Data-driven simulation and characterisation of gold nanoparticle melting

Claudio Zeni,[1,3,8] Kevin Rossi,[1,4] Theodore Pavloudis,[2,5] Joseph Kioseoglou,[5] Stefano de Gironcoli,[3] Richard E. Palmer,[6] and Francesca Baletto[1,7]

[1])*Department of Physics, King's College London, London, WC2R 2LS, UK*

[2])*College of Engineering, Swansea University, Bay Campus, Fabian Way, Swansea, SA1 8EB,*
UK

[3])*Present address: International School for Advanced Studies, Via Bonomea, 265, 34136, Trieste,*
IT

[4])*Present address: Laboratory of Nanochemistry, Institute of Chemistry and Chemical Engineering,*
Ecole Polytechnique Fédérale de Lausanne, Lausanne, CH

[5])*Present address: Department of Physics, Aristotle University of Thessaloniki, Thessaloniki GR-54124,*
GR

[6])*Present address: College of Engineering, Swansea University, Bay Campus, Fabian Way, Swansea, SA1 8EB,*
UK

[7])*Present address: DIPC, Paseo Manuel de Lardizabal, 20018 San Sebastian, Spain*

[8])*email: czeni@sissa.it*

ABSTRACT

The simulation and analysis of the thermal stability of nanoparticles, a stepping stone towards their application in technological devices, require fast and accurate force fields, in conjunction with effective characterisation methods. In this work, we develop efficient, transferable, and interpretable machine learning force fields for gold nanoparticles based on data gathered from Density Functional Theory calculations. We use them to investigate the thermodynamic stability of gold nanoparticles of different sizes (1 to 6 nm), containing up to 6266 atoms, concerning a solid-liquid phase change through molecular dynamics simulations. We predict nanoparticle melting temperatures in good agreement with available experimental data. Furthermore, we characterize the solid-liquid phase change mechanism employing an unsupervised learning scheme to categorize local atomic environments. We thus provide a data-driven definition of liquid atomic arrangements in the inner and surface regions of a nanoparticle and employ it to show that melting initiates at the outer layers.

Keywords: Machine Learning, Force Fields, Molecular Dynamics, Nanoparticles, Phase Change, Melting, Gold, Au

INTRODUCTION

Gold (Au) nanoparticles (NPs) find widespread application in many technological areas, such as in optics,[1,2] nanomedicine,[3,4] and catalysis[5–9]. As all chemo-physical properties of Au NPs depend on their shape, the analysis of their structural stability has attracted a lot of attention in the past years. A deep understanding of the liquid-solid phase change mechanisms of Au NP, also accounting for surface rearrangements, may, in particular, be crucial for catalytic applications, where the reaction conditions often demand the nanocatalysts to work at high temperatures while preserving their size and shape.

Numerical simulations can, in principle, offer a platform to investigate and characterize phase change mechanisms of NPs at an atomistic level. However, two long-standing challenges must be overcome to improve the numerical predictions of NPs' thermal stability. The first concerns the difficulty of defining an unbiased characterisation of the phase change mechanism. Indeed the identification of order parameters to characterise solid-liquid phase changes at the nanoscale is an active topic of debate with a long tradition.[10–13] Widely used methods often rely on chemical-intuition and heuristic approaches, and can therefore lead to descriptive order parameters which are neither fully general nor robust to parameter tuning. For example, changes in the first neighbour distance distribution affect the definition of coordination number

too drastically,[14,15] and little research has been carried out on the characterization of local atomic environments peculiar of NP's surface atoms.

The second challenge is related to the development of accurate and fast interparticle potentials, which reproduce the complexity of the NPs' energy landscape. In so far, atomistic modelling methods have offered a strict trade-off between computational speed and accuracy. While simulations based on electronic structure methods, such as density functional theory (DFT), provide quantitative accuracy, their computational cost severely limits the capabilities to generate dynamical trajectories of large systems and for long times. On the contrary, large systems and long simulation timescales are easily accessible when employing semi-empirical potentials. Nevertheless, such methods do not necessarily provide a quantitative insight on the chemistry of NPs' phase changes[16] because their analytical functional form limits their predictive power and flexibility. Furthermore, these potentials are often fitted to bulk properties, which poses an additional limit to their accuracy when simulating nanoscale systems[17].

In this work, we tackle these two challenges by adopting data-driven methods to generate an accurate and efficient description of interatomic potentials, and by developing an automated routine that classifies the atomic environments observed during Au NPs' phase change. To obtain long, i.e., hundreds of ns, and accurate trajectories during melting of Au NPs of variable sizes, we develop a set of machine learn-

ing force fields (ML-FFs)[18–25] using the innovative framework of mapped Gaussian processes[26–28]. ML-FFs can approximate the force-energy predictions yielded by the reference DFT method they are trained upon while being many orders of magnitude faster to compute. Here, we train ML-FFs on LDA-DFT data, rPBE-DFT data, and contrast our results with experimental results, and with predictions found using a semi-empirical interatomic potential. To characterise the melting kinetics, we adopt an unsupervised machine learning clustering scheme which discriminates in an automatic fashion locally liquid from locally solid environments, surface from inner environments, and high-coordination from low-coordination surface environments. We then obtain a route to estimate the NPs melting temperature by monitoring the relative population of liquid atoms in the nanoparticle, and the melting mechanism by recording the spatial distribution of locally liquid environments as a function of temperature. We employ these data-driven tools to study the melting of Au NPs with diameters between 1 and 6 nm, and various initial geometrical shapes. We univocally show that melting initiates in the outermost layer of Au NPs first, and occurs in the NPs' core second.

RESULTS

Machine Learning Force Fields

To construct a training database, we extract 7 random decorrelated frames from ab initio molecular dynamics trajectories where an Au NP containing 309 atoms (∼2 nm of diameter) with an initial face centred cubic (FCC) morphology undergoes melting. We calculate, for each frame, forces and energies at DFT LDA and DFT GGA-rPBE levels, and utilize a 2+3-body mappable Gaussian process regression framework[26,27] to fit two ML-FFs, one for each DFT method; further detail is provided in the Methods and in the Supplementary Methods. When training on 2100 local atomic environments, our ML-FFs incur in a mean absolute error (MAE) on the force components of 0.09 ± 0.04 eV/Å (LDA ML-FF) and 0.07 ± 0.03 eV/Å (rPBE ML-FF), and in a MAE on the atomic energy differences of 2.65 ± 2.02 meV/atom (LDA ML-FF) and 1.98 ± 1.76 meV/atom (rPBE ML-FF), on validation sets disjointed from the training sets. The reported accuracy is comparable to the ones quoted in previous studies[29–35], and is deemed satisfactory. This training dataset, albeit small, contains a heterogeneous set of local atomic environments, as shown in Supplementary Figure 1; and we, therefore, consider it to be representative for Au NPs in the size range of interest.

We test the accuracy of the two ML-FFs on a more complex dataset, which encompasses NPs' architectures of different sizes and shapes (see Supplementary Methods). Supplementary Table 1 reports the mean absolute errors (MAEs) on force components and atomic energies incurred by each of the ML-FFs developed on these validation datasets. The MAEs on force components are again consistently around 0.1 eV/Å, and the MAEs on atomic energy differences are consistently lower

than 10 meV/atom. The ML-FFs are, therefore, considered accurate enough and, more importantly, transferable across different NPs' sizes and shapes. This holds regardless of the DFT level of theory used to train the ML-FF (GGA PBE and LDA) and its implementation (VASP projector augmented-wave and CP2K Gaussian plane wave).

When validating the ML-FFs against the experimental bulk cohesive energy (Supplementary Figure 4), we observe that LDA (rPBE) based ML-FF overestimates (underestimates) this quantity. We then adopt a parametric mixing of the two ML-FFs (see also the Methods section) and generate a third ML-FF, labelled hybrid, which, by construction, has cohesive energy in the bulk phase that matches the experimental one. The 2- and 3-body FFs forming the three ML-FFs present some noticeable differences; in Supplementary Figure 5 we show how the LDA ML-FF is more bound and stiffer than the rPBE ML-FF, and how the hybrid ML-FF has, as expected, a shape that is in-between the one of the other two ML-FFs.

Phase Change Characterization

Following the successful training and validation of our ML-FFs, we employ them to study the size-dependent melting temperature of Au NPs. We consider NPs whose diameter ranges from 1 to 6 nm, corresponding to NPs containing 147, 309, 561, 923, 2869, and 6266 atoms. We sample the NPs' evolution in a temperature range between 400K and 1600K when subject to a heating rate of 20 K/ns. We also test 5 K/ns and 10 K/ns heating rates and do not observe significant changes in the melting temperature estimate (see Supplementary Figure 6). For each NP size and ML-FF, we simulate Au NPs for a total of 2.4 ms, a time-length not accessible to electronic structure calculations, even for the largest state-of-the-art computational facilities. We refer the interested reader to the Methods section for further details on the numerical setup used to perform the simulations.

To characterize the solid-liquid phase transition, while distinguishing between surface and inner melting, we adopt an unsupervised machine learning approach that hinges on a small database of configurations randomly extracted from the phase change trajectories we simulated and a local atomic density representation. In particular, we employ a modified version[36] of the 3-body local atomic cluster expansion descriptor[37] to associate a 40-dimensional set of features to each atom. We then exploit a hierarchical k-means clustering scheme to isolate six classes of local atomic environments (see also the Methods and Supplementary Methods). This labels the local atomic environments as being in solid or liquid phase, and as belonging to the inner, high-coordination surface, and low-coordination surface motifs. As illustrated in panels b and c of Figure 1 and Supplementary Figures 6 and 13, both the number of nearest neighbours (#NN) within a predefined cut-off and the nominal MD simulation temperature at which these are sampled, correlate with the labels assigned by the clustering algorithm.

The six local atomic environment classes showcased in Figure 1 and Supplementary Figures 8 and 9, can be characterised
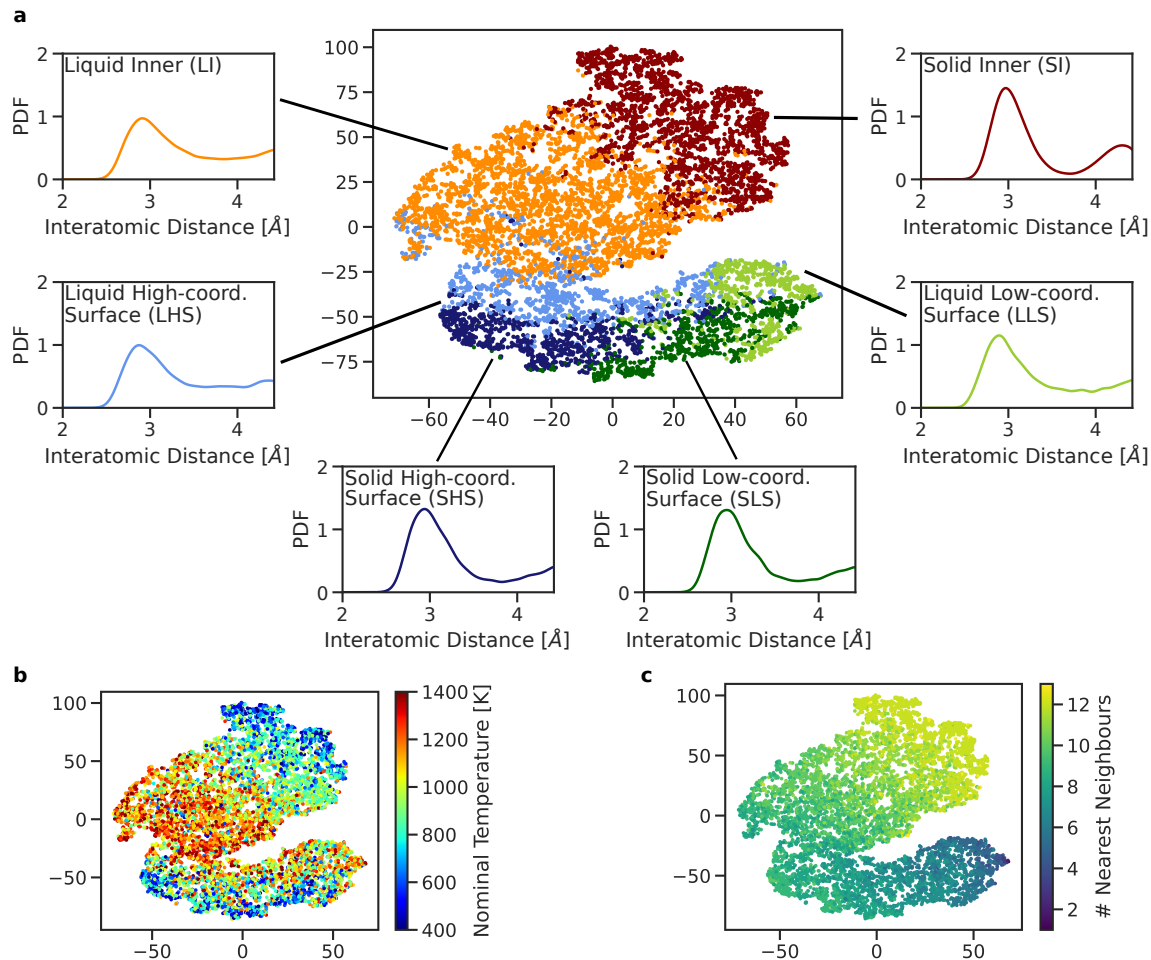
FIG. 1. **Features of the six classes of local atoms environments identified through clustering.** Visualization of the hierarchical k-means clustering results for MD simulations of Au nanoparticles with 147, 309, 561, 923, 2869 and 6266 atoms, carried out using the ML-FF trained on rPBE-DFT data. Panel a: 1st and 2nd component (x- and y-axis) of the t-sne projection of the atomic expansion coefficients of 10^4 local atomic environments randomly sampled from melting MD simulations. The colours label the six classes assigned by the hierarchical k-means clustering algorithm, as defined in the main text. The normalized average pair-distance distribution function (PDF) belonging to each class is shown and coloured accordingly. Panels b and c: same t-sne projection as in panel a). In panel b, the colours indicate the nominal simulation temperature at which the local atomic environment was taken from, in panel c, the number of nearest neighbours (#NN) computed using a cut-off of 3.6 Å.

from the number of neighbours within a given cut-off (here taken as 3.50 Å for the LDA ML-FF, 3.75 Å for the rPBE ML-FF, and 3.60 Å for the hybrid ML-FF) they display, and from features in their pair-distance distribution function (PDF). In detail:

- Solid inner (SI) atoms have 12 NN within the chosen cut-off, and their PDF displays a well-defined peak at a second NN distance consistent with the one of bulk FCC Au. SI local atomic environments comprise FCC-like motifs, as well as motifs with 5-fold- or icosahedral-symmetry.

- Liquid inner (LI) atoms have, on average, 11 NNs. The PDF for this class of local atomic environment presents the first peak at distances lower than the one for bulk lattice and lacks a pronounced second peak in corre-

spondence to the bulk lattice one.

- Solid High-coordination Surface (SHS) atoms present, on average, 8 NNs, and peaks its PDF in correspondence to the second nearest-neighbours (lattice bulk).

- Liquid High-coordination Surface (LHS) atoms also have 8 NNs on average, yet the PDF lacks a peak at the bulk lattice.

- Solid Low-coordination Surface (SLS) atoms find an average of 6.9 atoms at a distance consistent with the bulk NNs distance.

- Liquid Low-coordination Surface (LLS) atoms have, on average, 6.0 atoms at a distance lower than the bulk nearest-neighbours distance; furthermore, the PDF does not display any peak for the second NNs.

This unsupervised approach enables an original and data-driven definition of liquid atomic arrangements in the inner part of the NPs and at the surface. Local atomic environments in a liquid phase are all characterized by the absence of a peak of their PDF in correspondence to their bulk lattice distance (i.e., the one predicted by the reference interatomic potential). This observation holds regardless of whether they lie in the inner part or at the surface of the NP, and their coordination. Furthermore, this result confirms and rationalises the universal signature of melting for the whole NP we recently proposed[13].

Having discriminated in automated fashion atoms in liquid and solid environments, also as a function of their spatial location in the NP, we draw novel definitions to determine melting phase changes in the nanoparticle. To this end we monitor the time evolution of the occurrence of liquid environments, their rate of variation, and at which temperature their relative population increases above the 0.4 of the total, also as a function of their distance from the center of mass of the NP. In the following, we refer to the melting temperature of the NP ($T_{\text{melt}}^{\text{NP}}$) as the temperature at which the number of inner atoms that are identified as liquid (#LI) by the clustering algorithm displays the maximum positive derivative. This melting temperature estimation method yields equivalent results w.r.t. other well-established algorithms to calculate the melting temperature, such as the caloric curve maximum derivative and the heat capacity peak (see Supplementary Figure 13). This observation further corroborates our trust in the clustering algorithm as a tool to characterise Au nanoparticles melting.

Size-dependent melting

Figure 2 reports the $T_{\text{melt}}^{\text{NP}}$ for NPs of different sizes as a function of the NPs' reciprocal radius, as found during MD simulations carried out with the LDA, rPBE and hybrid ML-FFs. The reported $T_{\text{melt}}^{\text{NP}}$ is averaged over the 4 (2 for NPs with more than 2500 atoms) independent MD simulations carried out for each NP and each ML-FF. For an immediate comparison, we report the experimental melting temperature of bulk FCC Au at atmospheric pressure (T_{melt}^{bulk}), and the experimental melting temperatures of Au NPs as a function of the NP size[38,39]. For reference, we add the $T_{\text{melt}}^{\text{NP}}$ estimates obtained using a classical MD where the interatomic interaction is derived in the second-moment approximation of the tight-binding (TB-SMA)[13]. All the ML-FFs lead to $T_{\text{melt}}^{\text{NP}}$ predictions which are (as expected) lower than the ones found during experiments for C-supported Au NPs (pink squares in Figure 2). On average, the rPBE-derived ML-FF predicts $T_{\text{melt}}^{\text{NP}}$ 250±50 K lower than the ones predicted by the LDA-derived ML-FF, and 180±40 K lower than the ones predicted by the hybrid ML-FF. Interestingly, the $T_{\text{melt}}^{\text{NP}}$ predicted by the hybrid ML-FF are less than 50 K away from the melting temperatures found experimentally via differential scanning calorimetry measurements.[39]

Melting mechanism characterisation

In the previous section we established the quantitative agreement between the ML-FFs' predictions and the experimental melting temperatures of Au NPs, also as a function of their size. It is then natural to proceed further and analyse the mechanism by which phase changes occur.

To this end, we display in Figure 3 example snapshots of an Au 6266 NP at different temperatures (panel a), and the temperature-dependent radial distribution of the fraction of LI (#LI/#tot, panel b) and of LS (#LS/#tot, panel c) local atomic environments. The # symbol indicates the number of atoms belonging to a certain class, where we define: $\#LS = \#LHS + \#LLS$, and $\#tot = \#LHS + \#LLS + \#LI + \#SHS + \#SLS + \#SI$. The results we report are found by averaging over the set of independent MD melting simulations employing the rPBE-based ML-FF. We refer the interested Reader to Supplementary Figures 14 and 15 for the same plots for all systems with 147, 309, 561, 923, 2869 and 6266 atoms and using the three ML-FFs.

In Figure 3 and Supplementary Figures 14 and 15, the large majority of the local atomic environments are correctly labelled as solid (liquid) at the start (end) of each MD simulation. The average occurrence of all LI atoms increases with temperature, reaching around 0.5 at the $T_{\text{melt}}^{\text{NP}}$ independently
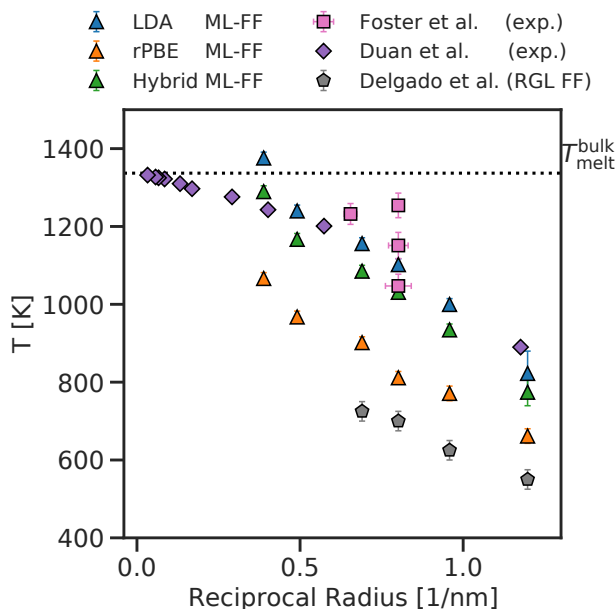

FIG. 2. **Melting temperatures of Au NPs of different sizes.** Average $T_{\text{melt}}^{\text{NP}}$ as a function of NP's reciprocal radius computed for MD simulations employing the LDA-trained (blue triangles), rPBE-trained (orange triangles), and hybrid (green triangles) ML-FFs. Experimental data for size-selected Au NPs supported on carbon (pink squares) and spherical Au NPs (purple diamonds), is taken from Foster et al.[38] and Duan et al.[39], respectively. Grey pentagons refer to the $T_{\text{melt}}^{\text{NP}}$ estimates from TB-SMA iterative MD melting simulations from Delgado-Callico et al.[13]. Error bars indicate the standard deviation of the melting temperature estimations, and of the NP sizes for experimental data taken from Foster et al.[38] (pink squares).
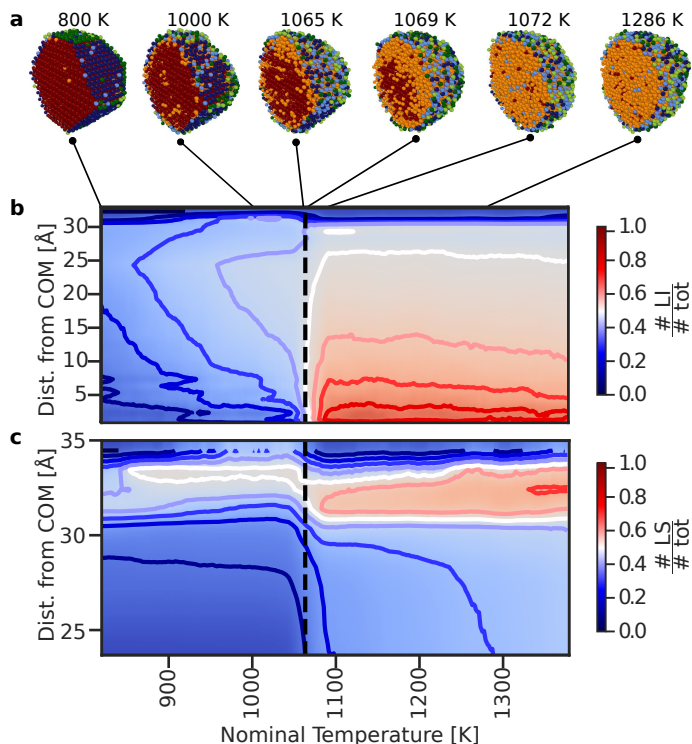
FIG. 3. **Distribution of liquid environments in an Au 6266 NP.**
Panel a: snapshots of Au 6266 simulated using the r-PBE ML-FF at different nominal simulation temperatures, with atoms coloured according to the clustering algorithm, and using the same colour scheme as in Figure 1. Panels b and c: average fraction of #LI (b) and #LS (c) local atomic environments as a function of the radial distance from the COM (y coordinate), and of the nominal system temperature (x coordinate). The bold coloured lines in panels b and c indicate the isosurfaces in the plot, from 0 to 1 every 0.1, while the black dashed line indicates the T_{melt}^{NP} of 1065 K.

of their distance from the NPs' centre of mass (COM). Areas located few Å below the NPs' surface instead display significant abundances of LI atoms also at temperatures below T_{melt}^{NP}. Such observation is in line with experimental results by Foster et al. [38], where a surface melting ($T_{melt}^{surf.}$) temperature below the T_{melt}^{NP} was observed for Au NPs of sizes comparable to the ones we analyse. This $T_{melt}^{surf.}$ was determined in Foster et al. [38] by taking the average between the onset temperature for shape changes visible via aberration-corrected scanning transmission electron microscope and the highest temperature for which these did not occur.

To compare our results with available experimental data, we would like to introduce a numerical definition of $T_{melt}^{surf.}$. In analogy to the T_{melt}^{NP} definition, $T_{melt}^{surf.}$ should be defined as the temperature at which a clear discontinuity appears in the temperature-dependent evolution of the abundances of liquid-like atoms at the surface of the nanoparticle. This is, however, not advisable. While the number of LI atoms has a clear and distinct positive jump – which allows us to define a T_{melt}^{NP} (Supplementary Figure 12) – the temperature-dependent evolution of the number of LS does not show such a clear first-order
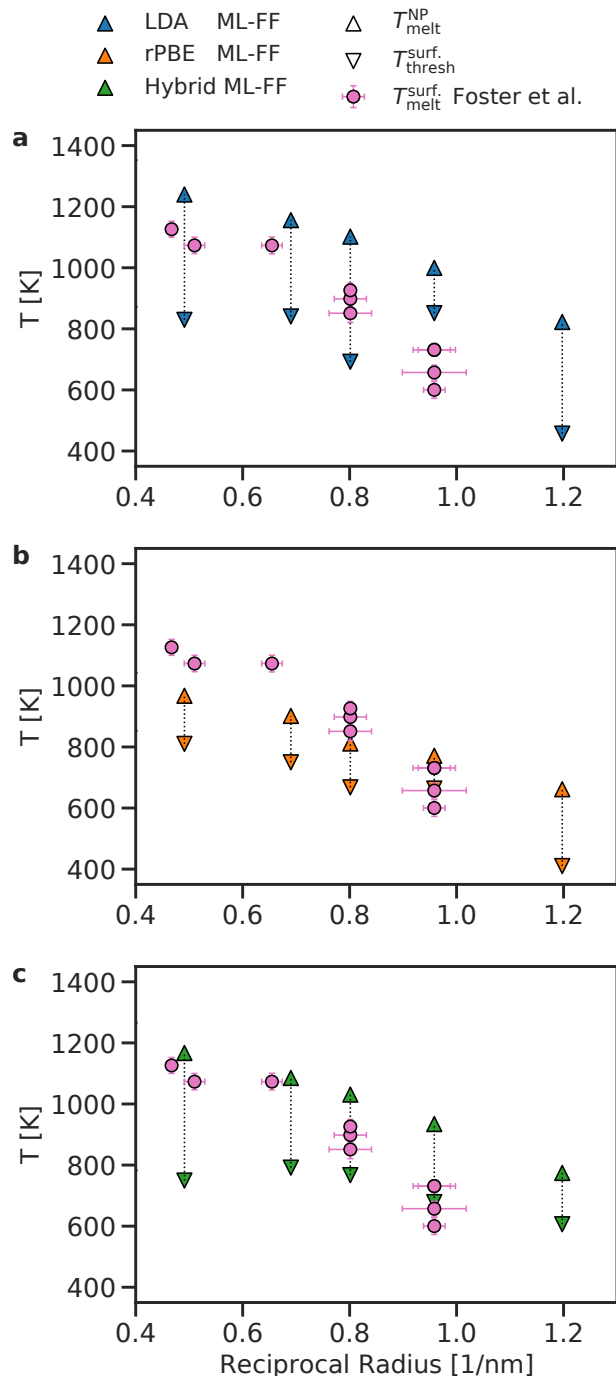

FIG. 4. **Surface phase change temperatures of Au NPs of different sizes.** $T_{melt}^{surf.}$ (downward triangles) and T_{melt}^{NP} (upward triangles) as a function of NPs' reciprocal radius, for MD simulations carried out using the LDA ML-FF (panel a), rPBE ML-FF (panel b), and hybrid ML-FF (panel c). Experimental estimates of $T_{melt}^{surf.}$ from high-resolution TEM measurements are taken from Foster et al. [38], and are shown as pink circles. Error bars indicate the standard deviation of the melting temperatures and of the NP sizes for experimental data taken from Foster et al. [38] (pink circles).

transition (Supplementary Figure 17). The relative amount of LS atoms increases gradually with temperature for all NP

sizes and all ML-FFs, and reaches values around 0.5 (white line in panel (c) of Figure 3 and Supplementary Figures 14 and 15) for atoms in the surface layer at temperatures approaching $T_{\mathrm{melt}}^{\mathrm{NP}}$.

We, thus, abandon the search for an unbiased definition of $T_{\mathrm{melt}}^{\mathrm{surf.}}$ and introduce the quantity $T_{\mathrm{thresh}}^{\mathrm{surf.}}$, which provides an indication of the temperature at which significant surface rearrangement occurs. The latter is defined as the lowest temperature where at least 0.4 of the local atomic environments in the surface of the NP are classified as liquid (see the Methods section for additional details). Figure 4 reports the values of $T_{\mathrm{thresh}}^{\mathrm{surf.}}$ and the $T_{\mathrm{melt}}^{\mathrm{NP}}$ for all our MD simulations, and the experimental $T_{\mathrm{melt}}^{surf.}$ as reported in Foster et al.[38]. The temperature ranges comprised between $T_{\mathrm{thresh}}^{\mathrm{surf.}}$ and $T_{\mathrm{melt}}^{\mathrm{NP}}$ are in line with the experimentally reported $T_{\mathrm{melt}}^{\mathrm{surf.}}$, this is especially true for the case of the hybrid ML-FF.

To deepen our understanding of the melting mechanism, we also calculate the mean first-passage temperature (MFPT) required to observe the transition to a liquid phase of 0.4 of the atoms which initially resided at a nanoparticle edge, on a (100) surface, on a (111) surface, or in a bulk environment. These families of atoms are discriminated against according to their number of first NNs at the beginning of the MD simulations. Edge atoms (#NN=6) are more likely to move into a liquid phase than atoms on a (100) facet (#NN=8), which in turn are more prone to end into a liquid phase than the atoms on a (111) facet (#NN=9). The inner atoms (#NN=12) present the overall largest MFPT. This finding is coherent with the melting initiating from the outer layers of the NP (additional details are available in the Methods, and in the Supplementary Methods).

The trends observed during the melting characterization indicate that local phase changes in the outermost layer of a NP start to occur at temperatures a few hundred of K below the $T_{\mathrm{melt}}^{\mathrm{NP}}$. For Au NPs, the proposed characterization protocol establishes that local solid to liquid changes first initiate at low-coordinated atoms at the vertices and edges, then propagate to atoms on (100) and (111) facets, and finally proceed to the inner region of the nanoparticle.

DISCUSSION

We characterize the melting mechanism in gold nanoparticles of size 1-6 *nm*, and predict the melting temperatures in good agreement with experimental data using molecular dynamics. These simulations employ machine learning force fields, under the mapped Gaussian process framework, to surpass the trade-off between accuracy and cost in traditional atomistic modelling methods. We showcase that accurate, efficient, and size-transferable force fields can be trained using small training datasets. We additionally generate a hybrid 2+3-body force field by linearly combining two machine learning force fields fitted on data computed using different density functional theory functionals; this force field is parametrised to reproduce the bulk cohesive energy and yields predictions of melting temperatures of Au nanoparticles in striking agreement with available experimental data.

To elucidate the melting mechanism, we subsequently develop a general unsupervised clustering approach to differentiate between inner and surface layers and to characterise the phase change at the atomistic level. Thanks to the insight offered by the proposed clustering algorithm, we demonstrate that the melting transition initiates at the outer layer, and later spreads to the inner region. The increase in locally liquid environments in the outer region of the nanoparticle before the melting of its core finds a parallel with what is generally referred to in the literature as surface melting. The predicted trend is in very good agreement with our experimental observations, where melting was found to start at the outermost layer, at a temperature few hundred of K lower than the NP melting. We verify that such a melting mechanism occurs regardless of the force-field used to model interatomic interactions, but we also find that different force fields predict different surface and nanoparticle melting temperatures. We expect that the data-driven simulation and characterisation methods developed here, and the insight we obtain, will stimulate and benefit other research aimed at addressing the complexity of phase changes (solid-to-liquid and liquid-to-solid alike) at the nanoscale.

METHODS

Database Construction

To construct the training set, we randomly sample 7 frames from a set of 60 frames extracted at regular time intervals from an *ab initio* MD trajectory where an Au NP containing 309 atoms (approx 2 nm in diameter) with an initial FCC morphology undergoes melting from 300 K to 1200 K. Atomic forces and energy associated with each configuration are calculated within the density functional theory framework, and employing LDA and GGA-rPBE pseudopotentials to generate the training sets for the LDA an rPBE ML-FFs, respectively. The training sets we employ therefore contain 2163 local atomic environments and associated forces, and 7 total energy values, one for each structure. When assessing learning curves (Supplementary Figure 2) we find, in agreement with previous reports[26,27], that the MAE on force prediction converges for training databases which encompass a few hundreds of local atomic environments, and energy predictions do so when energies of a handful of configurations are utilized. We furthermore note (Supplementary Figure 3) that the shape of the 2-body part of the ML-FFs resulting from training encompassing few hundreds of local atomic environments remains, in essence, unchanged when the number of training points is increased.

We generate a validation set by extracting de-correlated frames from MD trajectories previously reported in Delgado-Callico et al.[13], and from *ab initio* MD trajectories previously reported in Foster et al.[38]. We sample the melting MD trajectories reported in Delgado-Callico et al.[13], carried out using a second-moment tight-binding potential, from 400 K up to 1200 K, and increasing iteratively the temperature of 25 K every 5 ns. For this setup, we consider NPs contain-

ing 146, 147, 192, and 201 atoms which present initial different closed-shell geometries, namely octahedron (146 atoms), icosahedron (147 atoms), Marks decahedron (192 atoms), and regular-truncated octahedron (201 atoms). The NPs undergo both solid-solid and solid-liquid rearrangements during these MD trajectories (for more details see also the original reference[13]). The melting MD trajectories reported in Foster et al.[38] are carried out via NVT simulations, as in the VASP suit, performed at temperatures from 300 to 1200 K with a 150 K interval using LDA DFT, for Au NPs containing 147, 309, and 561 atoms starting from a cuboctahedron.

Machine Learning Force Fields Construction

We construct the ML-FFs for Au by applying the framework of mappable few-body FFs trained via Gaussian Process regression (GPR)[26,27] using the FLARE Python Package.[25,28] GPR FFs hinge on the nearsightedness principle of quantum mechanics to predict total energies for a system of atoms S as a sum of local atomic energy contributions $\varepsilon_i(\rho_i)$:

$$E(S) = \sum_{i \in S} \varepsilon_i(\rho_i), \qquad (1)$$

where the local atomic energy is predicted as:

$$\varepsilon_i(\rho_i) = \sum_n k(\rho_i, \rho_n)\alpha_n. \qquad (2)$$

In Eq. 2, $k(\rho_i, \rho_n)$ is the kernel (or similarity) function computed between two local atomic environments, the weights α are analytically calculated during the training process, and n is the index that runs from 0 to the number of training data points employed. We employ 2- and 3-body kernels for local atomic environments, which compare local atomic environments ρ_i based on their distances of pairs and triplets of atoms, respectively.[26,27,40] A local atomic environment ρ_i is defined as the collection of relative positions $r_{ij} = r_j - r_i$ of all atoms j contained within a sphere of radius r_{cut} centered on atom i. While traditional GPR FFs are faster to compute than the electronic structure methods they are trained on, they are still orders of magnitude slower than traditional parametrised FFs. The GPR FFs are therefore transformed into tabulated FFs, which retain the accuracy of the original GPR FFs while being extremely fast to compute, on par with other classical FFs. The ability to map the GPR FFs follows from the explicit 2- and 3-body nature of the representations we adopt, and takes place via spline interpolation, following the procedure introduced by Glielmo et al.[26] and first applied to MD simulations in Zeni et al.[27]. The hyper-parameters used to train the ML-FFs are, following the notation employed in Vandermause et al.[28], $\sigma_{s,2} = 0.02$, $l_2 = 0.4$, $\sigma_{s,3} = 7.0$, $l_2 = 8.6$, $\sigma_n = 0.12$, $r_{\text{cut, 2}} = 8.0$ Å, $r_{\text{cut, 3}} = 4.5$ Å.

Hybrid ML-FFs

We generate a third ML-FF, named hybrid, by linearly combining the 2- and 3-body FFs of the ML-FFs derived from LDA and rPBE. This is done through a parameter β that weights the two ML-FFs so that the energy $\varepsilon^{\text{hybrid}}$ for a local atomic environment ρ, is:

$$\varepsilon^{\text{hybrid}}(\rho) = \beta \varepsilon^{\text{LDA}}(\rho) + (1 - \beta)\varepsilon^{\text{rPBE}}(\rho). \qquad (3)$$

The parameter β is tuned to match the experimental cohesive energy of bulk Au (3.81 eV/atom) and is set to 0.61 for our ML-FF. The resulting hybrid ML-FF is a 2+3-body FF, and it has cohesive energy and equilibrium bulk lattice parameter that are intermediate between the LDA and rPBE ML-FFs ones, as can be seen in Supplementary Figure 1. We remark that the generation of such hybrid ML-FF is possible because of the strictly 2+3-body nature of the ML-FFs employed, and because of the similar functional forms the LDA and rPBE ML-FFs display. Furthermore, the hybrid ML-FF can be easily fitted to match the experimental cohesive energy of bulk Au solely because this energy is overestimated (underestimated) by the LDA (rPBE) ML-FF.

DFT calculations setup

We employ training data calculated under the Local Density (LDA) or Generalized Gradient Approximation (GGA - rPBE pseudopotentials) to the exchange-correlation term. We carry out LDA[41] calculations using the Vienna Ab initio Simulation Package[42,43] with projector-augmented wave pseudopotentials.[44,45] The energy cut-off of the plane-wave basis set was 240 eV, and the tolerance for self-consistency for the electronic steps was set at 10^{-6} eV. We calculate GGA rPBE[46] reference energies and forces using CP2K 6.1.[47] All elements are described with the DZVP-MOLOPT basis set[48] with cores represented by the dual-space Goedecker-Teter-Hutter pseudopotentials.[49] The plane-waves cut-off is set to 500 Ry with a relative cut-off of 50 Ry. The self-consistent cycle converges when a change of less than 10^{-6} eV is observed in the estimate of the system's energy.

Molecular Dynamics calculations setup

To study via ML-FFs the melting of Au NPs, we perform several independent MD simulations at fixed volume in periodic boundary conditions (box width = 100Å). We employ LAMMPS[50] as our MD engine, and the FLARE[28] add-on for calculating the energies and forces predicted by the mapped ML-FF. The temperature of the system, controlled using a Langevin thermostat with a 100 fs noise, continuously increases at a rate of 20 K/ns, with starting temperatures that range between 400K and 700K and ending temperatures that range between 1200K and 1500K, depending on the NP's sizes. Newton's equations of motions are integrated via a velocity-Verlet algorithm with a 1 fs time step for systems with less than 1000 atoms, and 2 fs for systems above 1000 atoms in the case of the LDA and rPBE ML-FFs. All simulations employing the hybrid ML-FF are carried on using a 5 fs integration time step.

Local Atomic Environment Descriptor

We employ a local atomic density descriptor to feature each atomic environment in a NP as a function of the relative positions of the other atoms within a cut-off set to 1.75 times the average nearest neighbour distance, and therefore set to 4.24 Å for simulations employing the LDA ML-FF, to 4.42 Å for simulations employing the r-PBE ML-FF, and to 4.30 Å for simulations employing the hybrid ML-FF. A sensitivity analysis shows that the featurisation associated with the representation is marginally affected by the choice of the cut-off radius, as long as the latter is larger than the bulk second nearest neighbours distance (see Supplementary Methods for further detail). We adopt the 2+3-body atomic cluster expansion representation with 4 radial and 4 angular components and employ Bessel functions of the first kind as radial basis functions.[36,37,51]

Clustering Algorithm for Phase Change Characterisation

To apply the clustering algorithm to data generated through the use of a ML-FF, we first gather 10000 randomly chosen local atomic environment representations from among MD simulations of all NP sizes. We then run a hierarchical k-means clustering[52] algorithm to group similar representations, applying two to three iterations of k-means clustering to partition the local atomic environment sampled during the MD simulations into the six classes described previously (additional details can be found in the Supplementary Methods).

Melting Temperature Estimation

We estimate the $T_{\mathrm{melt}}^{\mathrm{NP}}$ as the temperature for which the maximum positive derivative of the fraction of inner atoms labelled as liquid w.r.t. the nominal simulation temperature (or, equivalently, the simulation time) is observed. The $T_{\mathrm{melt}}^{\mathrm{NP}}$ is commonly defined as the temperature where the highest value of the heat capacity is observed $T_{\mathrm{melt}}^{\mathrm{NP}}$, or as the temperature here the highest standard deviation in the total energy is found.[13,53,54] Supplementary Figure 13 shows the striking correspondence that exists between the $T_{\mathrm{melt}}^{\mathrm{NP}}$ estimated using the three aforementioned methods. This result confirms that the $T_{\mathrm{melt}}^{\mathrm{NP}}$ estimation methods we introduce are accurate for the systems we consider and reinforces our belief that the characterization offered by our clustering method is valid.

Surface Transition Temperature Calculation

To calculate $T_{\mathrm{thresh}}^{\mathrm{surf.}}$, we analyse the spatial distribution of LS atoms. We subdivide the NP in spherical shells of width 1 Å centered at the COM of the NP. We define the crust radius, R_{crust}, as the distance from the NP COM of the spherical shell where the highest fraction of LS atoms resides. This generally coincides with the outermost radial shell of atoms in the
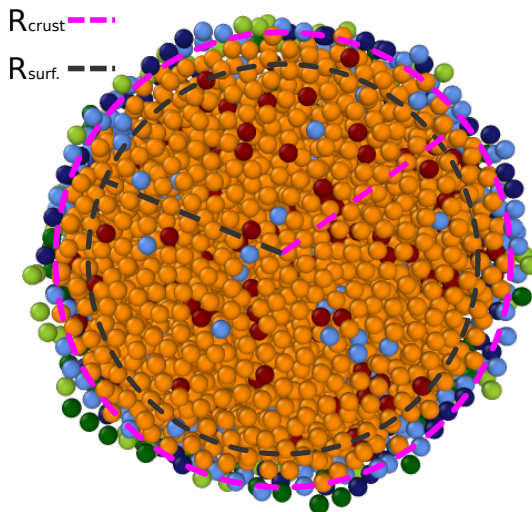

FIG. 5. **Crust and surface radiuses in an Au 6266 NP.** Depiction of the values of R_{crust} (pink) and $R_{\mathrm{surf.}}$ (grey) for an Au 6266 NP. Atoms are colour-coded according to their class as yielded by the clustering algorithm, and mirror the ones of Figures 1 and 3.

NP. We then aim to define a surface shell, and consider a second distance, $R_{\mathrm{surf.}} = R_{\mathrm{crust}} - 3$ Å. The choice of a 3 Å buffer represents an arbitrary but educated guess to incorporate, approximately, a second shell of atoms in our statistics. Finally, we define $T_{\mathrm{thresh}}^{\mathrm{surf.}}$ as the lowest temperature at which the liquid local atomic environments in the surface of the NP amount for the 0.4 of the total number of local atomic environments in the surface shell. To exemplify the protocol, Figure 5 displays the values of $R_{\mathrm{surf.}}$ and R_{crust} for a snapshot extracted from an MD trajectory sampled using the rPBE ML-FF.

Mean First-Passage Temperature

To evaluate the mean first-passage temperature (MFPT), we monitor the label assigned to each atom in the system by the hierarchical clustering scheme at each time step. The MFPT is then defined as the lowest temperature at which at least 0.4 atoms of given initial coordination are labelled as liquid environments. Since MFPTs depend on the $T_{\mathrm{melt}}^{\mathrm{NP}}$, for all MD trajectories we normalize each MFPT by the average $T_{\mathrm{melt}}^{\mathrm{NP}}$ for that particular NP size and ML-FF employed.

Statistical Information

Simulation results are obtained as averages over 4 independent simulations for NPs containing less than 1000 atoms, and over 2 independent simulations for NPs containing more than 1000 atoms. The $T_{\mathrm{melt}}^{\mathrm{NP}}$ reported for each NP size and ML-FF are the average $T_{\mathrm{melt}}^{\mathrm{NP}}$ computed across the 4 (or 2) independent MD simulations. The error bars for $T_{\mathrm{melt}}^{\mathrm{NP}}$ reported for the y-axis of Figure 2 and Supplementary Figures 6 and 13, are calculated as the maximum between 25 K - the temperature window (see also Methods) used to individuate the peak

of the positive derivative of the fraction of liquid atoms w.r.t. simulation temperature - and the standard deviation of T_{melt}^{NP} computed for the 4 (or 2) independent MD simulations for each NP size and ML-FF used to simulate them. The mean absolute errors (MAEs) on energy differences (force components) reported in Supplementary Table 1 are computed on a variable number of observations, determined by the NP size, from 9 (15147) for Au 561 Co to 50 (22050) for Au 147 Co. On average, MAEs on energy differences (force components) are calculated on 33 ± 14 (21417 ± 12531) samples.

DATA AVAILABILITY

The tabulated Au ML-FFs, Au NPs MD trajectories, and ab initio training data for Au NPs generated in this study have been deposited in the Materials Cloud database under accession code https://archive.materialscloud.org/record/2021.131.[55] Example MD trajectories are also stored in the same repository. Source data for Figures 2 and 4 are provided with this paper. Other data are available from the authors upon request.

CODE AVAILABILITY

A majority of the code used in this calculation is open source. ML-FF training and mapping are carried on using FLARE (https://github.com/mir-group/flare). DFT data are gathered using CP2K (https://www.cp2k.org) and VASP (https://www.vasp.at - licence number 5-867). MD Simulations are run via LAMMPS (https://lammps.sandia.gov). The computation of local atomic environment descriptors, and the clustering characterization are carried on using the Raffy Python package[56]. K-means clustering is done in Python via the SciPy library.

REFERENCES

[1] V. Amendola, R. Pilot, M. Frasconi, O. M. Maragò, and M. A. Iatì, Journal of Physics Condensed Matter **29**, 203002 (2017).

[2] L. Jauffred, A. Samadi, H. Klingberg, P. M. Bendix, and L. B. Oddershede, Chemical Reviews **119**, 8087–8130 (2019).

[3] E. C. Dreaden, A. M. Alkilany, X. Huang, C. J. Murphy, and M. A. El-Sayed, Chemical Society Reviews **41**, 2740 (2012).

[4] C. Carnovale, G. Bryant, R. Shukla, and V. Bansal, ACS Omega (2019).

[5] H. Ha, S. Yoon, K. An, and H. Y. Kim, ACS Catalysis **8**, 11491–11501 (2018).

[6] Y. Zhang, J. Zhang, B. Zhang, R. Si, B. Han, F. Hong, Y. Niu, L. Sun, L. Li, B. Qiao, K. Sun, J. Huang, and M. Haruta, Nature Communications **11**, 558 (2020).

[7] J. H. Carter, X. Liu, Q. He, S. Althahban, E. Nowicka, S. J. Freakley, L. Niu, D. J. Morgan, Y. Li, J. W. Niemantsverdriet, S. Golunski, C. J. Kiely, and G. J. Hutchings, Angewandte Chemie - International Edition **56**, 16037 (2017).

[8] T. Mitsudome and K. Kaneda, Green chemistry **15**, 2636 (2013).

[9] J. Zhao, L. Ge, H. Yuan, Y. Liu, Y. Gui, B. Zhang, L. Zhou, and S. Fang, Nanoscale **11**, 11429 (2019).

[10] J. J. Gilvarry, Physical Review **308**, 102 (1956).

[11] J. D. Honeycutt and H. C. Andersen, Journal of Physical Chemistry (1987).

[12] W. Lechner and C. Dellago, Journal of Chemical Physics **129**, 114707 (2008).

[13] L. Delgado-Callico, K. Rossi, R. Pinto-Miles, P. Salzbrenner, and F. Baletto, Nanoscale **13**, 1172 (2021).

[14] Z. Li and D. Truhlar, Journal of The American Chemical Society **130**, 12698–12711 (2008).

[15] T. Fukuya and Y. Shibuta, Computational Materials Science **184**, 109880 (2020).

[16] T. Ellaby, J. Aarons, A. Varambhia, L. Jones, P. Nellist, D. Ozkaya, M. Sarwar, D. Thompsett, and C. K. Skylaris, Journal of Physics Condensed Matter **30**, 155301 (2018).

[17] F. Baletto, R. Ferrando, A. Fortunelli, F. Montalenti, and C. Mottet, Journal of Chemical Physics **116**, 3856 (2002).

[18] K. Rossi, V. Jurásková, R. Wischert, L. Garel, C. Corminbœuf, and M. Ceriotti, Journal of Chemical Theory and Computation **16**, 5139–5149 (2020), 2006.12597.

[19] B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, Proceedings of The National Academy of Sciences of The United States of America **116**, 1110 (2019).

[20] A. M. Cooper, J. Kästner, A. Urban, and N. Artrith, npj Computational Materials **6**, 54 (2020), 2002.04172.

[21] V. L. Deringer, M. A. Caro, and G. Csányi, Advanced Materials **31**, 1902765 (2019).

[22] J. Behler, Angewandte Chemie International Edition **56**, 12828 (2017).

[23] C. Zeni, K. Rossi, A. Glielmo, and F. Baletto, Advances in Physics: X **4**, 1654919 (2019).

[24] R. Lot, F. Pellegrini, Y. Shaidu, and E. Küçükbenli, Computer Physics Communications **256**, 107402 (2020).

[25] Y. Xie, J. Vandermause, L. Sun, A. Cepellotti, and B. Kozinsky, npj Computational Materials **7**, 40 (2021), 2008.11796.

[26] A. Glielmo, C. Zeni, and A. De Vita, Physical Review B **97**, 184307 (2018), 1801.04823.

[27] C. Zeni, K. Rossi, A. Glielmo, Á. Fekete, N. Gaston, F. Baletto, and A. De Vita, Journal of Chemical Physics **148**, 241739 (2018).

[28] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, npj Computational Materials **6**, 1 (2020).

[29] S. Jindal, S. Chiriki, and S. S. Bulusu, Journal of Chemical Physics **146**, 204301 (2017).

[30] S. Jindal and S. S. Bulusu, Journal of Chemical Physics **194101**, 149 (2018).

[31] S. Jindal and S. S. Bulusu, Journal of Chemical Physics **152**, 154302 (2020).

[32] A. Thorn, J. Rojas-Nunez, S. Hajinazar, S. E. Baltazar, and A. N. Kolmogorov, Journal of Physical Chemistry C **123**, 30088–30098 (2019).

[33] T. D. Loeffler, S. Manna, T. K. Patra, H. Chan, B. Narayanan, and S. Sankaranarayanan, ChemCatChem **12**, 4796 (2020).

[34] J. Li, T. Chen, K. Lim, L. Chen, S. A. Khan, J. Xie, and X. Wang, Advanced Intelligent Systems **1**, 1900029 (2018), 1811.02771.

[35] S. Chiriki, S. Jindal, and S. S. Bulusu, Journal of Chemical Physics **146**, 084314 (2017).

[36] C. Zeni, K. Rossi, A. Glielmo, and S. De Gironcoli, Journal of Chemical Physics **154**, 224112 (2021).

[37] R. Drautz, Phys. Rev. B **99**, 014104 (2019).

[38] D. M. Foster, T. Pavloudis, J. Kioseoglou, and R. E. Palmer, Nature Communications **10**, 2583 (2019).

[39] H. Duan, Y. Xue, Z. Cui, Q. Fu, X. Chen, and R. Zhang, Particle & Particle Systems Characterization **35**, 1800156 (2018).

[40] A. Glielmo, P. Sollich, and A. De Vita, Physical Review B **95**, 214302 (2017), 1611.03877.

[41] J. P. Perdew and A. Zunger, Physical Review B **23**, 5048 (1981).

[42] G. Kresse and J. Furthmüller, Physical Review B - Condensed Matter and Materials Physics **54**, 11169 (1996).

[43] G. Kresse and J. Furthmüller, Computational Materials Science **6**, 15 (1996).

[44] P. E. Blöchl, Physical Review B **50**, 17953 (1994).

[45] D. Joubert, Physical Review B - Condensed Matter and Materials Physics **59**, 1758 (1999).

[46] B. Hammer, L. B. Hansen, and J. K. Nørskov, Physical Review B - Condensed Matter and Materials Physics **59**, 7413 (1999).

[47] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, Wiley Interdiscip. Rev. Comput. Mol. Sci. **4**, 15 (2014).

[48] J. VandeVondele and J. Hutter, J. Chem. Phys. **127**, 114105 (2007).

[49] M. Krack, Theor. Chem. Acc. **114**, 145 (2005).

[50] S. Plimpton, Journal of Computational physics **117**, 1 (1995).

[51] R. Drautz, Phys. Rev. B **102**, 024104 (2020).

[52] J. Macqueen, Berkeley Symposium on Mathematical Statistics and Probability **5.1**, 281 (1967).

[53] J. Chen, X. Fan, J. Liu, C. Gu, Y. Shi, D. J. Singh, and W. Zheng, Journal of Physical Chemistry C **124**, 7414 (2020).

[54] C. Dai, P. Saidi, H. Song, Z. Yao, M. Daymond, and J. Hoyt, Acta Materialia **136**, 11 (2017).

[55] C. Zeni, K. Rossi, T. Pavloudis, J. Kioseoglou, S. de Gironcoli, R. E. Palmer, and F. Baletto, (2021), 10.24435/materialscloud:s0-24.

[56] C. Zeni, "Ridge-regression atomistic force fields in python," (2021).

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

F.B, T.P. and R.E.P. originally envisioned the project. C.Z., K.R., and F.B. designed and interpreted the simulations and their post-processing. C.Z. carried on the machine learning force-field fitting and performed the ML-FFs melting simulations. C.Z. implemented the characterisation algorithm. K.R. and T.P. performed the electronic structure calculations. C.Z., K.R., T.P, J.K., S.d.G, R.E.P. and F.B. contributed to the writing of the paper.

COMPETING INTERESTS

The authors declare no competing interests.

Supplementary Information

The supplementary information contain further details on the ML-FF training, validation and volume-energy curves, on the hierarchical clustering procedure we develop, and on the melting and surface phase changes we observe during the MD simulations.