

# Data-driven simulation and characterization of gold nanoparticles melting

Claudio Zeni (✉ [czeni@sissa.it](mailto:czeni@sissa.it))

International School for Advanced Studies <https://orcid.org/0000-0002-6334-2679>

Kevin Rossi

Swiss Federal Institute of Technology Lausanne

Theodore Pavlouidis

Aristotle University of Thessaloniki

Joseph Kioseoglou

Aristotle University of Thessaloniki <https://orcid.org/0000-0002-6933-2674>

Stefano de Gironcoli

International School for Advanced Studies

Richard Palmer

Swansea University <https://orcid.org/0000-0001-8728-8083>

Francesca Baletto

King's College London

---

## Article

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

**Posted Date:** July 13th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-677091/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Nature Communications on October 18th, 2021. See the published version at <https://doi.org/10.1038/s41467-021-26199-7>.

# Data-driven simulation and characterisation of gold nanoparticles melting

Claudio Zeni,<sup>1,2,a)</sup> Kevin Rossi,<sup>3,1</sup> Theodore Pavloudis,<sup>4,5</sup> Joseph Kioseoglou,<sup>5</sup> Stefano de Gironcoli,<sup>2</sup> Richard Palmer,<sup>4</sup> and Francesca Baletto<sup>1</sup>

<sup>1)</sup>Department of Physics, King's College London, London, WC2R 2LS, UK

<sup>2)</sup>International School for Advanced Studies, Via Bonomea, 265, 34136, Trieste, IT

<sup>3)</sup>Laboratory of Nanochemistry, Institute of Chemistry and Chemical Engineering, Ecole Polytechnique Fédérale de Lausanne, Lausanne, CH

<sup>4)</sup>College of Engineering, Swansea University, Bay Campus, Fabian Way, Swansea, SA1 8EB, UK

<sup>5)</sup>Department of Physics, Aristotle University of Thessaloniki, Thessaloniki GR-54124, GR

We develop efficient, accurate, transferable, and interpretable machine learning force fields for Au nanoparticles, based on data gathered from Density Functional Theory calculations. We then use them to investigate the thermodynamic stability of Au nanoparticles of different sizes (1 to 6 nm), containing up to 6266 atoms, with respect to a solid-liquid phase change through molecular dynamics simulations. We predict nanoparticle melting temperatures in good agreement to available experimental data. Furthermore, we characterize in detail the solid to liquid phase change mechanism employing an unsupervised learning scheme to categorize local atomic environments. We thus provide a rigorous and 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

Gold nanoparticles (Au-NPs) find widespread application in many technological areas, such as in optics,<sup>1,2</sup> nanomedicine,<sup>3,4</sup> and catalysis<sup>5–9</sup>. 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.<sup>10–13</sup> 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,<sup>14,15</sup> 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 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<sup>16</sup> 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<sup>17</sup>.

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 learning force fields (ML-FFs)<sup>18–25</sup> using the innovative framework of mapped Gaussian processes<sup>26–28</sup>. 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 temper-

<sup>a)</sup>Electronic mail: czeni@sissa.it

ature 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 NP's core second.

## I. RESULTS

### A. Machine Learning Force Fields

To construct a training database, we extract random de-correlated frames from *ab initio* molecular dynamics trajectories where an Au NP containing 309 atoms ( $\sim 2$  nm of diameter) with an initial face centred cubic (FCC) morphology undergoes melting. More information is available in Section I of the Supplementary Materials (SM). 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<sup>26,27</sup> to fit two ML-FFs, one for each DFT method (see Method section for further details). When training on 2100 local atomic environments, our ML-FFs incur in a mean absolute error (MAE) on the force components of  $0.09 \pm 0.04$  eV/Å (LDA ML-FF) and  $0.07 \pm 0.03$  eV/Å (rPBE ML-FF), and in a MAE on the atomic energy of  $2.65 \pm 2.02$  meV/atom (LDA ML-FF) and  $1.98 \pm 1.76$  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<sup>29–35</sup>, and is deemed satisfactory. This training dataset, albeit small, contains a heterogeneous set of local atomic environments, as shown in Figure S1; 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 Suppl. Mat., Section II). Table S1 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 (Figure S2), 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 form-

ing the three ML-FFs present some noticeable differences; in Figure S3 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.

### B. 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 Figure S4). For each NP size and ML-FF, we simulate Au-NPs for a grand 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 Method 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<sup>36</sup> of the 3-body local atomic cluster expansion descriptor<sup>37</sup> 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 Method, Section E and Suppl. Mat. Figure S5). 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 the lower panels of Figure 1, and in Figures S6-S7, both the number of nearest neighbours (#NN) within a pre-defined 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 Figures 1 and S5-S6, can be characterised 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

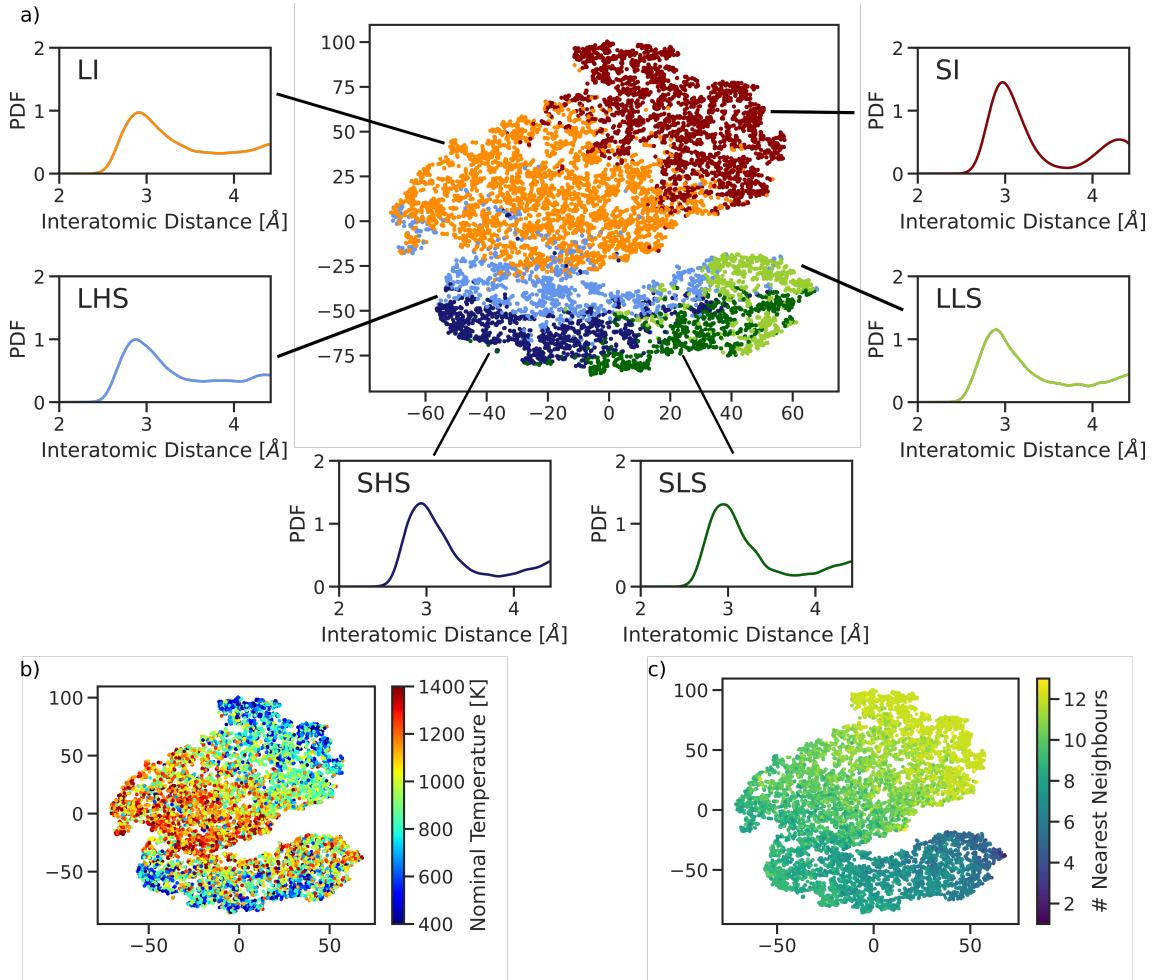


FIG. 1. Visualization of the hierarchical k-means clustering results for MD simulations of Au nanoparticles with 147, 309, 561, 923, 2869 and 6266 atoms, carried out using the ML-FF trained on rPBE-DFT data. Top panel (a): 1st and 2nd component (x- and y-axis) of the t-SNE projection of the atomic expansion 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. Bottom panels (b, 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.

presents the first peak at distances lower than the one for bulk lattice, and lacks a pronounced second peak in correspondence 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

198 the bulk nearest-neighbours distance; furthermore, the  
 199 PDF does not display any peak for the second NNs.  
 200

201 This unsupervised approach enables an original and data-  
 202 driven definition of liquid atomic arrangements in the inner  
 203 part of the NPs and at the surface. Local atomic environ-  
 204 ments in a liquid phase are all characterized by the absence  
 205 of a peak of their PDF in correspondence to their bulk lattice  
 206 distance (i.e., the one predicted by the reference interatomic  
 207 potential). This observation holds regardless of whether they  
 208 lie in the inner part or at the surface of the NP, and of their  
 209 coordination. Furthermore, this result confirms and ratio-  
 210 nalises the universal signature of melting for the whole NP  
 211 we recently proposed<sup>13</sup>.

212 Having discriminated in automated fashion atoms in liq-  
 213 uid and solid environments, also as a function of their spatial  
 214 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 ( $T_{melt}^{NP}$ ) 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 (Figures S8-S9). This observation further corroborates our trust in the clustering algorithm as a tool to characterise Au nanoparticles melting.

### C. Size-dependent melting

Figure 2 reports the  $T_{melt}^{NP}$  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  $T_{melt}^{NP}$  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

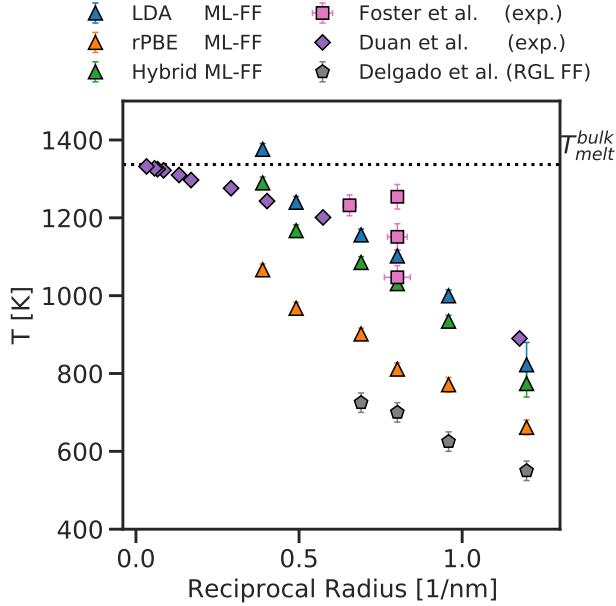


FIG. 2. Average  $T_{melt}^{NP}$  as a function of NP's reciprocal radius computed for MD simulations employing the LDA-trained (blue triangles), rPBE-trained (orange triangles), and hybrid (green triangles) ML-FFs. Experimental data for size-selected Au NPs supported on carbon (pink squares) and spherical Au NPs (purple diamonds), is taken from Foster *et al.*<sup>38</sup> and Duan *et al.*<sup>39</sup>, respectively. Grey pentagons refer to the  $T_{melt}^{NP}$  estimates from Delgado-Callico *et al.*<sup>13</sup>.

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<sup>38,39</sup>. For reference, we add the  $T_{melt}^{NP}$  estimates obtained using a classical MD where the interatomic interaction is derived in the second-moment approximation of the tight-binding (TB-SMA)<sup>13</sup>. All the ML-FFs lead to  $T_{melt}^{NP}$  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  $T_{melt}^{NP}$  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  $T_{melt}^{NP}$  predicted by the hybrid ML-FF are less than 50 K away from the melting temperatures found experimentally via differential scanning calorimetry measurements.<sup>39</sup>

### D. Melting mechanism characterisation

After having established a quantitative agreement between the ML-FFs' predictions and the experimental melting temperatures of Au NPs, also as a function of their size, it is natural to proceed further and analyse the mechanism by which phase changes occur. To this end, Figure 3 displays 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, averaged over independent MD melting simulations employing the rPBE-based ML-FF, where # indicates the number of atoms belonging to a certain class, #LS=#LHS+#LLS, and #tot=#LHS+#LLS+#LI+#SHS+#SLS+#SI. We refer the interested Reader to Figures S10-S11 for the same plots for all systems with 147, 309, 561, 923, 2869 and 6266 atoms and using the three ML-FFs. In Figures 3, S10 and S11, 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  $T_{melt}^{NP}$  independently of their distance from the NP centre of mass (COM). Areas located few Å below the NP 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.*<sup>38</sup>, 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 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.<sup>38</sup>

To better compare our results with available experimental data, we would therefore like to numerically define a  $T_{melt}^{surf.}$  similarly to how we defined the  $T_{melt}^{NP}$ . This is however not advisable, as while the #LI atoms has a clear and distinct positive jump which allows us to identify the  $T_{melt}^{NP}$  (See Figure S8), the #LS does not show such a distinct transition (See Figure S13). The relative amount of LS atoms does instead

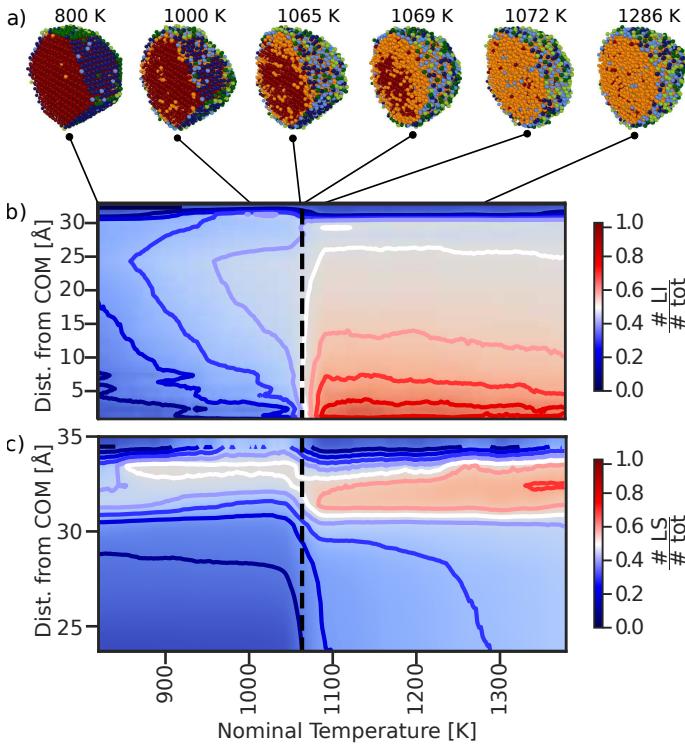
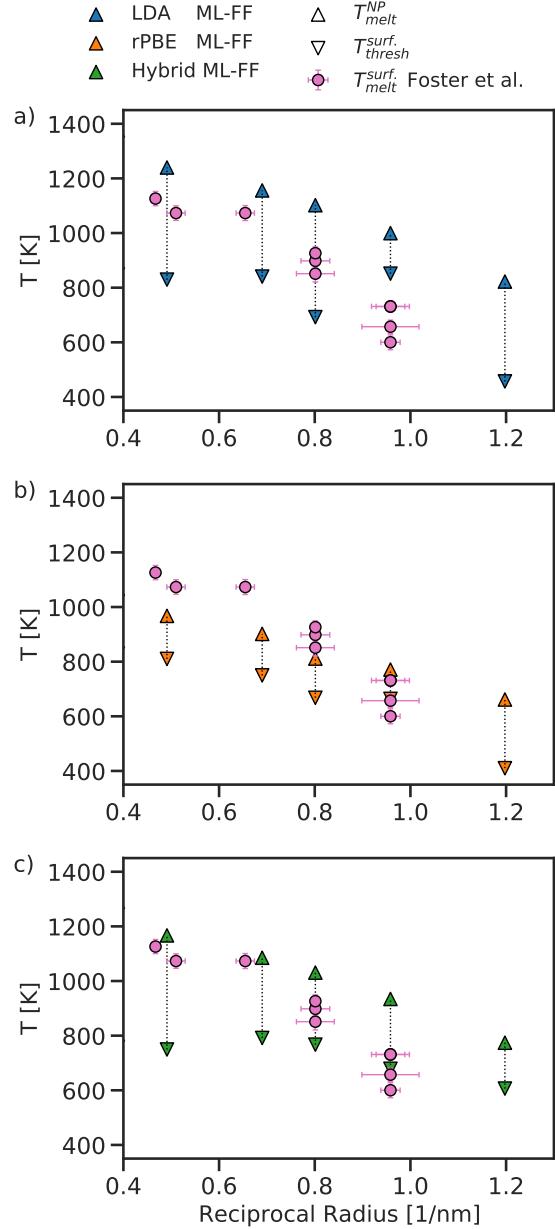


FIG. 3. a) Snapshots of Au 6266 simulated using the r-PBE ML-FF at different nominal simulation temperatures, with atoms coloured according to the clustering algorithm, and using the same colour scheme as in Figure 1. b), c) Average fraction of #LI (b) and #LS (c) local atomic environments as a function of the radial distance from the COM (y coordinate), and of the nominal system temperature (x coordinate). The bold colored lines in 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.

increase gradually with temperature for all NP sizes and all ML-FFs, and usually reaches values around 0.5 (white line in panel (c) of Fig. 3) for atoms in the surface layer at temperatures approaching  $T_{melt}^{NP}$ . Therefore, we must provide an alternative definition for the temperature at which surface phase change can be observed. To do so, we first define the surface of each NP as the region outside a sphere of radius  $R_S$  centered at the NP's COM, where  $R_S$  is calculated as a function of the spatial distribution of LS environments (see the Methods section and Figure 5 for additional details). Afterward, we compute the lowest temperatures where at least 0.4 of the local atomic environments in the surface of the NP are classified as liquid ( $T_{thresh}^{surf.}$ ). Figure 4 reports the values of  $T_{thresh}^{surf.}$  and the  $T_{melt}^{NP}$  for all our MD simulations, and the experimental  $T_{melt}^{surf.}$  as reported in Foster et al.<sup>38</sup>. The temperature ranges comprised between  $T_{thresh}^{surf.}$  and  $T_{melt}^{NP}$  are in line with the experimentally reported  $T_{melt}^{surf.}$ , 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



308  
309  
310  
311  
312  
313  
314  
315  
316 FIG. 4.  $T_{thresh}^{surf.}$  (downward triangles) and  $T_{melt}^{NP}$  (upward triangles)  
317 as a function of NPs' reciprocal radius, for MD simulations carried  
318 out using the LDA ML-FF (panel a), rPBE ML-FF (panel b), and  
319 hybrid ML-FF (panel c). Experimental estimates of  $T_{melt}^{surf.}$  from  
320 high-resolution TEM measurements are taken from Foster et al.<sup>38</sup>,  
321 and are shown as pink circles.

322  
323  
324 edge, on a (100) surface, on a (111) surface, or in a bulk  
325 FCC environment. These families of atoms are discrimi-  
326 nated according to their number of first NNs at the begin-  
327 ning of the MD simulations. Edge atoms (#NN=6) are more  
328 likely to move into a liquid phase than atoms on a (100) facet  
329 (#NN=8), which in turn are more prone to end into a liquid

330  
331  
332  
333  
334  
335

phase than the atoms on a (111) facet (#NN=9). The inner atoms (#NN=12) present the overall largest MFPT, this is coherent with the melting initiating from the outer layers of the NP (additional details are available in the Methods, and in Section XI of the Suppl. Mat.).

The trends observed during the melting characterization indicate that local phase changes in the outermost layer of a NP start to occur at temperatures few hundred of K below the  $T_{melt}^{NP}$ . 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.

## II. DISCUSSION AND CONCLUSION

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 parametrized 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 and original 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 our novel 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 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 at the nanoscale.

## III. METHODS

### A. Database Construction

To construct the training set, we extract 7 random de-correlated frames 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 to each configuration are calculated within the density functional theory framework. We generate a validation set by extracting de-correlated frames from MD trajectories previously reported in Delgado-Callico et al.<sup>13</sup>, and from *ab initio* MD trajectories previously reported in Foster et al.<sup>38</sup>. We sample the melting MD trajectories reported in Delgado-Callico et al.<sup>13</sup>, 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 containing 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<sup>13</sup>). The melting MD trajectories reported in Foster et al.<sup>38</sup> 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.

### B. 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)<sup>26,27</sup> using the FLARE Python Package.<sup>28</sup> 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  $\varepsilon_i(\rho_i)$ :

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

where the local atomic energy is predicted as:

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

In Eq. 2,  $k(\rho_i, \rho_n)$  is the kernel (or similarity) function computed between two local atomic environments, and the weights  $\alpha$  are analytically calculated during the training process. We employ 2- and 3-body kernels for local atomic environments, which compare local atomic environments  $\rho_i$  based on their distances of pairs and triplets of atoms, respectively.<sup>26,27,40</sup> A local atomic environment  $\rho_i$  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.<sup>26</sup> and first applied to MD simulations in Zeni et al.<sup>27</sup>. The hyper-parameters used to train the ML-FFs are, following the notation employed in Vandermause et al.<sup>28</sup>,  $\sigma_{s,2} = 0.02$ ,  $l_2 = 0.4$ ,  $\sigma_{s,3} = 7.0$ ,  $l_2 = 8.6$ ,  $\sigma_n = 0.12$ ,  $r_{cut,2} = 8.0 \text{ \AA}$ ,  $r_{cut,3} = 4.5 \text{ \AA}$ .

### 1. 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  $\beta$  that weights the two ML-FFs so that the energy  $\epsilon^{hybrid}$  for a local atomic environment  $\rho$ , is:

$$\epsilon^{hybrid}(\rho) = \beta \epsilon^{LDA}(\rho) + (1 - \beta) \epsilon^{rPBE}(\rho). \quad (3)$$

The parameter  $\beta$  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 Fig. S1. 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.

### C. 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<sup>41</sup> calculations using the Vienna Ab initio Simulation Package<sup>42,43</sup> with projector-augmented wave pseudopotentials.<sup>44,45</sup> 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<sup>46</sup> reference energies and forces using CP2K 6.1.<sup>47</sup> All elements are described with the DZVP-MOLOPT basis set<sup>48</sup> with cores represented by the dual-space Goedecker-Teter-Hutter pseudopotentials.<sup>49</sup> 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.

### D. 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 = 100 Å). We employ LAMMPS<sup>50</sup> as our MD engine, and the FLARE<sup>28</sup> 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.

### E. Phase Change Characterisation

#### 1. 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. 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.<sup>36,37,51</sup>

#### 2. Clustering Algorithm

To construct the clustering algorithm for a given 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<sup>52</sup> 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 in Section IB (additional details can be found in Section X of the Suppl. Mat.).

#### 3. Melting Temperature Estimation

We estimate the  $T_{melt}^{NP}$  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  $T_{melt}^{NP}$  is commonly defined as the temperature where the highest value of the heat capacity is observed  $T_{melt}^{NP}$ , or as the temperature here the highest standard deviation in the total energy

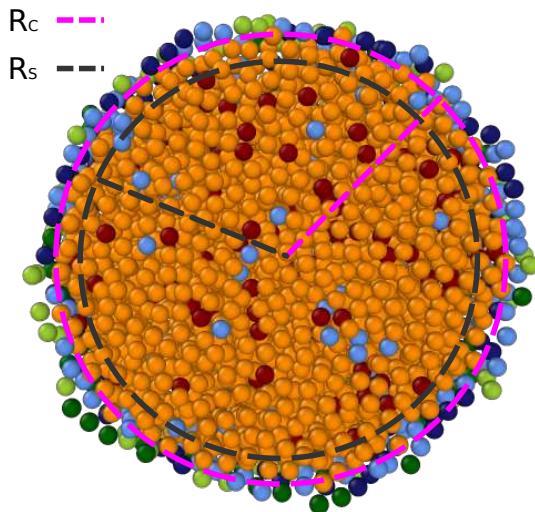


FIG. 5. Depiction of the values of  $R_C$  (pink) and  $R_S$  (grey) for an Au-6266 NP. Atoms are colour-coded according to their class as yielded by the clustering algorithm, and mirror the ones of Figures 1 and 3.

is found<sup>13,53,54</sup>. In Section 7 of the Suppl. Mat., we show the striking correspondence that exists between the  $T_{melt}^{NP}$  estimated using the three aforementioned methods. This result confirms that the  $T_{melt}^{NP}$  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.

#### 4. Surface Distance Calculation

The calculation of  $T_{thresh}^{surf}$ , shown in Figure 4 requires the estimation of a distance  $R_S$  that separates the inner from the surface regions of the NPs. We first define  $R_C$  as the “crust radius”, i.e. the radius of the spherical shell of width 1 Å centered on the COM for which the highest fraction of LS atoms is observed, which typically coincides with the average distance of the outermost layer of atoms from the COM. We then increase the width of the surface region by adding a 3 Å buffer zone, obtaining  $R_S = R_C - 3$  Å. The choice of a 3 Å buffer is an educated guess to incorporate, approximately, a second shell of atoms under the liquid “crust”. Figure 5 displays the values of  $R_S$  and  $R_C$  for a snapshot extracted from an MD trajectory sampled using the rPBE ML-FF.

#### 5. 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 a given initial coordination are labeled as “liquid” environments. Since MFPTs depend on the  $T_{melt}^{NP}$ ,

for all MD trajectories we normalize each MFPT by the average  $T_{melt}^{NP}$  for that particular NP size and ML-FF employed.

#### F. 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  $T_{melt}^{NP}$  reported for each NP size and ML-FF are the average  $T_{melt}^{NP}$  computed across the 4 (or 2) independent MD simulations. The error bars for  $T_{melt}^{NP}$  reported for the y-axis of Figures 2, S4, and S9, are calculated as the maximum between 25 K - the temperature window (see also Method Section, subsection E3) 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 energies (force components) reported in Table S1 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 energies (force components) are calculated on  $33 \pm 14$  ( $21417 \pm 12531$ ) samples.

#### IV. CODE AND DATA 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). 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 (<https://github.com/ClaudioZeni/Raffy>). K-means clustering is done in Python via the SciPy library. Tabulated Au ML-FFs trained on LDA-DFT, rPBE-DFT, and the hybrid ML-FF are deposited in the Materials Cloud (10.24435/materialscloud:s0-24). Example MD trajectories are also stored in the same repository. Other data are available from the authors upon request.

<sup>1</sup>V. Amendola, R. Pilot, M. Frasconi, O. M. Maragò, and M. A. Iati, Journal of Physics Condensed Matter **29**, 203002 (2017).

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

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

<sup>4</sup>C. Carnovale, G. Bryant, R. Shukla, and V. Bansal, ACS Omega (2019).

<sup>5</sup>H. Ha, S. Yoon, K. An, and H. Y. Kim, ACS Catalysis **8**, 11491–11501 (2018).

<sup>6</sup>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).

<sup>7</sup>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).

- <sup>8</sup>T. Mitsudome and K. Kaneda, Green chemistry **15**, 2636 (2013).
- <sup>9</sup>J. Zhao, L. Ge, H. Yuan, Y. Liu, Y. Gui, B. Zhang, L. Zhou, and S. Fang, Nanoscale **11**, 11429 (2019).
- <sup>10</sup>J. J. Gilvarry, Physical Review **308**, 102 (1956).
- <sup>11</sup>J. D. Honeycutt and H. C. Andersen, Journal of Physical Chemistry (1987).
- <sup>12</sup>W. Lechner and C. Dellago, Journal of Chemical Physics **129**, 114707 (2008).
- <sup>13</sup>L. Delgado-Callico, K. Rossi, R. Pinto-Miles, P. Salzbrenner, and F. Baletto, Nanoscale **13**, 1172 (2021).
- <sup>14</sup>Z. Li and D. Truhlar, Journal of The American Chemical Society **130**, 12698–12711 (2008).
- <sup>15</sup>T. Fukuya and Y. Shibuta, Computational Materials Science **184**, 109880 (2020).
- <sup>16</sup>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).
- <sup>17</sup>F. Baletto, R. Ferrando, A. Fortunelli, F. Montalenti, and C. Mottet, Journal of Chemical Physics **116**, 3856 (2002).
- <sup>18</sup>K. Rossi, V. Jurášková, R. Wischert, L. Garel, C. Corminboeuf, and M. Ceriotti, Journal of Chemical Theory and Computation **16**, 5139–5149 (2020), 2006.12597.
- <sup>19</sup>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).
- <sup>20</sup>A. M. Cooper, J. Kästner, A. Urban, and N. Artrith, npj Computational Materials **6**, 54 (2020), 2002.04172.
- <sup>21</sup>V. L. Deringer, M. A. Caro, and G. Csányi, Advanced Materials **31**, 1902765 (2019).
- <sup>22</sup>J. Behler, Angewandte Chemie International Edition **56**, 12828 (2017).
- <sup>23</sup>C. Zeni, K. Rossi, A. Glielmo, and F. Baletto, Advances in Physics: X **4**, 1654919 (2019).
- <sup>24</sup>R. Lot, F. Pellegrini, Y. Shaidu, and E. Küçükbenli, Computer Physics Communications **256**, 107402 (2020).
- <sup>25</sup>Y. Xie, J. Vandermause, L. Sun, A. Cepellotti, and B. Kozinsky, npj Computational Materials **7**, 40 (2021), 2008.11796.
- <sup>26</sup>A. Glielmo, C. Zeni, and A. De Vita, Physical Review B **97**, 184307 (2018), 1801.04823.
- <sup>27</sup>C. Zeni, K. Rossi, A. Glielmo, Á. Fekete, N. Gaston, F. Baletto, and A. De Vita, Journal of Chemical Physics **148**, 241739 (2018).
- <sup>28</sup>J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, npj Computational Materials **6**, 1 (2020).
- <sup>29</sup>S. Jindal, S. Chiriki, and S. S. Bulusu, Journal of Chemical Physics **146**, 204301 (2017).
- <sup>30</sup>S. Jindal and S. S. Bulusu, Journal of Chemical Physics **194**, 101, 149 (2018).
- <sup>31</sup>S. Jindal and S. S. Bulusu, Journal of Chemical Physics **152**, 154302 (2020).
- <sup>32</sup>A. Thorn, J. Rojas-Nunez, S. Hajinazar, S. E. Baltazar, and A. N. Kolmogorov, Journal of Physical Chemistry C **123**, 30088–30098 (2019).
- <sup>33</sup>T. D. Loeffler, S. Manna, T. K. Patra, H. Chan, B. Narayanan, and S. Sankaranarayanan, ChemCatChem **12**, 4796 (2020).
- <sup>34</sup>J. Li, T. Chen, K. Lim, L. Chen, S. A. Khan, J. Xie, and X. Wang, Advanced Intelligent Systems **1**, 1900029 (2018), 1811.02771.
- <sup>35</sup>S. Chiriki, S. Jindal, and S. S. Bulusu, Journal of Chemical Physics **146**, 084314 (2017).
- <sup>36</sup>C. Zeni, K. Rossi, A. Glielmo, and S. De Gironcoli, Journal of Chemical Physics **154**, 224112 (2021).
- <sup>37</sup>R. Drautz, Phys. Rev. B **99**, 014104 (2019).
- <sup>38</sup>D. M. Foster, T. Pavloudis, J. Kioseoglou, and R. E. Palmer, Nature Communications **10**, 2583 (2019).
- <sup>39</sup>H. Duan, Y. Xue, Z. Cui, Q. Fu, X. Chen, and R. Zhang, Particle & Particle Systems Characterization **35**, 1800156 (2018).
- <sup>40</sup>A. Glielmo, P. Sollich, and A. De Vita, Physical Review B **95**, 214302 (2017), 1611.03877.
- <sup>41</sup>J. P. Perdew and A. Zunger, Physical Review B **23**, 5048 (1981).
- <sup>42</sup>G. Kresse and J. Furthmüller, Physical Review B - Condensed Matter and Materials Physics **54**, 11169 (1996).
- <sup>43</sup>G. Kresse and J. Furthmüller, Computational Materials Science **6**, 15 (1996).
- <sup>613</sup><sup>44</sup>P. E. Blöchl, Physical Review B **50**, 17953 (1994).
- <sup>614</sup><sup>45</sup>D. Joubert, Physical Review B - Condensed Matter and Materials Physics **59**, 1758 (1999).
- <sup>615</sup><sup>46</sup>B. Hammer, L. B. Hansen, and J. K. Nørskov, Physical Review B - Condensed Matter and Materials Physics **59**, 7413 (1999).
- <sup>616</sup><sup>47</sup>J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, Wiley Interdiscip. Rev. Comput. Mol. Sci. **4**, 15 (2014).
- <sup>617</sup><sup>48</sup>J. VandeVondele and J. Hutter, J. Chem. Phys. **127**, 114105 (2007).
- <sup>618</sup><sup>49</sup>M. Krack, Theor. Chem. Acc. **114**, 145 (2005).
- <sup>619</sup><sup>50</sup>S. Plimpton, Journal of Computational physics **117**, 1 (1995).
- <sup>620</sup><sup>51</sup>R. Drautz, Phys. Rev. B **102**, 024104 (2020).
- <sup>621</sup><sup>52</sup>J. Macqueen, Berkeley Symposium on Mathematical Statistics and Probability **5**, 1, 281 (1967).
- <sup>622</sup><sup>53</sup>J. Chen, X. Fan, J. Liu, C. Gu, Y. Shi, D. J. Singh, and W. Zheng, Journal of Physical Chemistry C **124**, 7414 (2020).
- <sup>623</sup><sup>54</sup>C. Dai, P. Saidi, H. Song, Z. Yao, M. Daymond, and J. Hoyt, Acta Materialia **136**, 11 (2017).
- <sup>624</sup>  
<sup>625</sup>  
<sup>626</sup>  
<sup>627</sup>  
<sup>628</sup>  
<sup>629</sup>
- V. ACKNOWLEDGMENTS**
- <sup>630</sup>  
<sup>631</sup>  
<sup>632</sup>  
<sup>633</sup>  
<sup>634</sup>  
<sup>635</sup>  
<sup>636</sup>  
<sup>637</sup>  
<sup>638</sup>  
<sup>639</sup>  
<sup>640</sup>  
<sup>641</sup>  
<sup>642</sup>  
<sup>643</sup>  
<sup>644</sup>  
<sup>645</sup>  
<sup>646</sup>  
<sup>647</sup>  
<sup>648</sup>  
<sup>649</sup>  
<sup>650</sup>  
<sup>651</sup>  
<sup>652</sup>  
<sup>653</sup>  
<sup>654</sup>  
<sup>655</sup>  
<sup>656</sup>  
<sup>657</sup>  
<sup>658</sup>  
<sup>659</sup>  
<sup>660</sup>  
<sup>661</sup>  
<sup>662</sup>  
<sup>663</sup>  
<sup>664</sup>  
<sup>665</sup>  
<sup>666</sup>  
<sup>667</sup>  
<sup>668</sup>  
<sup>669</sup>  
<sup>670</sup>  
<sup>671</sup>  
<sup>672</sup>  
<sup>673</sup>  
<sup>674</sup>  
<sup>675</sup>  
<sup>676</sup>  
<sup>677</sup>  
<sup>678</sup>  
<sup>679</sup>  
<sup>680</sup>  
<sup>681</sup>  
<sup>682</sup>  
<sup>683</sup>  
<sup>684</sup>  
<sup>685</sup>  
<sup>686</sup>  
<sup>687</sup>  
<sup>688</sup>  
<sup>689</sup>  
<sup>690</sup>  
<sup>691</sup>  
<sup>692</sup>  
<sup>693</sup>  
<sup>694</sup>  
<sup>695</sup>  
<sup>696</sup>  
<sup>697</sup>  
<sup>698</sup>  
<sup>699</sup>
- <sup>700</sup>  
<sup>701</sup>  
<sup>702</sup>  
<sup>703</sup>  
<sup>704</sup>  
<sup>705</sup>  
<sup>706</sup>  
<sup>707</sup>  
<sup>708</sup>  
<sup>709</sup>  
<sup>710</sup>  
<sup>711</sup>  
<sup>712</sup>  
<sup>713</sup>  
<sup>714</sup>  
<sup>715</sup>  
<sup>716</sup>  
<sup>717</sup>  
<sup>718</sup>  
<sup>719</sup>  
<sup>720</sup>
- CZ acknowledge funding by the Engineering and Physical Sciences Research Council (EPSRC) through the Centre for Doctoral Training Cross-Disciplinary Approaches to Non-Equilibrium Systems (CANES, Grant No. EP/L015854/1) and by the European Union's Horizon 2020 research and innovation program (Grant No. 824143, MaX 'MAterials design at the eXascale' Centre of Excellence). KR has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, (Grant agreement No. 890414)" We are grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/P020194/1) and to the Supercomputing Wales project, which is partially funded by the European Regional Development Fund (ERDF) via the Welsh Government. SdG acknowledges funding from European Union's Horizon 2020 research and innovation program (Grant No. 824143, MaX 'MAterials design at the eXascale' Centre of Excellence). FB acknowledges the financial support offered by the Royal Society under project number RG120207.**
- VI. AUTHOR CONTRIBUTIONS**
- <sup>721</sup>  
<sup>722</sup>  
<sup>723</sup>  
<sup>724</sup>  
<sup>725</sup>  
<sup>726</sup>  
<sup>727</sup>  
<sup>728</sup>
- <sup>729</sup>  
<sup>730</sup>
- R.P., T.P. and F.B. originally envisioned the project. C.Z., K.R., and F.B. designed and interpreted the simulations and their post-processing. C.Z. developed the machine learning force-field fitting and performed the ML-FFs melting simulations. C.Z. implemented the novel characterisation algorithm. K.R. and T.P. performed the electronic structure calculations. All authors contributed to the writing of the paper.**
- VII. COMPETING INTERESTS**
- <sup>730</sup>  
<sup>731</sup>
- The authors declare no competing interests.

731 **VIII. ADDITIONAL INFORMATION**

736 Phase changes.

732 **A. Supplementary Materials**

733 Is available for this paper at .... The latter contains fur-  
734 ther detail on ML-FF training and validation, The Clustering  
735 of local atomic environments, the inner Melting and Surface

737 **B. Correspondence**

738 Should be addressed to C.Z.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.pdf](#)