristallogr. Mineral. 449 (1901).

[59] M. Gjerding, T. Skovhus, A. Rasmussen, F. Bertoldo, A. H. Larsen, J. J. Mortensen, and K. S. Thygesen, Atomic simulation recipes: A python framework and library for automated workflows, Comput. Mater. Sci. **199**, 110731 (2021).

[60] P. Bobadova-Parvanova, K. A. Jackson, S. Srinivas, M. Horoi, C. Köhler, and G. Seifert, Scanning the potential energy surface of iron clusters: A novel search strategy, J. Chem. Phys. **116**, 3576 (2002).

[61] C. Köhler, G. Seifert, and T. Frauenheim, Density functional based calculations for Fe$_n$ ($n \leqslant 32$), Chem. Phys. **309**, 23 (2005).

[62] Q. M. Ma, Z. Xie, J. Wang, Y. Liu, and Y. C. Li, Structures, binding energies and magnetic moments of small iron clusters: A study based on all-electron DFT, Solid State Commun. **142**, 114 (2007).

[63] A. Aktürk and A. Sebetci, BH-DFTB/DFT calculations for iron clusters, AIP Adv. **6**, 055103 (2016).

[64] A. P. Sutton and J. Chen, Long-range Finnis–Sinclair potentials, Philos. Mag. Lett. **61**, 139 (1990).

[65] J. A. Elliott, Y. Shibuta, and D. J. Wales, Global minima of transition metal clusters described by Finnis–Sinclair potentials: A comparison with semi-empirical molecular orbital theory, Philos. Mag. **89**, 3311 (2009).

[66] T. D. Liu, T. E. Fan, J. W. Zheng, G. F. Shao, Q. Sun, and Y. H. Wen, Structural optimization of Fe nanoclusters based on multi-populations differential evolution algorithm, J. Nanopart. Res. **18**, 77 (2016).

[67] D. J. Wales, J. P. K. Doye, A. Dullweber, M. P. Hodges, F. Y. Naumkin, F. Calvo, J. Hernández-Rojas, and T. F. Middleton, The Cambridge Energy Landscape Database, http://www-wales.ch.cam.ac.uk/CCD.html, accessed: 2022-07-05.

[68] Zenodo repository containing the nanoparticles in our convex hull, https://doi.org/10.5281/zenodo.7632315 (2023).

[69] R. Ibragimova and M. A. Caro, `ase_tools`: Fast Fortran libraries to be used in combination with ASE, https://github.com/mcaroba/ase_tools (accessed January 27, 2023).

[70] C. Bauckhage, *Numpy/scipy Recipes for Data Science: k-Medoids Clustering*, Tech. Rep. (University of Bonn, 2015).

[71] M. A. Caro, A. Aarva, V. L. Deringer, G. Csányi, and T. Laurila, Reactivity of amorphous carbon surfaces: Rationalizing the role of structural motifs in functionalization using machine learning, Chem. Mater. **30**, 7446 (2018).

[72] M. A. Caro and P. Hernández-León, fast-kmedoids, https://github.com/mcaroba/fast-kmedoids (accessed March 16, 2022).

[73] P. Hernández-León and M. A. Caro, Cluster-based multi-dimensional scaling embedding tool for data visualization, arXiv:2209.06614.

[74] J. Kloppenburg, L. B. Pártay, H. Jónsson, and M. A. Caro, A general-purpose machine learning Pt interatomic potential for an accurate description of bulk, surfaces and nanoparticles, J. Chem. Phys. **158**, 134704 (2023).

[75] D. Li, M. Freitag, J. Pearson, Z. Q. Qiu, and S. D. Bader, Magnetic Phases Of Ultrathin Fe Grown On Cu(100) As Epitaxial Wedges, Phys. Rev. Lett. **72**, 3112 (1994).

[76] D. J. Keavney, D. F. Storm, J. W. Freeland, I. L. Grigorov, and J. C. Walker, Site-Specific Mössbauer Evidence of Structure-Induced Magnetic Phase Transition in fcc Fe(100) Thin Films, Phys. Rev. Lett. **74**, 4531 (1995).

[77] Zenodo repository containing the GAP potential files, https://doi.org/10.5281/zenodo.7630369 (2023).

[78] J. P. Darby, J. R. Kermode, and G. Csányi, Compressing local atomic neighbourhood descriptors, npj Comput. Mater. **8**, 166 (2022).

[79] J. P. Darby, D. P. Kovács, I. Batatia, M. A. Caro, G. L. W. Hart, C. Ortner, and G. Csányi, Tensor-reduced atomic density representations, arXiv:2210.01705.

[80] soap_turbo, https://github.com/libAtoms/soap_turbo (last accessed: 2022-11-08).