

Cite this: DOI: 00.0000/xxxxxxxxxx

Structural characterisation of nanoalloys for (photo)catalytic applications with the Sapphire library

Robert M. Jones,^a Kevin Rossi ^b, Claudio Zeni,^c Mirko Vanzan,^d Igor Vasiljevic ^e, Alejandro Santana-Bonilla ^a, and Francesca Baletto ^{a,e}

Received Date

Accepted Date

DOI: 00.0000/xxxxxxxxxx

A non-trivial interplay rules the relationship between the structure and the chemophysical properties of a nanoparticle. In this context, characterization experiments, molecular dynamics simulations and electronic structure calculations may allow to pinpoint the variables that determine a given property. Conversely, a rigorous computational characterization of the geometry and chemical ordering of metallic nanoparticles and nanoalloys enables to discriminate which descriptors could be linked with their stability and performance. We introduce a modular and open-source library, **Sapphire**, which classify the structural characteristics of a given nanoparticle through several structural analysis and order parameters. A special focus is geared towards geometrical descriptors of catalytic activity.

1 Introduction

Mono-, bi-, and poly-metallic nanoparticles (MNPs), or nanoalloys (NAs), find a wealth of potential uses across disciplines ranging from sensing^{1,2} to drug-delivery,³ memory-storage^{4,5} to optics^{6,7}. Among the most prominent applications, MNPs further play a significant role as thermal,^{8–10} electro-chemical,^{11,12} and photo-chemical^{13,14} catalysts.

The need of a detailed characterization of the MNP's morphology stems from the delicate interplay among size, geometrical features, chemical composition and ordering, and the chemophysical (e.g., optical, catalytic, etc., etc.) properties of the MNP itself^{1,15–17}. Focusing on catalytic applications, the role of the MNP's and NA's surface is central. In this context, electronic structure calculations represent an established route to infer the structure-property relationships which rule the activity, selectivity, and stability of the catalyst itself^{18–21}. A knowledge of robust structure-property relationships, and of the finite temperature probability of observing MNPs and NAs with a given architecture, in turn, allows to draw design rules, to predict the activity and, to forecast the ageing of a nanocatalysts. One challenge in such process is the lack of an automated and agnostic characterization of the MNP's and NA's morphology, with a special focus on its surface. Developing a comprehensive and robust set of taxonomy rules to classify and char-

acterise MNPs and NAs could similarly be beneficial across their large number of multi-disciplinary applications. The objective of this communication is to offer the community an open-source and user-friendly package for the structural and chemical analysis of MNPs and NAs, released as **Sapphire**. Whilst there exists software and libraries to characterise structural motifs in molecular and periodic systems^{22–24}, we believe that **Sapphire** fills a gap in the computational chemical physics community by providing an open-source, modular, and documented library dedicated to the computational characterisation of metallic nanoparticles and nanoalloys. **Sapphire** is a Python library publicly available at <https://github.com/kcl-tscm/Sapphire>²⁵, which provides a unified framework to calculate various structural descriptors, from the pair-distance distribution function to the identification of surface atoms. **Sapphire** extends the common neighbour analysis with the introduction of atomic patterns, allowing a classification into the morphological families, e.g. icosahedra (Ih), decahedra (Dh) and FCC (both with or without stacking fault), as well as amorphous/molten-like motifs. **Sapphire** comes with a robust statistics through the use of ensemble averages and well-known statistical quantities, such as the Kullback-Liebler divergence, and statistical tests, like the Kolmogorov-Smirnov one. We recommend the use of **Sapphire** only for metallic nanoparticles and nanoalloys. Although some algorithms are independent of the chemical species, the application of this software across the periodic table is not tested yet.

^a Physics Department, King's College London, WC2R 2LS, UK

^b Swiss Federal Institute of Technology Lausanne, Rte Cantonale, 1015 Lausanne, Switzerland

^c International School for Advanced Studies, Via Bonomea, 265, 34136 Trieste TS, Italy

^d University of Padua, Via VIII Febbraio, 2, 35122 Padova PD, Italy

^e Physics Department, Università di Milano "La Statale", I-20133, Italy

The next section of the manuscript is dedicated to discuss the philosophy and workflow of **Sapphire**, focusing primarily on the nature of its hierarchical design, and how we implement the library to be flexible and facile in its use. We provide a retrospec-

tive overview of **Sapphire** capabilities. Each section presents a didactic introduction to the theory and practical use of each of the characterisation tools. Among them, it is worthy to note that **Sapphire** provides a Metadata section, as starting to be requested by FAIR (Findable, Accessible, Interoperable and Re-usable)²⁶. We thus hope **Sapphire** will help this community to easily create FAIR databases in their future contributions. As paradigmatic examples, we apply **Sapphire** to individual snapshots as well as molecular dynamics trajectories of Au, Pt, AuPt, CuPt, and AuRh nanoalloys.

2 Workflow

The aim of this section is to briefly discuss **Sapphire**'s architecture. Details on how to use the software are available within the tutorial at `/main/Sapphire/Tutorials/`. These web resources will allow the prospective user to reproduce the figures in this manuscript. The interested reader is referred to the Github main page, and the tutorial folder in particular, to obtain a complete overview of how **Sapphire** practically exploits each of its modules.

Fig. 1 provides the structure of **Sapphire** and shows the nature of the decentralised analysis tools. **Sapphire** requires as input only the atomic coordinates for the MNP. Given that **Sapphire** leverages the input / output (IO) stream of ASE,²⁷ many common formats (e.g., `.xyz`, `.exyxyz`, etc...) are compatible. Note, the coordinate file can be obtained from experimental reconstructions – e.g., from a tomography experiment – as well as from numerical simulations. As for the nature of the analysis, one may provide a single snapshot, an entire trajectory, or a set of independent trajectories, i.e. as usually occurs using Molecular Dynamics tools.

From here an interconnected nest of sub-modules, organised by common theme, may be accessed. It is worth noting that because the `Process` class does not directly interact with the IO stream, one may simply call a sub-module to begin the desired analysis.

After gathering the set of configurations to be analysed, **Sapphire** creates an *extended xyz* file, which encodes the user selected single-atom labels, e.g., its radial distance from the centre of mass, coordination number, and functions thereof. Global quantities are both stored in separate files plain text files, to be plotted immediately via Gnuplot or the user's plotting tool of choice; and in the dedicated Metadata object. We dedicate a section for the main quantities. A crucial feature of **Sapphire** is its data storage philosophy. Each calculation using **Sapphire**'s `Process` creates a Metadata in agreement with EMMO²⁸ and EMMC²⁹ rules, and the FAIR²⁶ principles. **Sapphire** is amongst the first materials modelling post-processing tools to automatically write metadata in a standardised format. Contained within the metadata are, but not limited to, the following:

- Compressed forms of the written output to better facilitate a streamlined integration with potential subsequent interactive MNP and NA databases.
- User specific information to enable the community to better monitor the provenance of a given data set.
- Specifically chosen parameters for post-processing tools.

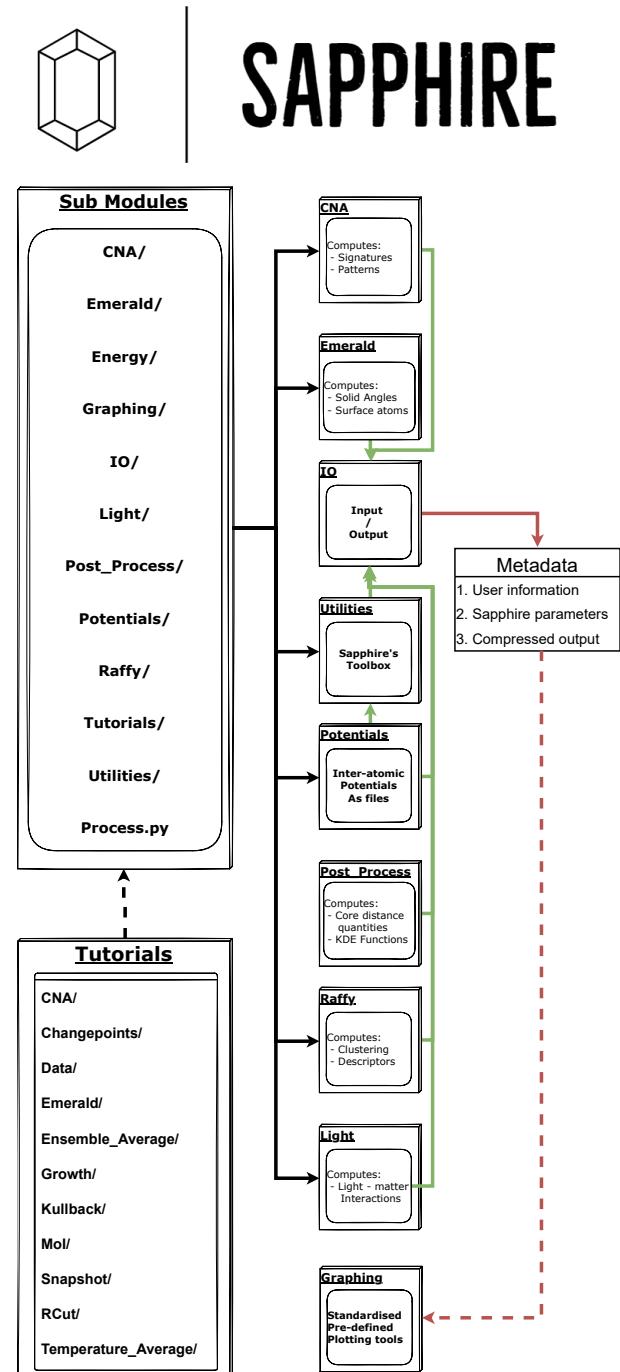


Fig. 1 Flowchart detailing the scheme under which **Sapphire** is executed. Black lines indicate module import directionality. Green lines indicate IO streams, red lines indicate data flow, solid lines indicate hard-coded streams, dashed lines highlight user choice.

Each parameter is described in details in the online manual.

The final stage of the flowchart in Fig. 1 lies in the use of `Graphing` tools. This is a library of pre-prepared matplotlib templates which may be called by the `Plot_Funcs` object defined within the library. From this, the `Figures` object is constructed from the collected metadata and the user requested input quantities. This parses the lists of input parameters for each requested

plotting function and introduces Sapphire's defaults, should any arguments be omitted. We then call the `Make_Plots` function of the `Figures` object which iterates over all of the plotting functions, passing into each one the relevant list of input arguments.

3 Distance distribution Functions

The distribution of pair-atomic distances (pair-distance distribution function, PDDF) is a crucial quantity to characterise the geometry and chemical ordering of an MNP or NA. Further, the PDDF is a directly measurable quantity via X-ray techniques³⁰.

To define the PDDF, let d_{ij} be the pair-distance between atoms i and j :

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} . \quad (1)$$

We then calculate the PDDF from the Kernel density estimates (KDE), constructed from n observations:

$$PDDF [K(d_{ij}, d; h)] = \frac{1}{Nh} \sum_i^N \sum_{j \neq i} K\left(\frac{d_{ij} - d}{h}\right) \quad (2)$$

where $K(d_{ij}, d; h)$ labels the kernel function over the d_{ij} variable. The parameter h is the bandwidth that defines the tightness of the kernel function. KDE assumes that each inter-atomic distance has been randomly drawn from a given distribution, which the user may define via input. The Gaussian distribution is the default choice. Alternatives, such as the Epanechnikov, and the uniform distributions, are also currently supported.

Fig. 2 shows the effect of the impact of the K and h choice in the PDDF calculation as a function of the distance d , reported in terms of the lattice bulk a_0 . Setting the bandwidth h is the most delicate step as it should balance a too fine resolution – where each distance might be present a single time, hence loosing all the information – and a too low one – where different neighbour shells are projected onto the same distance width. We suggest to consider values below half of the distance between the first and second nearest neighbour as a proper choice for distinguishing nearest neighbours peaks in solids. As a default, we have set the bandwidth to be 0.05 Å, which we have found to be sufficiently broad to smooth the sharp Dirac peaks from having a finite sample, and to sufficiently resolve key features, *i.e.*, the first peak, minimum, and the second peak (orange line in Fig. 2).

Given the analytical form of the KDE, derivatives can be easily calculated, and the approximate location of minima (and maxima) in the distribution can be identified. This is a key utility. The first PDDF minimum is instrumental to the robust determination of nearest neighbour distance. The position of the second maximum can identify whether the MNP undergoes to a phase change, if associate to an overall breathing of the longest pair-distance.

By monitoring the evolution of the pair-distance distribution function (PDDF), independently of their chemical label and even at the same time, enables the:

- Identification of a cutoff distance (R_{cut}) set as where the first minimum in the PDDF falls. The subsequent analyses depend on the R_{cut} value as it precipitates down to adjacency matrix compilation, as it determines the first shell of neigh-

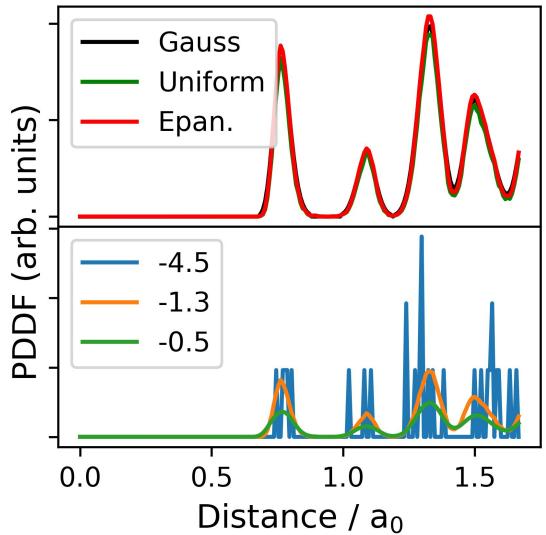


Fig. 2 The three possible parameterisations for calculating the PDDF using Eq. 2. Left panel demonstrates the agreement between available kernels. Right panel illustrates the dependence on the bandwidth, h , appropriately – values in the legend correspond to $\log_{10}(h)$.

bours around each atom.

- Detection of low symmetry shapes, depending on the position of the second peak of the PDDF. Where we define as "amorphous/ molten-like" a structure where there is not a maximum in correspondence of the bulk lattice, a_0 ³¹.

To clarify further this points, Fig. 3 details the PDDFs of pure MNP with 1415 atoms of Au, Pt, and AuPt with two different compositions. We refer to simulations data obtained from iterative molecular dynamics runs where the temperature is increased by 50 K each 1 ns, from 300 K to 1000 K^{32,33}.

We show cases both at cold (solid lines in panel (a)-(d) Fig. 3) and high temperatures (dashed lines in the same panels). We note a broadening of the first peak at higher T, both pure and alloyed nanosystems the position of the first peak does not change significantly with the temperature. We highlight the position of the first minimum (full circle dots) in the PDDF at both temperatures. From here, as it depends mildly on the temperature, a cut-off can be defined^{30,34}. The position of the second peak instead clear changes between cold and hot temperatures. This is more evident in the case of pure Au and Au-based nanoalloys. Pt-MNPs just show a minor effect due to its higher melting point. Indeed, the pure Pt-NP still presents a clear peak in correspondence of its bulk lattice. We also note that for NAs with a large amount of Pt, the second peak still present a shoulder in correspondence of the bulk lattice, suggesting that the phase change starts mainly in the gold subcluster (see panel (f) of Fig. 3 PDDF-line corresponding to Au-Au and Au-Pt bonding).

3.1 Radial distribution function

For nanoalloys, Sapphire provides three radial distributions: how the atoms are distributed from the centre of mass of the whole

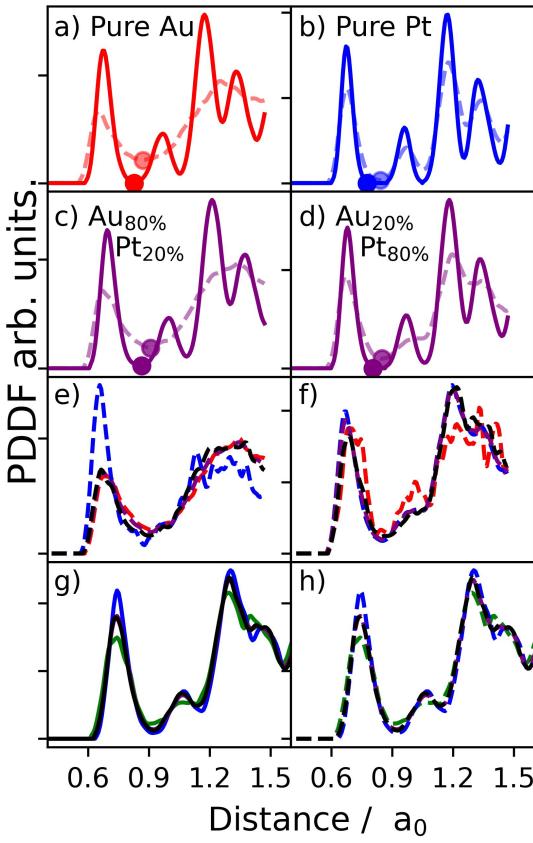


Fig. 3 PDDF computed with Sapphire for a set of pure Au, Pt, and AuPt NAs containing 1415 atoms and adopting initially an icosahedral (Ih) shape. Distances are scaled with respect to the bulk lattice distance a_0 . For NAs, this measure is taken as the average of the two pure metals' bulk lattice distance. Panels (a) and (b) refer to pure Au and pure Pt MNPs, respectively. Panels (c) and (d) refer to randomly alloyed AuPt NAs at different chemical compositions, as indicated in the inset. Solid lines correspond to the PDDF sampled at 300 K while dashed lines are taken at 1000 K. Dots represent the position of the first PDDF minimum, corresponding to R_{cut} . Panels e), f), g), and h) present the PDDF for homo interactions (Au - red, Pt - blue, Cu - green), and hetero interactions (black).

system; distribution of atoms-A from their centre of mass; and similarly for B-atoms.

The chemical ordering of a bi- or multi-metallic nanoalloy can be qualitatively, if not quantitatively, extracted from the radial distribution function (RDF) of the elements in the MNP. Similarly to the PDDF, this quantity can be readily extracted from both numerical simulations and (line-scan) experiment³⁵.

The RDF counts the number of atoms falling in concentric shells from the centre of mass of the nanoparticle,

$$r_\alpha(i) = \sqrt{(\hat{x}(i)_\alpha)^2 + (\hat{y}(i)_\alpha)^2 + (\hat{z}(i)_\alpha)^2} \quad (3)$$

where the coordinates $\hat{x}_\alpha, \hat{y}_\alpha, \hat{z}_\alpha$ of the atom- i and chemical species $\alpha = A, B$ are rescaled w.r.t the centre of mass of the whole MNP.

As was the case for computing the PDDF, we may also apply the KDE approach to compute the RDF for each of the three types

of radial distribution described above.

$$RDF [K(d_{ij}, d; h)] = \frac{1}{Nh} \sum_i^N \sum_{j \neq i} K(dist(CoM)) \quad (4)$$

Alternatively, Sapphire also supports the radial distribution as used to describe long-range order in liquids³⁶. We have elected to include this quantity, in supplement to the community standard definition, as it allows us to monitor order within the cluster, and is also obtainable via experimental techniques such as powder diffraction³⁶.

$$g(r) = \frac{dn(r)}{4\pi r^2 \rho dr} \quad (5)$$

Where ρ is the number density of atoms, $dn(r)$ is the density of atoms within a spherical shell thickness of dr .

Behind the radial distribution of atoms, a useful quantity to characterise the chemical ordering within a NA are the relative distance of the centre of mass of the two sub-clusters,

$$dist(CoM) = \sum_{i \in A} \frac{r_{i,A}}{N_A} - \sum_{i \in B} \frac{r_{i,B}}{N_B} \quad (6)$$

where $r_{i,A \in A}$ is the radial distance of the atom- i of species A taken from the centre of mass of the sub-cluster A. Similarly, for the B sub-cluster. $N_A(B)$ is the number of atoms A(B) in the nanoparticle. $dist(CoM)$ is expected to be close to zero for alloyed, core-shell, and multi-shell ordering, but not zero when the phase-segregation breaks the radial symmetry. This quantity has been shown to be significant to characterize the quasi-Janus chemical ordering for the first time in AuCo nanoalloys³⁷. We will show an example of $dist(CoM)$ for AuPt, averaged over the time in the Section 6.

4 Adjacency matrix based descriptors

The definition and characterization of nearest neighbour networks has been largely adopted to classify MNPs' morphology and draw structure-property relationships³⁸⁻⁴⁰.

To evaluate nearest neighbour networks and quantities deriving from the latter, solely according to a distance criterion, the adjacency matrix A is defined as:

$$A(r_{ij}) = \begin{cases} 1, & \text{if } r_{ij} \leq R_{cut} \\ 0, & \text{if } r_{ij} > R_{cut} \end{cases} \quad (7)$$

4.1 Coordination Numbers

One of the first quantity mentioned in any solid state books is the number of neighbours for the various Bravais lattice. With a focus on catalysis, the coordination of an adsorption site has been often used as a descriptor to rationalize its activity. Still related to the ability of a system to bind other molecule, in biophysics and soft matter, several algorithms have been developed to estimate the number of nearest neighbours of a macromolecules⁴¹. A simple chemical intuition suggests that low-coordinated atoms are more likely to form chemical bonds than highly-coordinated ones. The coordination number of an atom i follows from the

adjacency matrix:

$$CN_j = \sum_{i \neq j} \mathbf{A}(r_{ij}) . \quad (8)$$

Nonetheless, this definition might suffer of finding a suitable cutoff that might not be always easy to be selected, especially for nanoalloys with a large mismatch.

Sapphire contains another algorithm to defined the coordination number of atom-*i* which is based on the Van Meel's algorithm⁴¹ based on the solid angle (VMCN). The VMCN calculates the nearest-neighbour attributing to each possible neighbor a solid angle and determines the cutoff radius by the requirement that the sum of the solid angles must equal $4\pi^{41}$. First we calculate the shell-radius containing *m* neighbouring of the atom *i*, which is at a radial distance r_i

$$R_i^{(m)} = \sum_{j=1}^m \frac{r_{ij}}{m-2} . \quad (9)$$

The minimum of *m* is 3. In a FCC bulk, the maximum is expected to be 12. The second step is to calculate the solid angle as

$$\Omega_i = \pi \left(\frac{R_i^{(m)}}{r_i} \right)^2 . \quad (10)$$

Both algorithms can be used to analyse 3D snapshot, obtained from tomography as well as from numerical simulations.

The nominal CN definition and the one from the van Meel's algorithm agree generally well. The latter tends to show higher value of the coordination than the former, see less dark blue atoms in Fig. 4.

In both cases, it is possible to have atoms that are coordinated more than 12, even for metals that are FCC in the bulk. This is due to internal distortion and not for a faulty algorithm. No differences are found for geometrical closed geometries.

4.1.1 coordination numbers based descriptors

Other functions of the coordination number lead to useful descriptors. Among them, **Sapphire** provides the calculus of the atop generalised coordination number (aGCN)⁴² defined as :

$$aGCN_i = \sum_j \frac{CN_j}{CN_{max}} , \quad (11)$$

with CN_{max} set to 12, as this is the maximal coordination of an FCC atom in the bulk. The reason why we suggest to look at the aGCN is three-fold. First, it has been shown to provide a robust linear relationship with the adsorption energy of small molecules (e.g., O, CO, OH)⁴³. Second, the aGCN is able to characterise the MNPs surface sites -avoiding the basic classification into face, edge, vertex- and to classify a geometry on the basis of its aGCN-genome⁴⁴. Finally, the use of the aGCN offers a route to estimate the surface area beyond a spherical approximation assumption. The MNP's area can be well approximate by writing it as a sum over atomic contributions, which are a function of the atomic radius r_{at} weighted with their aGCN, as the latter is a measure of

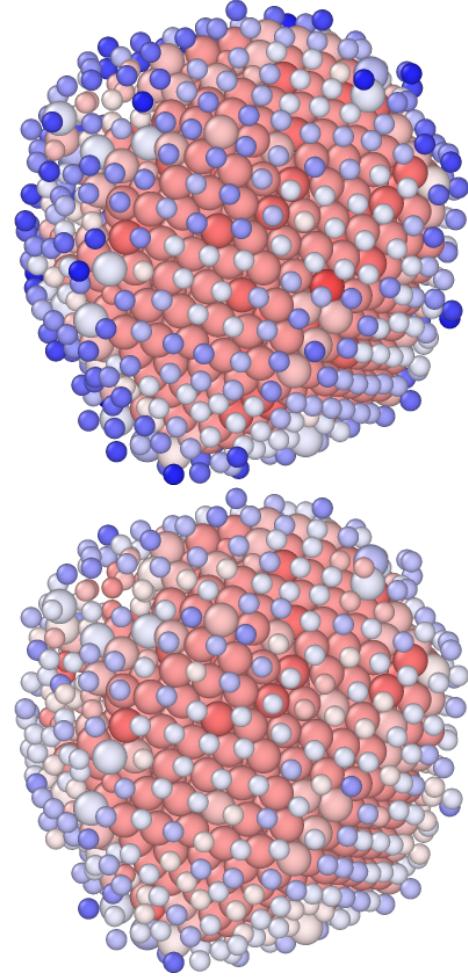


Fig. 4 Au₂₈₃Pt₁₁₃₂ NA where atoms are colored based on their nominal CN (left, with a cutoff of 3.55 Å) and the VMCN (right). Atoms are coloured with respect to their CN (or VMCN), ranging from 16 (full red) to 3 (dark blue). Au atoms are represented as smaller to Pt, to improve visualization. Minor discrepancies between the two algorithms occur. Using the solid angle condition, we count 561 atoms with a CN less than 10, 170 with CN=11, and 684 with a CN above or equal 12, Using the van Meel's algorithm, the occurrence is 477, 216, and 722, respectively.

how much they are "exposed":³¹

$$A_{MNP} = 4\pi \sum_i r_{at,i}^2 \left(1 - \frac{aGCN_i}{CN_{max}} \right) \quad (12)$$

A full comparison of the surface area calculated for closed shell geometries using cluster spherical approximation, geometrical consideration, and based on the aGCN is reported in Ref.³¹.

In the case of nanoalloys, we can easily separate the local environment of each atom between homo-pairs and hetero pairs. Counting the overall contribution summing over A-A pairs, NB_{AA} , B-B NB_{BB} , as well as A-B pairs, NB_{AB} enables the mixing parameter μ , a useful parameter for a fast classification of the chemical ordering into mixed or segregating NAs:

$$\mu = \frac{NB_{AA} + NB_{BB} - NB_{AB}}{NB_{AA} + NB_{BB} + NB_{AB}} . \quad (13)$$

μ tends to -1 when the NA is fully alloyed, and to +1 when there

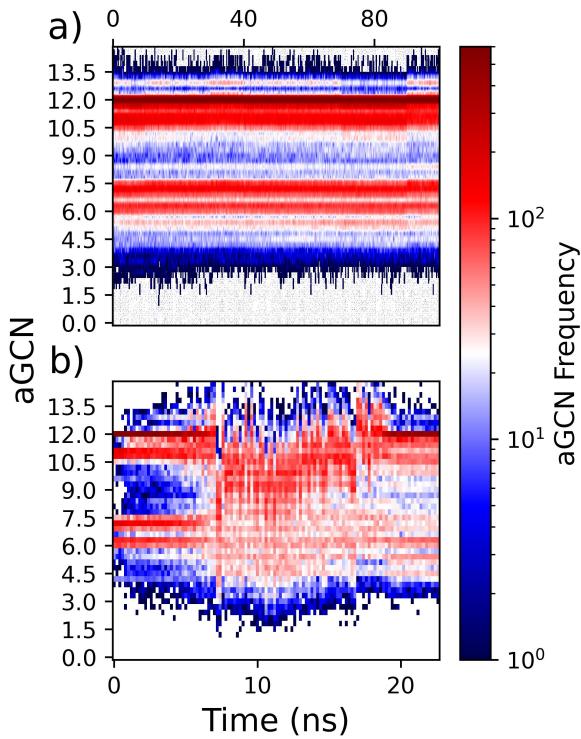


Fig. 5 Heat-maps demonstrating the evolution of the aGCN distribution, as calculated via the adjacency matrix, with respect to time for two types of dynamical processes. Panel a): two Pt_{55}^{lh} NAs deposited atop an Au_{2057}^{lh} MNP maintained at 600 K for 100 ns. Panel b): randomly alloyed $\text{Au}_{1283}\text{Pt}_{132}$ NAs with the initial morphology of an Ih heated from 300 K ($t=0$) to 1000 K $t=12$ ns and then cooled back to 300 K ($t=24$ ns).

is a complete phase separation.

4.2 Concertedness and collectivity of a rearrangement

Monitoring time-dependent changes in the adjacency matrix, over successive time-steps, allows to characterize whether structural rearrangements took place, and if these involve concerted and/or collective rearrangements. In this context, we adopt the definitions first put forward in Ref.³³.

Let $AA_{ij}(\Delta t)$ be a matrix counting the absolute number of bonds formed or lost between each ij pair of atoms, within a characteristic time length Δt :

$$AA_{ij}(\Delta t) = |A_{ij}(t + \Delta t) - A_{ij}(t)| \quad (14)$$

From this quantity, the system mobility, $R(t, t + \Delta t)$, is measured by summing over single atom mobility:

$$\begin{aligned} R(t, t + \Delta t) &= \sum_i R_i(t, t + \Delta t) \\ R_i(t, t + \Delta t) &= \sum_{j \neq i} AA_{ij}(\Delta t) \end{aligned} \quad (15)$$

To measure the concertedness of a mechanism, H , one then counts the ratio of atoms which change at least one neighbour within a

Δt interval, and the total number of atoms in the MNP:

$$H = \frac{\sum_i \Theta(R_i(t, t + \Delta t))}{N} \quad (16)$$

Where Θ labels an Heaviside step function and N the number of atoms in the MNP. By definition, the magnitude of H can vary between 0 and 1. The level of concertedness, C , is defined as the change in the number of atoms involved in the process between $(t - \Delta t, t)$, and between $(t, t + \Delta t)$

$$C(t - \Delta t, t, t + \Delta t) = |H(t - \Delta t, t) - H(t, t + \Delta t)| \quad (17)$$

When all the atoms in the MNP change their local connectivity time t , but the latter remains stable at the successive one $t + \Delta t$, $C(t - \Delta t, t, t + \Delta t)$ reaches its maximum value, 1.

Note, all the descriptors of mobility, concertedness, and collectivity discussed in this section display a dependence on the magnitude of $\Delta(t)$. A too long Δt may affect the H and C estimate by coarsening many atomic rearrangements into a single one. A too short Δt may unfaithfully describe a single step rearrangement as a multi-step one. The suggested (and default) value for the choice of this quantity is, $\Delta(t) = 10$ ps, which is consistent with the time-scale of adatom diffusion on low Miller index surfaces. We consider the jump of adatom as one of the fastest mechanisms that can take place during MNP's structural rearrangements.

All the quantities discussed above can be readily modified to account for the presence of multiple chemical species, $chem = AA, BB, AB$, in the NA:

$$R^{chem}(t, t + \Delta t) = \sum_i R_i^{chem}(t, t + \Delta t) \quad (18)$$

where R_i^{chem} is given by

$$R_i^{chem} = \sum_{j \neq i} AA_{ij}^{chem}(\delta t) \quad (19)$$

$$C^{chem}(t - \Delta t, t, t + \Delta t) = |H^{chem}(t - \Delta t, t) - H^{chem}(t, t + \Delta t)|. \quad (20)$$

Single-step mechanisms involving only a section of the cluster, or multi-step processes are characterized by lower values of $C(t - \Delta t, t, t + \Delta t)$, while continuous atomic rearrangements display $C \sim 0$. Indeed, this is precisely what we see in Fig. 6, as the C statistic is consistently near 0 whilst the H statistic monotonically increases as a function of temperature. This is indicative of a highly dynamical object whose constituent atoms are regularly rearranging with respect to one another, as at between each subsequent 10 ps window beyond 600 K, half of the atoms have a minimum of one change to their nearest neighbour network. Conversely, we draw attention to the sharp peak in the C statistic at the beginning of the dynamics. As with any statistical measure, this demonstrates that something interesting may be occurring at the beginning of the process. In this instance, this is representative of the collective motion of the Pt atoms within the cluster as their relatively large (to Au) cohesive energy creates an incentive to aggregate. This behaviour is indeed what we capture in the early stages of these dynamics.

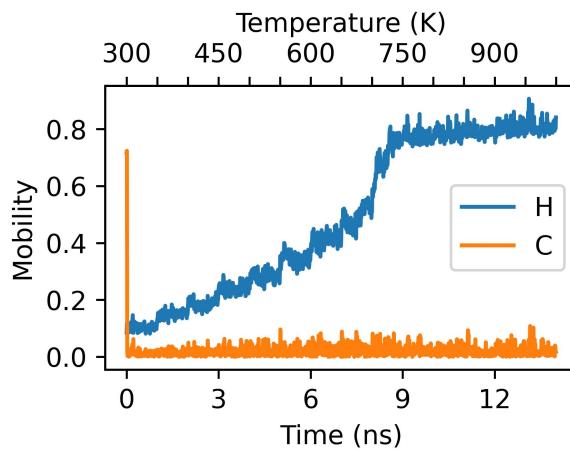


Fig. 6 Mobility descriptors for a randomly alloyed $\text{Au}_{1149}\text{Pt}_{266}$ NA in the configuration of an Ih which has been rapidly heated from 300 K to 1000 K over the span of 14 ns. Δt has been set to 10 ps.

4.3 Common Neighbour Analysis signatures and patterns

To classify the geometrical environments of core and surface atoms, we first evaluate all of the common neighbour analysis (CNA) signatures attributed to pairs of nearest neighbour atoms and averaged over the number of pairs. As seen for structural rearrangements, the nearest neighbour pair can be labelled as homo- (A-A and B-B type), and hetero- (A-B) when needed. However, to describe the overall geometry of a MNP and NA, we weight in the same manner A and B atoms. CNA signatures are of the form (rst) such that r is the number of nearest neighbours common to both atoms in the pair; s is the number of bonds between shared neighbours and t is the longest chain which can be made from bonding s atoms if they are nearest neighbours. By definition, CNA signatures describe the local environment of each pair of neighbours, summing and averaging typical contributions⁴⁵. While proven to be successful in many cases, the CNA is not a property of an atom i . Furthermore, the classification of structures, as available in commercial software as Ovito⁴⁶, is based on a few typical values of CNA signatures, and hence limiting the comparison to well-known shapes, as bulk geometries, icosahedra, and decahedra. Usually it is only meaningful for core atoms.

Sapphire extends the common neighbour analysis to become a property of each atom i . We list all the possible CNA signatures an atom is part of, and we count them (frequency chart). The local environment of each atom- i is, therefore, the collection of those signatures and their frequency into a CNA-pattern, as labelled into the first column of Table 1. A MNP with N atoms display between a few to a maximum of N CNA-patterns, depending on its symmetry. An octahedron presents only five different CNA patterns while a Marks decahedra has at least 11. As a rule of thumb, lowering the symmetry of the MNP shape increases the number of different CNA-pattern. In the extreme scenario, every atom will have a different local environment. Table 1 reports the most important features for common geometries, a short description of each pattern. We do stress that no restriction to any specific

signature occurs here.

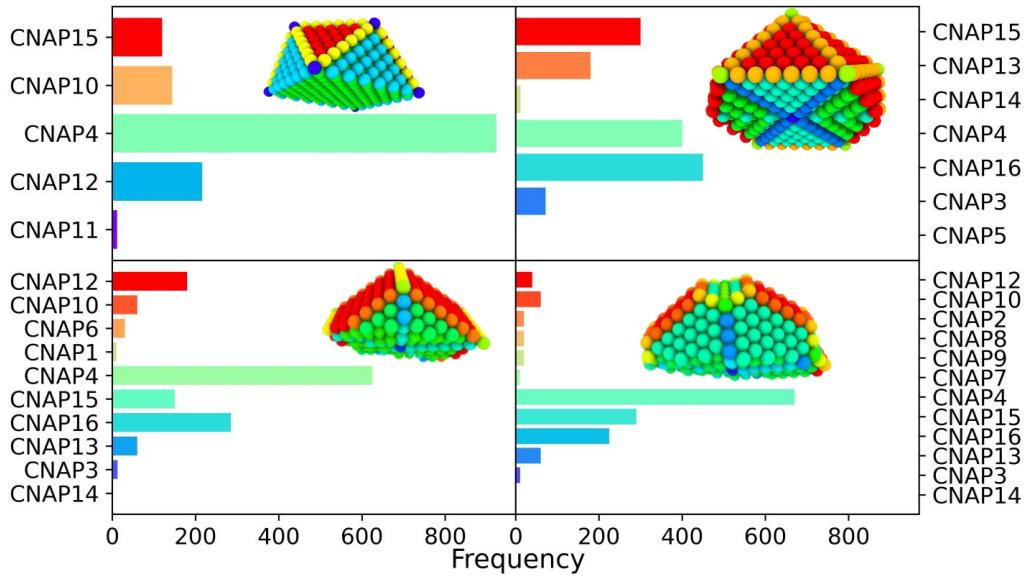


Fig. 7 Frequency chart of regularly-occurring CNA patterns in closed-shell geometries, namely a cubooctahedron (Co), an icosahedron (Ih), a Ino-decahedron (InoDh) with 1415 atoms, and a Marks-decahedron (mDh) with 1428 atoms. Snapshots of cut-in-half MNPs are shown in the insets, and atoms are coloured in agreement with their CNA-pattern, as listed in Table 1.

Table 1 Definition of regularly observed CNA Patterns and descriptions of the local environment in which they are found. We have identified whether or not a given pattern is associated with surface features to assist with the classification of regular NAs.

Pattern	CNAP-label	Description	Surface
[(1, (1, 0, 0)), (2, (2, 1, 1)), (1, (3, 2, 2)), (1, (4, 2, 2))]	CNAP1	Re-entrance centre bounded by (111) facets	Y
[(1, (2, 0, 0)), (2, (2, 1, 1)), (2, (3, 1, 1)), (1, (4, 2, 1))]	CNAP2	Corners of box bounded by between re-entrance lines, (100) and (111) Facets.	Y
[(10, (4, 2, 2)), (2, (5, 5, 5))]	CNAP3	HCP twinning planes	N
[(12, (4, 2, 1))]	CNAP4	Bulk face centred cubic (FCC) structure	N
[(12, (5, 5, 5))]	CNAP5	Centre of an icosahedron	N
[(2, (1, 0, 0)), (2, (2, 1, 1)), (2, (4, 2, 2))]	CNAP6	Boundary between (100) Facets. facets	Y
[(2, (2, 0, 0)), (1, (3, 0, 0)), (2, (3, 1, 1)), (1, (3, 2, 2)), (1, (4, 2, 2))]	CNAP7	Corner atoms connecting the re-entrance line with the five-fold symmetry axes.	Y
[(2, (2, 0, 0)), (4, (3, 1, 1)), (1, (4, 2, 1))]	CNAP8	Vertex between re-entrance and five-fold symmetry axes	Y
[(2, (3, 0, 0)), (4, (3, 1, 1)), (2, (4, 2, 1)), (2, (4, 2, 2))]	CNAP9	Re-entrance centre bounded by (111) facets	Both
[(3, (2, 1, 1)), (2, (3, 1, 1)), (2, (4, 2, 1))]	CNAP10	Boundary between (100) and (111) facets	Y
[(4, (2, 1, 1)), (1, (4, 2, 1))]	CNAP11	Vertex between (100) and (111) Facets	Y
[(4, (2, 1, 1)), (4, (4, 2, 1))]	CNAP12	(100) Facet.	Y
[(4, (3, 1, 1)), (2, (3, 2, 2)), (2, (4, 2, 2))]	CNAP13	Five-fold symmetry axes separating (111) facets	Y
[(5, (3, 2, 2)), (1, (5, 5, 5))]	CNAP14	Tip of the five-fold symmetry axes	Y
[(6, (3, 1, 1)), (3, (4, 2, 1))]	CNAP15	(111) Facet	Y
[(6, (4, 2, 1)), (6, (4, 2, 2))]	CNAP16	Bulk hexagonal close packed (HCP) structure	N

Given the evident sensitivity to local geometry, the next question to ask is how sensitive are these patterns to imperfections in the crystalline structure?

Generally speaking, one can distinguish three main morphology families as FCC-like, decahedra (Dh), and icosahedra (Ih), depending the number and type of CNA-patterns. The two main

patterns to look at are the [10(422), 2(555)]; [4(311), 2(322), 2(422)]. Their appearance stands for a Dh and or a Ih. The [12(555)] is unique of an Ih, with the intercept of its fivefold axes lying within the MNP. Line displacement in a FCC arrangement arise as a mixture of [l(422) k(421)] with different l,k frequency, such as $l + k = 12$.

5 Surface Identification

As many physical and chemical processes occur at the surface,^{47–49} an automated classification of which atoms stay at the surface, subsurface, and core is welcome. The classification of core and subsurface atoms will occur "peeling-off" the first layer. core atoms are all those atoms that do not belong to surface and sub-surface. The Reader should note that combining the identification of atoms in the surface, subsurface, and core shells will allow to calculate other atomistic properties, i.e. related to the energy when calculated as atomistic contributions. Hence, the **Sapphire** users will have access to surface energy, local stress, local pressure. However, as those quantities depend on the choice of a proper energy calculator, see last Section, we will discuss this later. While on perfectly built geometries the identification of atoms at the surface follows their coordination number, i.e. imposing $CN \leq 10$, for unknown, low-symmetry, amorphous morphology, and mainly for elongated and amorphous this task is far from being trivial. Any classification simple based on the coordination number fails, independently on how the coordination number is calculated, as both low and high coordinated atoms can be present at the surface, see the heatmap in Fig.4. A combination of a condition on CN and radial distance will work only for spherical MNPs, but will fail for prolate/oblate systems. The need of finding other algorithms that do not require any pre-knowledge.

For this reasons, we implement other three algorithms to list and classify atoms at the surface. (i) The first criterion is fast and based on imposing a threshold on the solid angle Ω_i , as from Eq.10, combined with the radial distance of atoms i . This method reproduces data for closed shell geometries, and provides a good estimate for melted/low symmetry MNPs and NAs. On the other hand, it suffers from the introduction of a threshold value on the solid angle that depends on inversely on the size of the MNP. (ii) **Sapphire** offers the "Layers algorithm", usually adopted to calculate the accessible area in macromolecules citeKarampudi2015. This algorithm creates three cylinders per each atom along the x, y, z Cartesian components, and record the lowest, highest atom in each cylinder. In this case, there is only a free parameter tuning the radius of the cylinder. As a rule of thumb, choosing 1.5 of the atomic radius (or of the average atomic radius for binary systems) provides a good definition. A comparison of the two algorithms is shown in Fig.8 for Au and CuPt cases with 1415 atoms.

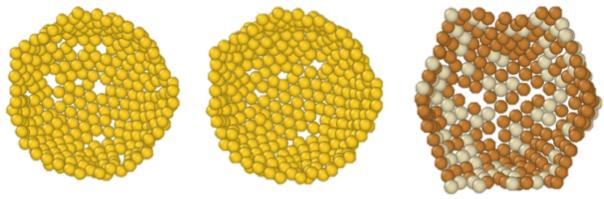


Fig. 8 Comparison between the solid-angle (left) and the Layers (middle and right) algorithms to identify the surface layer. here we display only half of the external shell of a low symmetry Au_{1415} . We use the condition on the solid angle as $\Omega_i \leq 0.05$ combined with $r_i \geq 15 \text{ \AA}$ (left). On the right, the Layer method applied to a low symmetry $CuPt_{1415}$ with a chemical composition of 0.3 for Pt (silver-ish coloured and Cu is orange).

Finally, we introduce and implement a clustering approach based on local atomic environment descriptors.³⁴ The clustering approach is used to classify atoms in Au NPs, and implemented in the RAFFY Python package⁵⁰, labels atoms in an unsupervised manner via hierarchical k-means clustering. The atoms from one or more MD snapshots are fingerprinted using local descriptors based on the atomic cluster expansion⁵¹ (ACE) framework. ACE descriptors encode information of local atomic environments by approximating the local atomic density via spherical harmonics and radial basis expansions, and then constructing rotation-invariant representations from the coefficients of such expansions. A cut-off radius, alongside other parameters, must be specified for the ACE descriptors; typical cut-off values are comprised between the distance of the second and third shell of neighbours. This approach unbiasedly distinguishes between low-coordination surface, highly coordinated surface, and bulk atoms.³⁴ Moreover, it enables the discernment of locally melted and locally solid atomic environments, therefore providing a robust measure of melting and surface rearrangement temperatures.

6 Averaging routes

There is support for creating plots for a single analysed trajectory or to consider the average over multiple, independent simulations. One can toggle between these alternatives by setting the `single_file` flag to be true and `iter_dir` to be false, or by setting `single_file` flag to be false and setting the `iter_dir` flag to be a list of relative paths from the defined base directory to each of the iterations to be considered. It should be noted that this is simply the arithmetic average for each of the requested quantities, and caution is advised to users who take advantage of this feature as it may not always be meaningful to compute or present such averages. Indeed, these meta-analyses are only meaningful if the ensemble being sampled across is expected to be approximately constant over time, and if the dynamics are near an equilibrium.

When reporting and presenting these ensemble averages, **Sapphire** only computes uncertainties across samples as the standard error. That is to say that across the ensemble, we assume that all deviations are independently and identically distributed (i.i.d.). Indeed, the former is trivial to guarantee. However, given the extensive phase space of dynamical systems, it cannot be guaranteed that there exists a distribution for a given descrip-

tor to sample. This is especially true if one is intending to probe non-equilibrium dynamics for which a steady state distribution cannot be uniquely defined at a given point in the dynamics.

Nonetheless, when the dynamics of the system are not far from equilibrium, this scheme provides a fast and intuitive method to visualise variance and spread in the descriptors.

Fig. 9 reports four ensemble averaged quantities for a coalesced $\text{Au}_{923}^{lh}\text{Pt}_{55}^{lh}$ held at 450 K for 500 ns. The temperature is too low to activate any structural rearrangements. We have elected to present this system to demonstrate the utility of the Sapphire averaging scheme. We define the local heterogeneous atomic environment (LAE) as being the percentage of specie A with LAE neighbours of specie B. $LAE = 0$ indicates that these atoms experience no heterogeneous bonding environments. $1 \leq LAE \leq 6$ describes an interface environment; while $LAE > 9$ suggest that these atoms are nearing total encapsulation by specie B. Initially, we note that all descriptors are found to have only small inter-sample variations at approximately 10% of the mean value. Moreover, we see evidence of steady dynamics in the relative constancy of the radius of gyration (RoG), the number of Pt at the surface (Pt_{surf}), and LAE of Pt versus time. However, looking at the distance between the subcluster centres of mass $dist(\text{CoM})$, a structural transformation occurs. In the presented example, this migration of clusters appears to occur consistently with variations of only a few Å between samples.

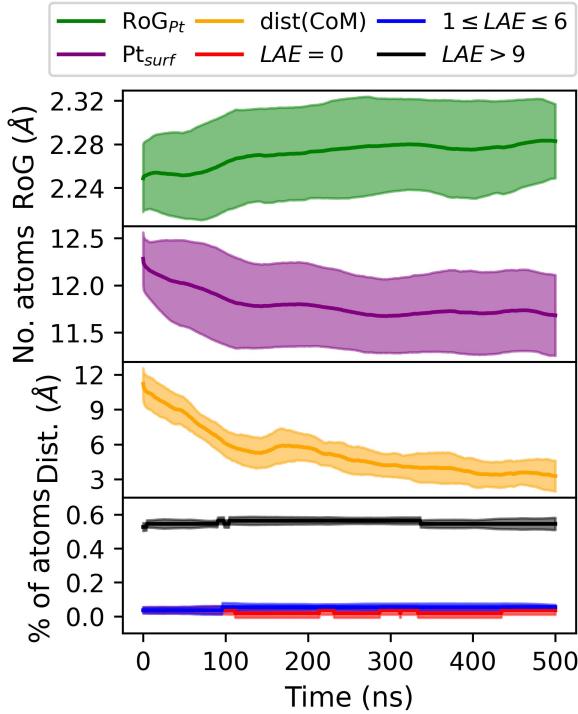


Fig. 9 Ensemble averaged evolution of atomic descriptors for a coalesced system of $\text{Au}_{923}^{lh}\text{Pt}_{55}^{lh}$ thermalised at 450 K for 500 ns. Each of these quantities is reporting the atomic environment as experienced by the Pt subcluster. I.e., The Radius of Gyration is reported for the Pt cluster only; and $dist(\text{CoM})$ reports the mean distance of Pt atoms to the system centre of mass.

We do not present these averaging techniques to serve as a replacement for an individual's choice of averaging routine, as these are unique and particular to the dynamics and quantities under consideration. Rather, we present this utility as a means to explore and present variations across i.i.d. samples.

7 Structure - Properties relationships

The Sapphire library focuses on the characterisation of MNPs and NAs morphology and to provide structural insights. Nonetheless, Sapphire can be used in synergy with available codes geared toward the prediction of chemophysical properties of MNPs and NAs. We briefly discuss three examples, where Sapphire will benefit from its ASE environment to move back to energy-related properties; the link with the NanoCHE⁵² to estimate its electrochemical properties; the pyGDM⁵³ to qualitatively estimate the optical properties.

7.1 From structure to energy

Often it is desirable to check the energy stability of a certain shape or chemical ordering. Currently, ASE supports effective medium (EMT), the embedded atom method (EAM), and the kimpyp calculators. EAM benefits from the Interatomic Potentials Repository Project available at <https://www.ctcms.nist.gov/potentials/>, while OpenKIM is a recent NSF project devoted to provide access to classical interatomic potentials. Unfortunately, second moment approximation of the tight binding (SMATB) calculators are not currently available. They were part of the ASAP and ASAP3 distribution, actually dismissed. On the Sapphire online manual we show how to call those ASE libraries.

7.2 Electrocatalytic properties via a microkinetic model

Let α be the value of a descriptors which encodes an explicit scaling between the structural properties of an adsorption site the reaction free energy associated to the rate-limiting step of an electrochemical reaction occurring at that site. A simple microkinetic model for the current density $j_{\text{nano particle}}(t, T, U)$ produced by a nanoparticle for such an electrochemical process can be written as:

$$j_{\text{nano particle}}(t, T, U) = \sum_{\alpha} \mathcal{C} \xi(t, T) \alpha e^{\beta \Delta G(U, T, \alpha)}, \quad (21)$$

where $\Delta G(U, T, \alpha)$ labels the reaction free energy at the applied potential U and temperature T , which is also a function of the descriptor α , β is the Maxwell-Boltzmann factor, $\xi(t, T) = \frac{\Omega(\alpha)}{N_{\text{site}}(t, T)}$ is the fraction of non-equivalent adsorption sites $F_{\text{nano particle}}$ to the total number of sites available, N_{site} , and the sum runs over the collection of the non equivalent sites appearing in the MNP under consideration. The latter are also characterized by their α values. Note that in 21 the effect of the potential is treated as an a posteriori correction, in line with the computational hydrogen electrode model⁵⁴. A suitable code is available at the Github page The aGCN, Eq. 11, of a surface site is an accurate descriptor to predict the activity of a metallic adsorption site, as demonstrated for Au, Cu, Pt, PtNi, catalysing the electrochemical oxygen or carbon dioxide reduction^{42,43,55}. Assuming a volcano-plot relation-

ship, the reaction free energy ΔG and the aGCN are related by the following general expression:

$$\Delta G = \begin{cases} +a1 \text{ aGCN}_n - b1 & \text{if } \text{aGCN}_n < \text{aGCN}^{\text{volcano-peak}} \\ -a2 \text{ aGCN}_n + b2 & \text{if } \text{aGCN}_n > \text{aGCN}^{\text{volcano-peak}} \end{cases} . \quad (22)$$

where the coefficients $a1$, $b1$, $a2$, $b2$, and the value of the $\text{aGCN}^{\text{volcano-peak}}$ should be derived from a small set of DFT calculations, if not available in the literature. For example, the case of ORR on Pt³¹.

Once the Gibbs free energy is defined in terms of the aGCN, a quantity that can be monitored during the lifetime of a MNP, we can establish relationships between the ageing of a MNP, referring to its structural stability, and its activity. Examples are reported previously by our group for ORR on Pt³¹ and for the conversion of CO₂ into methane on Cu NPs⁵². Finding a good volcano-like relationship, we do not see any limitation in using the scheme for NAs too.

7.3 Optical properties via semi-classical methods

For a fast evaluation and screening of the extinction spectrum of MNP and NA, one can adopt a classical approach, via the Green's Dyadic Method (GDM),^{53,56} as implemented in the pyGDM code^{57,58}. At its core, this library constructs a refractive environment via the construction of a 3D mesh of dipole oscillators whose physics is predicted within the quasi-static coupled dipole approximation. It then uses an efficient and generalised propagator to predict the extinction spectrum associated to the used-defined mesh. This mesh may be user defined from a series of coordinates consisting of multiple constitutive components.

To evaluate the extinction spectrum, we need only to consider the interaction of the dipole moment and total field in each discretised volume cell:

$$\sigma_{\text{ext}}(\omega) = \frac{8\pi^2}{n_{\text{env}}\lambda_0} \sum_j^{N_{\text{cells}}} \Im \left(\mathbf{E}_{0,j}^* \cdot \mathbf{P}_j \right), \quad (23)$$

wherein n_{env} , λ_0 , and \mathbf{P}_j are respectively the dielectric constant of the embedding environment, the incident wavelength, the dipole moment of cell j , and $E_{0,j} = E_0(r_j)$ is the incident field⁵⁹. In the described method, the near field may be found by self-consistently solving

$$\mathbf{E}(\mathbf{r}_i, \omega) = \mathbf{E}_0(\mathbf{r}_i, \omega) + \sum_j^{N_{\text{Cells}}} G_{\text{tot}}^{\text{EE}}(\mathbf{r}_i, \mathbf{r}_j, \omega) \chi \mathbf{E}'(\mathbf{r}_j, \omega) V_{\text{cell}},$$

where $G_{\text{tot}}^{\text{EE}}(\mathbf{r}_i, \mathbf{r}_j, \omega)$ is the Green Dyadic to be solved, χ is the metal's susceptibility, and V_{cell} is the volume of a unit cell. From this field, we then calculate the effective dipole moments, $\mathbf{P}_j = V_{\text{cell}} \chi_j \cdot \mathbf{E}_j$, from the internal field distribution. From a knowledge of the functional form of the incident field, and a sufficiently converged internal field distribution, one is able to compute the extinction spectrum for a given finite system held within a dispersive medium, as described by Eq. 23. All metallic dielectric parameters used in these calculations are provided by either experimental work^{60–62}, or from *ab initio* investigations⁶³. The interested

reader is referred to previous literature for further detail^{53,57,58}.

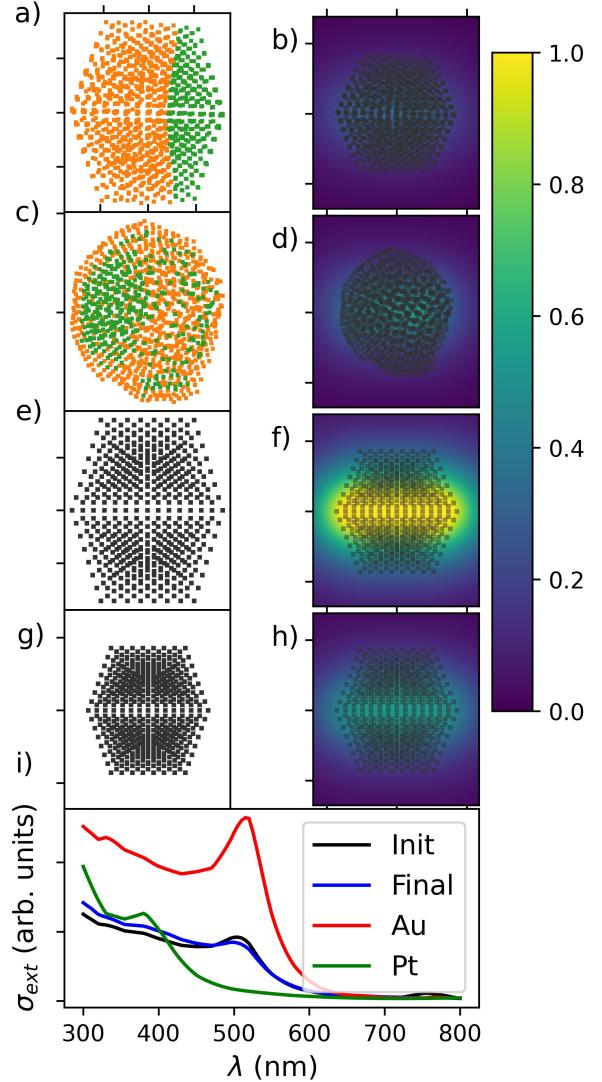


Fig. 10 An illustration of the interaction between Sapphire and pyGDM. Panel a): Au^{Ih}₁₁₃₂Pt^{Ih}₂₈₃ consisting of Au (orange), and Pt (green) showing a Janus chemical ordering (a). Panel c): Au^{Ih}₁₁₃₂Pt^{Ih}₂₈₃ consisting of Au (orange), and Pt (green) showing a shell-core chemical ordering. Panels e) and in g): pure MNPs of Au₁₄₁₅ and Pt₁₄₁₅, respectively. Panels b), d), f), h): near-field enhancement of the structure when illuminated by plane waves at 550 nm. Panel i)): photo-extinction spectra as computed via Eq. 23 for the examples showcased in Panels a, c, e, and g.

Any behaviour that may arise from quantum many-body effects is neglected. Furthermore, when considering structures with size ~ 1 nm illuminated in the ultraviolet-visible–near-infrared, the incident field will only weakly couple to the structure. This results in non-trivial internal field enhancement, which is not observed. We strongly suggest to use the GDM

Fig. 10 reports the dipole mesh reconstructed from atomic coordinates and species in the left set of panels, and the near-field enhancement at 520 nm plane wave illumination on the right side. Intensity, as described by the colour bar, is reported as $\mathbf{E}(\mathbf{r}) / \mathbf{E}_0$, the strength of the field relative to the amplitude of the illuminating wave. In the bottom panel, we present the extinc-

tion spectra computed via Eqn. 23 for four different systems. We have elected to present these systems to demonstrate the utility in being able to directly map a set of atomic coordinates, as may be parsed from a structural file, to a mesh of coupled dipoles with pre-defined dielectric properties.

8 Conclusions

The characterization and classification of the morphology and chemical ordering in mono-, bi-, and multi-metallic NAs is a key ingredient towards rationalizing their chemo-physical properties.

The necessity of an open-source robust, reproducible, and FAIR compliant post-processing tools, which caters the needs of the nanoalloy computational modelling community, is more and more evident. **Sapphire**, the open source software we here presented, is tailored specifically to address this need, by providing a library of standardised analysis tools for NAs characterization. **Sapphire** is indeed an open-source, modular, user-friendly platform to characterise an nanoparticle's and NAs' architecture observed during, e.g., atom coordinates reconstructions from experiment, molecular dynamics or Monte Carlo based sampling.

Within **Sapphire**, several order-parameters and descriptors can be calculated, namely: pair distance and radial distribution functions, inertia tensors and quantity derived from the latter, coordination distributions and other topological descriptors (generalized coordination and common neighbour analysis) derived from adjacency matrices. **Sapphire** offers a detailed analysis of the surface of MNPs and NAs. In the long term, it is our vision that **Sapphire** may serve to create a standardised, meaningful, characterisation and classification tool. This will be the first critical stage in creating a comprehensive community database for such complex systems.

Corresponding Author

Francesca Baletto

Physics Department, King's College London, Strand WC2R 2LS, United Kingdom

Physics Department, University of Milan, Via Celoria 16, 20133 Italy Email: francesca.baletto@unimi.it

Authors

Robert M. Jones

Physics Department, King's College London, Strand WC2R 2LS, United Kingdom Email: robert.m.jones@kcl.ac.uk

Kevin Rossi

Email: kevin.rossi@epfl.ch

Claudio Zeni

Email: czeni@sissa.it

Igor Vasiljevic

Email: igor.Vasiljevic@studenti.unimi.it

Alejandro Santana-Bonilla

Physics Department, King's College London, Strand WC2R 2LS, United Kingdom

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This project is funded by Engineering and Physical Sciences Research Council. Via our membership of the UK's HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202), this work used the UK Materials and Molecular Modelling Hub for computational resources, MMM Hub, which is partially funded by EPSRC (EP/P020194 and EP/T022213). This work used the ARCHER2 UK National Supercomputing Service (<https://www.archer2.ac.uk>). R. Jones acknowledges 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).

Notes and references

- 1 B. A. Prabowo, A. Purwidyantri, and K.-C. Liu, "Surface plasmon resonance optical sensor: A review on light source technology," *Biosensors*, vol. 8, no. 3, 2018.
- 2 A. A. Yaqoob, H. Ahmad, T. Parveen, A. Ahmad, M. Oves, I. M. I. Ismail, H. A. Qari, K. Umar, and M. N. Mohamad Ibrahim, "Recent advances in metal decorated nanomaterials and their various biological applications: A review," *Frontiers in Chemistry*, vol. 8, 2020.
- 3 M. Rahman, K. Alam, A. Hafeez, R. Ilyas, and S. Beg, "Chapter 7 - metallic nanoparticles in drug delivery and cancer treatment," in *Nanoformulation Strategies for Cancer Treatment* (S. Beg, M. Rahman, H. Choudhry, E. B. Souto, and F. J. Ahmad, eds.), Micro and Nano Technologies, pp. 107–119, Elsevier, 2021.
- 4 H. Chen, Y. Zhou, and S.-T. Han, "Recent advances in metal nanoparticle-based floating gate memory," *Nano Select*, vol. 2, no. 7, pp. 1245–1265, 2021.
- 5 H. Ditlbacher, J. R. Krenn, B. Lamprecht, A. Leitner, and F. R. Aussenegg, "Spectrally coded optical data storage by metal nanoparticles," *Opt. Lett.*, vol. 25, pp. 563–565, Apr 2000.
- 6 K. L. Kelly, E. Coronado, L. L. Zhao, and G. C. Schatz, "The optical properties of metal nanoparticles: The influence of size, shape, and dielectric environment," *The Journal of Physical Chemistry B*, vol. 107, no. 3, pp. 668–677, 2003.
- 7 E. A. Coronado, E. R. Encina, and F. D. Stefani, "Optical properties of metallic nanoparticles: manipulating light, heat and forces at the nanoscale," *Nanoscale*, vol. 3, pp. 4042–4059, 2011.
- 8 J.-D. Xiao and H.-L. Jiang, "Metal–organic frameworks for photocatalysis and photothermal catalysis," *Accounts of Chemical Research*, vol. 52, no. 2, pp. 356–366, 2019.
- 9 D. Mateo, J. L. Cerrillo, S. Durini, and J. Gascon, "Fundamentals and applications of photo-thermal catalysis," *Chem. Soc. Rev.*, vol. 50, pp. 2173–2210, 2021.
- 10 K. Lorber and P. Djinović, "Accelerating photo-thermal co₂ reduction to co, ch₄ or methanol over metal/oxide semiconduc-

- tor catalysts,” *iScience*, vol. 25, no. 4, p. 104107, 2022.
- 11 Y. Tang and W. Cheng, “Key parameters governing metallic nanoparticle electrocatalysis,” *Nanoscale*, vol. 7, pp. 16151–16164, 2015.
- 12 S. D. Lacey, Q. Dong, Z. Huang, J. Luo, H. Xie, Z. Lin, D. J. Kirsch, V. Vattipalli, C. Povinelli, W. Fan, R. Shahbazian-Yassar, D. Wang, and L. Hu, “Stable multimetallic nanoparticles for oxygen electrocatalysis,” *Nano Letters*, vol. 19, no. 8, pp. 5149–5158, 2019. PMID: 31313586.
- 13 J. Yu, Q. Liu, W. Qiao, D. Lv, Y. Li, C. Liu, Y. Yu, Y. Li, H. Niemantsverdriet, B. Zhang, and R. Su, “Catalytic role of metal nanoparticles in selectivity control over photodehydrogenative coupling of primary amines to imines and secondary amines,” *ACS Catalysis*, vol. 11, no. 11, pp. 6656–6661, 2021.
- 14 H. Song, “Metal hybrid nanoparticles for catalytic organic and photochemical transformations,” *Accounts of Chemical Research*, vol. 48, no. 3, pp. 491–499, 2015. PMID: 25730414.
- 15 T. P. Rossi, P. Erhart, and M. Kuisma, “Hot-carrier generation in plasmonic nanoparticles: The importance of atomic structure,” *ACS Nano*, vol. 14, no. 8, 2020.
- 16 V. Amendola, R. Pilot, M. Frasconi, O. M. Maragò, and M. A. Iati, “Surface plasmon resonance in gold nanoparticles: a review,” *Journal of Physics: Condensed Matter*, vol. 29, no. 20, 2017.
- 17 V. G. Rao, U. Aslam, and S. Linic, “Chemical requirement for extracting energetic charge carriers from plasmonic metal nanoparticles to perform electron-transfer reactions,” *Journal of the American Chemical Society*, vol. 141, no. 1, 2019.
- 18 W. Hong and C. W. Li, “Microstructural evolution of aupt core-shell nanoparticles under electrochemical polarization,” *ACS Appl. Mater. Interfaces*, vol. 11, 2019.
- 19 L.-M. Guo, Zhang, D.-F., and L. Guo, “Structure design reveals the role of au for orr catalytic performance optimization in ptco-based catalysts,” *Adv. Funct. Mater.*, vol. 30, p. 2001575, 2020.
- 20 M. S. P. Jagannath, S. Divi, and A. Chatterjee *J. Phys. Chem. C*, vol. 122, p. 26214–26225, 2018.
- 21 “Gold segregation improves electrocatalytic activity of icosahedron au@ptnanocluster: insights from machine,” *Chin. J. Chem.*, vol. 39, 2021.
- 22 V. Ramasubramani, B. D. Dice, E. S. Harper, M. P. Spellings, J. A. Anderson, and S. C. Glotzer, “freud: A software suite for high throughput analysis of particle simulation data,” *Computer Physics Communications*, vol. 254, p. 107275, 2020.
- 23 Richard J. Gowers, Max Linke, Jonathan Barnoud, Tyler J. E. Reddy, Manuel N. Melo, Sean L. Seyler, Jan Domański, David L. Dotson, Sébastien Buchoux, Ian M. Kenney, and Oliver Beckstein, “MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations,” in *Proceedings of the 15th Python in Science Conference* (Sebastian Benthall and Scott Rostrup, eds.), pp. 98 – 105, 2016.
- 24 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, “Mdanalysis: A toolkit for the analysis of molecular dynamics simulations,” *Journal of Computational Chemistry*, vol. 32, no. 10, pp. 2319–2327, 2011.
- 25 R. M. Jones, “Sapphire.” <https://github.com/kcl-tscm/Sapphire.git>, 2022.
- 26 C. Draxl and M. Scheffler, “Nomad: The fair concept for big-data-driven materials science,” 2018.
- 27 A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, “The atomic simulation environment—a python library for working with atoms,” *Journal of Physics: Condensed Matter*, vol. 29, p. 273002, jun 2017.
- 28 B. Stiller, T. Bocek, F. Hecht, G. Machado, P. Racz, and M. Waldburger, “Mobile Systems IV,” tech. rep., University of Zurich, Department of Informatics, 01 2010.
- 29 N. Adamovic, “Documentation on Materials Modelling Ontology in UML,” tech. rep., TU WIEN, 01 2016.
- 30 L. Delgado-Callico, K. Rossi, R. Pinto-Miles, P. Salzbrenner, and F. Baletto, “A universal signature in the melting of metallic nanoparticles,” *Nanoscale*, vol. 13, no. 2, 2021.
- 31 K. Rossi, G. G. Asara, and F. Baletto, “Structural screening and design of platinum nanosamples for oxygen reduction,” *ACS Catalysis*, vol. 10, no. 6, pp. 3911–3920, 2020.
- 32 K. Rossi, L. Pavan, Y. Soon, and F. Baletto, “The effect of size and composition on structural transitions in monometallic nanoparticles,” *The European Physical Journal B*, vol. 91, p. 33, Feb 2018.
- 33 K. Rossi, L. B. Pártay, G. Csányi, and F. Baletto, “Thermodynamics of cupt nanoalloys,” *Scientific Reports*, vol. 8, p. 9150, Jun 2018.
- 34 C. Zeni, K. Rossi, T. Pavloudis, J. Kioseoglou, S. de Gironcoli, R. E. Palmer, and F. Baletto, “Data-driven simulation and characterisation of gold nanoparticle melting,” *Nature Communications*, vol. 12, no. 1, 2021.
- 35 R. B. Neder and V. I. Korsunskiy, “Structure of nanoparticles from powder diffraction data using the pair distribution function,” *Journal of Physics: Condensed Matter*, vol. 17, pp. S125–S134, jan 2005.
- 36 R. E. Dinnebier and S. J. L. Billinge, eds., *Powder Diffraction*. The Royal Society of Chemistry, 2008.
- 37 B. Farkaš and N. H. de Leeuw, “Auco nanoparticles: ordering, magnetisation, and morphology trends predicted by dft,” *Phys. Chem. Chem. Phys.*, vol. 24, pp. 10451–10464, 2022.
- 38 S. Mourdikoudis, R. M. Pallares, and N. T. K. Thanh, “Characterization techniques for nanoparticles: comparison and complementarity upon studying nanoparticle properties,” *Nanoscale*, vol. 10, pp. 12871–12934, 2018.
- 39 D. E. Perea, I. Arslan, J. Liu, Z. Ristanović, L. Kovarik, B. W. Arey, J. A. Lercher, S. R. Bare, and B. M. Weckhuysen, “Determining the location and nearest neighbours of aluminium in zeolites with atom probe tomography,” *Nature Communications*, vol. 6, p. 7589, Jul 2015.

- 40 A. Stukowski, "Structure identification methods for atomistic simulations of crystalline materials," *Modelling and Simulation in Materials Science and Engineering*, vol. 20, no. 4, 2012.
- 41 J. A. van Meel, L. Filion, C. Valeriani, and D. Frenkel, "A parameter-free, solid-angle based, nearest-neighbor algorithm," *J. Chem. Phys.*, vol. 136, p. 234107, 2012.
- 42 F. Calle-Vallejo, J. I. Martínez, J. M. García-Lastra, P. Sautet, and D. Loffreda, "Fast prediction of adsorption properties for platinum nanocatalysts with generalized coordination numbers," *Angewandte Chemie - International Edition*, vol. 53, no. 32, 2014.
- 43 F. Calle-Vallejo, D. Loffreda, M. T. M. Koper, and P. Sautet, "Introducing structural sensitivity into adsorption-energy scaling relations by means of coordination numbers," *Nature Chemistry*, vol. 7, p. 403–410, 2015.
- 44 K. Rossi, G. G. Asara, and F. Baletto, "A genomic characterisation of monometallic nanoparticles," *Physical Chemistry Chemical Physics*, vol. 21, no. 9, 2019.
- 45 F. Baletto, "Structural properties of sub-nanometer metallic clusters," *Journal of Physics: Condensed Matter*, vol. 31, p. 113001, jan 2019.
- 46 A. Stukowski, "Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool," *MODELLING AND SIMULATION IN MATERIALS SCIENCE AND ENGINEERING*, vol. 18, JAN 2010.
- 47 J. Greeley, I. Stephens, A. Bandarenka, T. Johansson, H. Hansen, I. Chorkendorff, and J. Norskov, "Alloys of platinum and early transition metals as oxygen reduction electrocatalysts," *Nature chemistry*, vol. 1, p. 552, 2009.
- 48 T. P. Rossi, P. Erhart, and M. Kuisma, "Hot-carrier generation in plasmonic nanoparticles: The importance of atomic structure," *ACS Nano*, vol. 14, p. 9963–9971, 2020.
- 49 C. D. Paola, R. D'Agosta, and F. Baletto, "Geometrical effects on the magnetic properties of nanoparticles," *Nanoletters*, vol. 16, p. 2885, 2016.
- 50 C. Zeni, K. Rossi, A. Glielmo, and S. De Gironcoli, "Compact atomic descriptors enable accurate predictions via linear models," *The Journal of Chemical Physics*, vol. 154, no. 22, p. 224112, 2021.
- 51 R. Drautz, "Atomic cluster expansion for accurate and transferable interatomic potentials," *Phys. Rev. B*, vol. 99, p. 014104, Jan 2019.
- 52 E. Gazzarrini, K. Rossi, and F. Baletto, "Born to be different: The formation process of Cu nanoparticles tunes the size trend of the activity for CO₂to CH₄conversion," *Nanoscale*, vol. 13, no. 11, 2021.
- 53 O. J. F. Martin, C. Girard, and A. Dereux, "Generalized field propagator for electromagnetic scattering and light confinement," *Phys. Rev. Lett.*, vol. 74, pp. 526–529, Jan 1995.
- 54 M. Mavrikakis, B. Hammer, and J. K. Nørskov, "Effect of strain on the reactivity of metal surfaces," *Phys. Rev. Lett.*, vol. 81, pp. 2819–2822, Sep 1998.
- 55 Z. Zhao, Z. Chen, X. Zhang, and G. Lu, "Generalized Surface Coordination Number as an Activity Descriptor for CO₂ Reduction on Cu Surfaces," *Journal of Physical Chemistry C*, vol. 120, no. 49, 2016.
- 56 C. Girard, "Near fields in nanostructures," *Rep. Prog. Phys.*, vol. 68, pp. 1883–1933, jul 2005.
- 57 C. Girard, "pygdm." <https://pypi.org/project/pyGDM2/>, 2021.
- 58 P. R. Wiecha, "pygdm—a python toolkit for full-field electrodynamical simulations and evolutionary optimization of nanostructures," *Comp. Phys. Commun.*, vol. 233, p. 167–192, Dec 2018.
- 59 B. T. Draine, "The Discrete-Dipole Approximation and Its Application to Interstellar Graphite Grains," *American Physical Journal*, vol. 333, p. 848, Oct. 1988.
- 60 P. B. Johnson and R. W. Christy, "Optical constants of the noble metals," *Phys. Rev. B*, vol. 6, pp. 4370–4379, Dec 1972.
- 61 A. D. Rakić, A. B. Djurišić, J. M. Elazar, and M. L. Majewski, "Optical properties of metallic films for vertical-cavity optoelectronic devices," *Appl. Opt.*, vol. 37, pp. 5271–5283, Aug 1998.
- 62 J. H. Weaver, C. G. Olson, and D. W. Lynch, "Optical investigation of the electronic structure of bulk rh and ir," *Phys. Rev. B*, vol. 15, pp. 4115–4118, Apr 1977.
- 63 W. S. M. Werner, K. Glantschnig, and C. Ambrosch-Draxl, "Optical constants and inelastic electron-scattering data for 17 elemental metals," *Journal of Physical and Chemical Reference Data*, vol. 38, no. 4, pp. 1013–1092, 2009.