Journal Name

ARTICLE TYPE

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,*[a] Mirko Vanzan,*[a] Igor Vasiljevic *[a], Alejandro Santana-Bonilla [a], and Francesca Baletto *[a]

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 mono-,bi-, and multi-metallic nanoparticles enables to discriminate which descriptors could be linked with their stability and performance. In this work we introduce a modular and open-source library, which enables to classify the morphological and geometrical characteristics of a given NA through several structural analysis and order parameters previously introduced in the literature. 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[? ?] to drug-delivery,[? ?] memory-storage[? ?] to optics[? ?]. Among the most prominent applications, MNPs further play a significant role as thermal,[? ?] electro-chemical,[? ?] and photo-chemical[? ?] 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.[? ?] 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.[? ?] 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 characterise 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 metallic NPs and NAs. To this end, we present Sapphire, a Python library publicly available at `https://github.com/kcl-tscm/Sapphire`[1], which provides a unified framework to calculate various structural descriptors, and aimed to a fast classification into the main geometries, as icosahedra (Ih), decahedra (Dh) and FCC (both with or without stacking fault). KR: do we have this feature ? if yes, we are not agnostic because we decide the classes. also, why is this useful ?

A number of descriptors[? ?] and representations[? ?] have been adopted in the literature to describe the geometry and chemical ordering of mono-, bi-, and multi-metallic nanoparticles. Also, software and libraries to characterize structural motifs in molecular and periodic systems have been developed .[? ?] Yet, 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. Such tool, to our knowledge, has been absent in so far.

The next section of the manuscript is dedicated to discuss, in greater detail, the philosophy and workflow of Sapphire, focusing primarily on the nature of its hierarchical design, and how

[a] *Address, Address, Town, Country. Fax: XX XXXX XXXX; Tel: XX XXXX XXXX; E-mail: xxxx@aaa.bbb.ccc*

[b] *Address, Address, Town, Country.*

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 00.0000/00000000.

‡ Additional footnotes to the title and authors can be included *e.g.* 'Present address:' or 'These authors contributed equally to this work' as above using the symbols: ‡, §, and ¶. Please place the appropriate symbol next to the author's name and include a `\footnotetext` entry in the the correct place in the list.

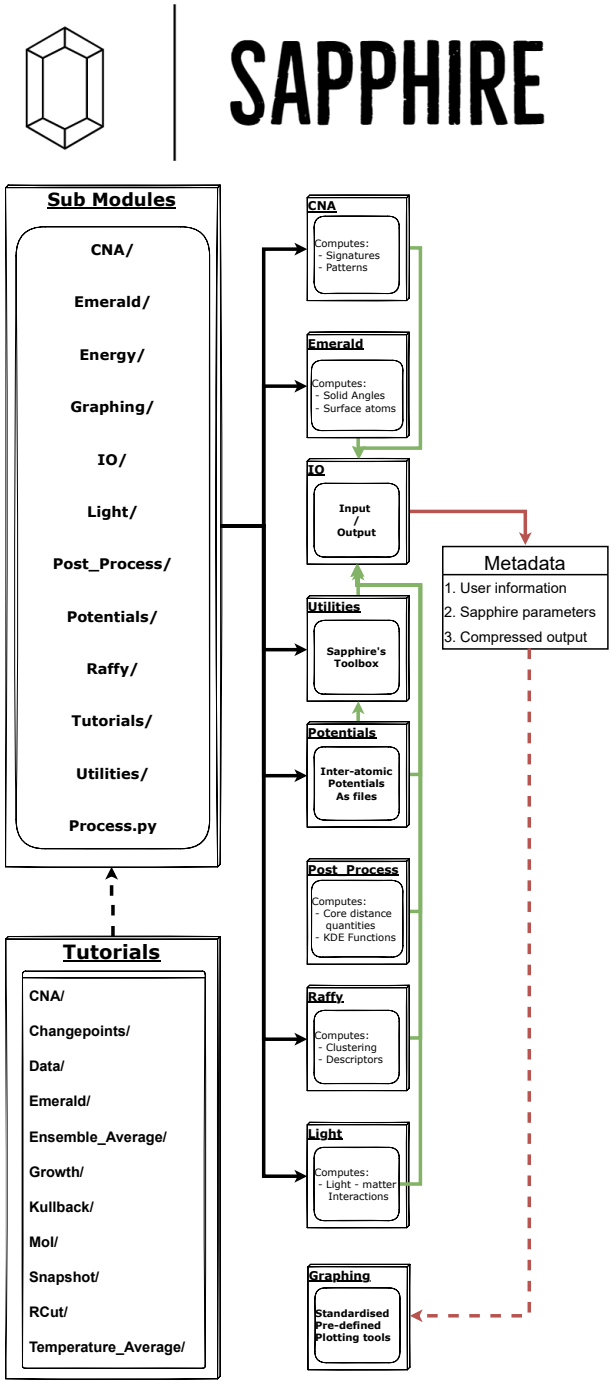

Fig. 1 Flowchart detailing the scheme under which Sapphire is executed. The primary Process class may be found in the root directory which may be called through the use of pre-written user input. Black lines indicate module import directionality. Green lines indicate IO streams. Red lines indicate data flow. Solid lines indicate hard-coded streams. Dashed - user choice.

we have developed the library to be flexible and facile in its use. Successive sections present a didactic introduction to the theory and practical use of each of the characterisation tools currently present in Sapphire.

As paradigmatic examples, we discuss the Sapphire analysis of trajectories sampling the dynamical evolution of Au, Pt, AuPt, and AuRh nanoparticles underdoing structural rearrangements. For each example, a tutorial is available in the dedicated folder `/main/Sapphire/Tutorials/`. These web resources will allow anybody to reproduce the figures in this manuscript exactly, respecting the Findable, Accessible, Interoperable and Re-usable (FAIR) requirements the scientific communities are asked for. We continue to describe the available strategies to obtain 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. To conclude, we provide a retrospective overview of Sapphire capabilities, and an outlook on its ongoing and future development.

2 Workflow

The aim of this section is to briefly discuss Sapphire architecture. 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. To begin analyses one needs only to provide KR: provide ... how ? command line ? calling which program, if any ? the atomic coordinates for the MNP they wish to analyse. We recommend the use of Sapphire only for metallic nanoparticles and nanoalloys, as currently there is no specific tools for systems displaying other chemistries. Given that Sapphire leverages the input / output (IO) stream of `ASE`,[?] many common formats (e.g., *.xyz*, *.exyxyz*, etc. etc.) is compatible with Sapphire. Note, the coordinate file can be obtain 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. In the current version of Sapphire, we in fact facilitate ensemble averaging over sets of molecular dynamics trajectories. Should one wish to, there exists the utility to integrate accompanying data files, e.g., collecting solely information on the energetics of each snapshot in a trajectory.

KR: we are missing a paragraph discussing in detail moment in which any structural analyis takes place. Robert: please write this asap that I'm unable to go further without it. Maaaaybe the bit below could be of help 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 an analysis.

After gathering from the user the set of configurations to be analyzed, we next create, if wished for, an extended xyz file, which is able to also encode the calculable single-atom labels calculated within Sapphire. Following the creation of this extended file, the time series analysis of the trajectory is performed, as described by Fig. 1. KR: in Fig 1 where ? please expand and clarify
KR: also, is this correct ? if the code – 1) reads the .xyz 2) writes the .extxyz 3) does other analysis – it is inefficient, 2 and 3 should happen simultaneously for best efficiency, or by time-series analysis you refer strictly and only to the statistical tests bit ?.
what happens if a single file is given ?
how do we create the .extxyz file with the single atom labels, if

we haven't calculated these yet ?

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.
A crucial feature of Sapphire is its data storage philosophy. Each calculation using Sapphire's `Process`, a metadata object is created in agreement with EMMO[?] and EMMC[?] rules, and the FAIR[?] principles. Indeed `Sapphire` is amongst the first materials modelling post-processing tools to automatically write metadata in a standardised format. Contained within the metadata are the following:

- Compressed forms of the written output to better facilitate a streamlined integration with potential subsequent interactive 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.

We thus hope `Sapphire` will help the community to easily create FAIR databases in their future contributions.

3 Distance distribution Functions

The distribution of inter-atomic distances in a MNP or in a NA is a crucial quantity to characterise the geometry and chemical ordering of an NP. Further, the PDDF is a directly measurable quantity via X-ray techniques.[?]

To define the PDDF of an MNP or a NA, ; let d_{ij} be the inter-atomic distance between atoms i and j:

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

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

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

where d labels the interatomic distance at which the density is estimated; $K\left(d_{ij},d;h\right)$ labels the kernel function over the d_{ij} variable; and the parameter h labels the bandwidth that defines the tightness of the kernel function.

More in detail, KDE assumes that each atomic position 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, the triangular, and the uniform distributions, are also currently supported. Given the analytical form of the KDE we impose, derivatives can be easily calculated, and the approximate location of minima (and maxima) in the distribution can be identified. This is a key utility because PDDF minima characterization is instrumental to the determination of robust nearest (and next nearest, etc.etc.) neighbour distances.

We report in Fig 3 panel x a practical example of the impact of the K and H choice in the PDDF calculation. 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 the 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.

I do not get the numbers in the inset. The good value is 0.05 why here we have -4.5 / -1.3/ -0.5? what does it mean "-"? FB:ALL DISTANCES MUST BE RESCALED BY THE BULK LATTICE or THE AVERAGE BULK LATTICE FOR BINARY. EASIER TO SEE sqrt(2)/2 and 1!!! @ROBERT: ADD HERE a 1415 CuPt at one chemical composition. ADD the PDDF of an amorphous-AuPt liquid droplet fastly quenched down to 300K. ROBERT: I will split the figure in two: panel (g) and (h) should be apart as they show setting the input and not results. Just use AuPt20-80 at 300K

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

- Identification of a cutoff distance (R_{cut}) to discriminate the first shell of neighbours around each atom by determining the location of the first minimum in the PDDF. The subsequent analyses depend on the R_{cut} value as it precipitates down to adjacency matrix compilation, and all the quantities deriving from the latter, e.g., coordination numbers and common neighbour analysis.
- Detection of a well-defined geometry or an amorphous shape, depending on the position of the second peak. Where we define as "amorphous" a structure where there is not a maximum in correspondence of the bulk lattice. @Robert: I will add as an output a string as "Amorphous is detected", if in correspondence of the bulk lattice there is not a maximum. This could also be used in conjunction with the CNA pattern
- Identification of the melting/freezing temperature, which is the one where the peak on correspondence of the bulk lattice disappears/appears.
- KR: the previous point and the one above are in contradiction one another: one can have a solid amorphous structure, a liquid amorphous structure, and a liquid structure... if we define amorphous an object where there is no second peak at bulk lattice – definition with which I'm against – we cannot say that we use the same definition also to say that it is liquid.

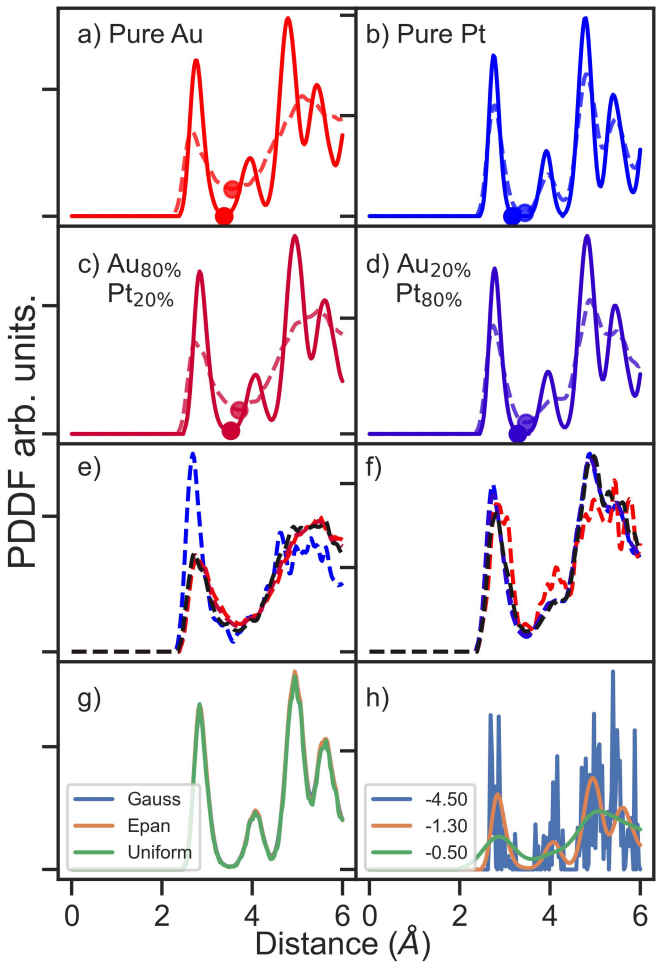

Fig. 2 Example of the PDDF from `Sapphire` for a set of pure and bimetallic NPs of 1415 atoms adopting initially an icosahedral (Ih) shape. (a) and (b) pure Au and pure Pt, respectively. (c) and (d) AuPt randomly alloyed at different chemical compositions, as indicated in the inset. Solid lines correspond to the PDDF of the system at 300 K while dashed lines are taken at 1000 K. Dots represent the position of the first PDDF minimum, corresponding to R_{cut}. (c) in orange and (d) in green. Panels e) and f) PDDF indentifying the chemical species: from homo-pairing Au-Au (red); from Pt-Pt (blue); heterogeneous bond Au-Pt (black), against the full system (orange and green respectively). Panels g) refers to the cold temperature PDDF of $Au_{80}Pt_{20}$ (panel c) and shows the weak dependence on the cthe kernel. h) refers to the case in panel c and reports the dependence on the h bandwidth parameter.

To clarify further this points, we report in Figure panel X and analyse in detail the PDDF of pure MNP of Au and Pt, and bimetallic NAs of AuPt and CuPt sampled over molecular dynamics runs where the system temperature is iteratively increased by 50 K each 1 ns, from 300 K to 1000 K.[?]

We show cases at cold and high temperature, so it is clear the effect of broadening of the first peak and the clear change in the position of the second peak, which signals a phase-change transition[2,3] organize the discussion better, state in which system the shift is clear. state in which system this is not. state why this is the case –> Pt has a larger melting temperature than Au and a Pt-rich alloy will preserve its stability In the case of AuPt there is a mild indication that the Pt-subcluster remains solid at higher temperatures than the Au one, as indicate in the appearance of the second peak in panel (f) at the bulk lattice. Instead, the position of the first peak is not affected by the temperature suggesting, for both mono- and bi- nanoparticles, the absence of any significant thermal dilatation.

3.1 Radial distribution functions

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 experiment.[?]

The radial distribution function is counting the number of atoms falling in concentric shells from the centre of mass of the nanoparticle,

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

where the coordinates $\hat{x}_\alpha, \hat{y}_\alpha, \hat{z}_\alpha$ of the atom-i and chemical species α are rescaled w.r.t the centre of mass of the whole nanoparticle. Again a binning procedure takes place. Here we need to define the shell-width s, which is related to the interlayer distance in the bulk. A reasonable value is about 0.2-0.3 the bulk lattice distance. Pls provide a reasonable value for that. from Bragg, the distance between (111) plane is a_{bulk}/sqrt(3). I took half of this value. WORKS ALWAYS IN TERMS OF THE LATTICE BULK!!!!

For nanoalloys, `Sapphire` provides three three radial distribution: how the atoms are distributed from the centre of mass of the whole system; distribution of atoms-A from their centre of mass; and similarly for B-atoms.

@ROBERT: just add a Figure for AuPt same composition but eventually different ordering to show the three different radial distribution. KR please put a figure here !!!

Still with a focus on binary NAs, it might be worthy to compute the distance between the centre of mass of the subclusters containing only atoms of type A and B, respectively:

$$\Delta R_{COM} = \sum_{i\in A} r_i/N_A - \sum_{i\in B} r_i/N_B \tag{4}$$

where $r_{i\in A(B)}$ is the radial distance of atoms of species A(B) from the centre of mass of the whole NA. $N_A(B)$ is the number of atoms A(B) in the NA.

4 Adjacency matrix based descriptors

The definition and characterization of nearest neighbour networks has been largely adopted to classify nanoparticles morphology and draw structure-property relationships.[? ? ?] To evaluate nearest neighbour networks and quantities deriving from the latter, let the adjacency matrix **A** be defined as:

$$\mathbf{A}\left(r_{ij}\right) = \begin{cases} 1, & \text{if } r_{ij} \le R_{cut} \\ 0, & \text{if } r_{ij} > R_{cut} \end{cases} \tag{5}$$

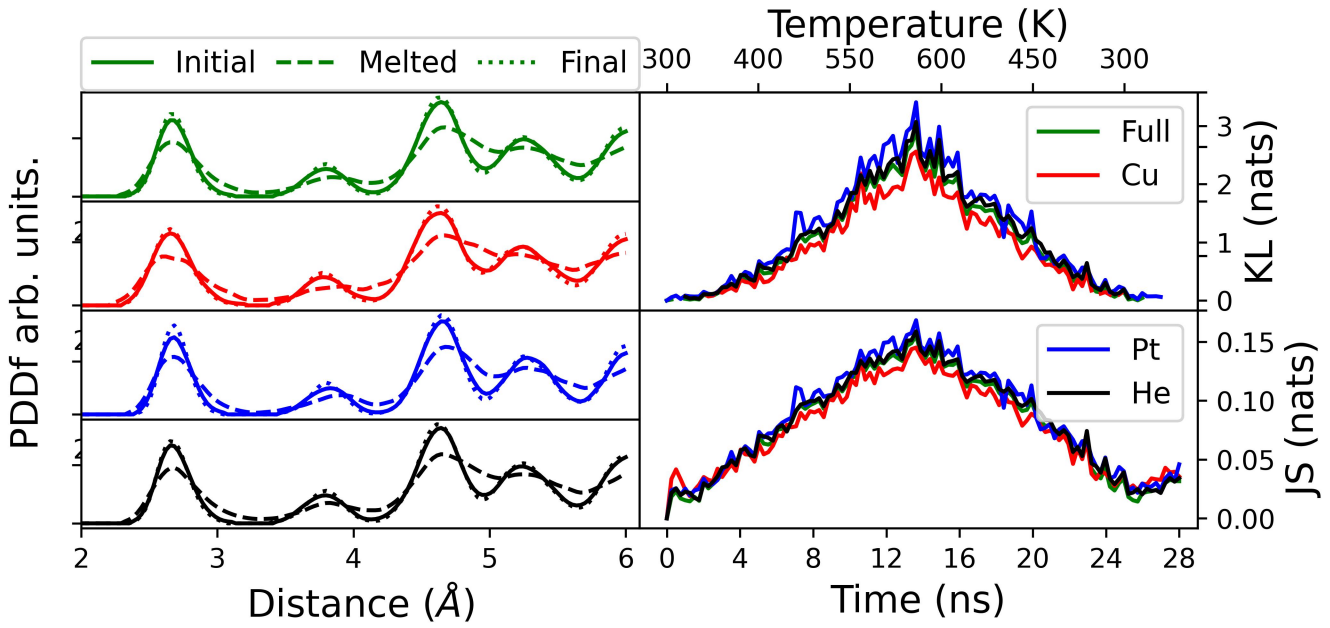

Fig. 3

4.1 Coordination Numbers

In the field of heterogeneous catalysis, the coordination of an adsorption site has been often used as a descriptor to rationalize its activity. Indeed simple chemical intuition, low-coordinated atoms are more likely to form stronger bonds than highly-coordinated ones. The coordination number of an atom i nominal is then defined as:

$$CN_j = \sum_{i \neq j} \mathbf{A}\left(r_{ij}\right) \tag{6}$$

We have also tested that the coordination number from Eq.6 and the Van Meel's algorithm based on the solid angle provide similar results. The nominal CN definition and the one from the van Meel's algorithm agree well but the latter is more robust in identifying the CN even of core atoms in very defected structures. Furthermore the van Meel's do not require any definition of cut-off. Rob this is in Emerald – add AuCu and AuPt Igor, send the instruction to Robert. Write a sentence and the reference for it

For the case of MNPs, a function of the coordination number of surface sites (i.e., the generalised coordination number[4,5]) can show a robust linear relationship with the adsorption energy of small molecules (e.g., O, CO, OH). The atop surface site generalised coordination number (aGCN) is defined as :[? ? ?]

$$aGCN_i = \sum_j \frac{CN_j}{CN_{max}} \tag{7}$$

with CN_{max} set to 12, as this is the maximal coordination of an FCC atom in the bulk. Furthermore, the aGCN is found to be a more robust descriptor than the nominal coordination, to characterize different nanoparticle surface sites, avoiding the too basic classification into face, edge, vertex. It could also be used to estimate the surface area in a more precise manner, more accurate than the spherical approximation and able to reproduce the analytical results in the case of geometrical built shapes[? ?] here I will take an Ih and a Dh of 1415 atoms and I will calculate the area using the A-agcn formula and geometry – see F.H Kaatz A. Bultheel , Nanoscale Research Letters 14, 150 (2019) Figure restrict to 3-11 aGCN. Figure **??** reports the distribution of sites during a coalescence of Au and Pt, with Pt half the size of Au-NP and same size of Au-NP. In the former, the formation of low coordinated sites is much higher than in the latter where highly coordinated sites prevails.

KR please put a figure here !!!

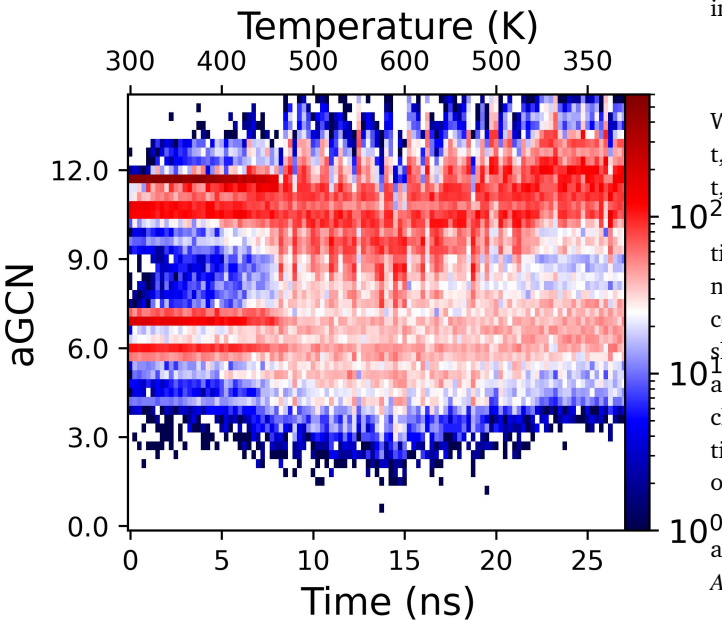

Fig. 4 Heatmap for the aGCN distribution of an $Au_{1149}Pt_{266}$.

@ROBERT: Plot where the mixing parameter over time is reported for the case of the system i mentioned in the RDF subsection. Just a AuPt alloyed and the Michey Mouse AuPt

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 **?**.

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:

$$AA_{ij}(\Delta t) = |A_{ij}(t+\Delta t)A_{ij}(t)| \tag{8}$$

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

$$R(t,t+\Delta t) = \sum_i R_i(t,t+\Delta t) \tag{9}$$

$$R_i(t,t+\Delta t) = \sum_{j\neq i} AA_{ij}(\Delta t) \tag{10}$$

To measure the collectivity 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 NP:

$$H = \frac{\sum_i \Theta(R_i(t,t+\Delta t))}{N} \tag{11}$$

Where Θ labels an Heaviside step function and N the number of atoms in the NP. By definition, the magnitude of H can vary between 0 and 1. The concertedness of a rerrangement, C, is then defined as the the change in the number of atoms involved in the process between t Δ t and t, and between t and t + Δ t:

$$C(t\Delta t,t,t+\Delta t) = |H(t\Delta t,t)H(t,t+\Delta t)| \tag{12}$$

When all the atoms in the NP change their local connectivitytime t, but the latter remains stable at the successive one t +Δ t, C(t Δ t, t, t +Δ 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 Δt. 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, Δt = 10 ps, which is consistent with the timescale of adatoms diffusion on low Miller index surfaces, one of the fastest rearrangement process in NP structural transitions. All the quantities discussed above can be readily modified to account for the presence of multiple chemical species, $chem = AA, BB, AB$, in the NP:

$$R^{chem}(t,t+\Delta t) = \sum_i R_i^{chem}(t,t+\Delta t) \tag{13}$$

where R_i^{chem} is given by

$$R_i^{chem} = \sum_{j\neq i} AA_{ij}^{chem}(\delta t) \tag{14}$$

$$C^{chem}(t\Delta t,t,t+\Delta t) = |H^{chem}(t\Delta t,t)H^{chem}(t,t+)|. \tag{15}$$

KR please put a figure here !!! we then use this text (paraphrased) below to explain it: Single-step mechanisms involving only a sector of the cluster, or multi-step processes are characterized by lower values of C(t Δ t, t, t +Δ t), while continuous atomic rearrangements display C $\sim$ 0.

4.3 Common Neighbour Analysis signatures and patterns

both total number and percentages are relevant

In this subsection, we shall illustrate the geometrical environments observed for bulk and surface atoms. To this end, we first evaluate all of the CNA signatures attributed to each individual atom in the system. 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. We then define the CNA pattern as a list of each appearing signature and the number of times it is detected.

patterns to identify surface and bulk features are common to clusters of different shapes. All of the patterns observed in Fig. **??** have been described in Fig. **??** and table **??** in the exact same context.

Given the evident sensitivity to local geometry, the next question to ask is how sensitive are these patterns to imperfections in the crystalline structure? Especially at the surface where such features are likely to have a profound impact on the catalytic performance of an NP. To probe this consideration, we have systematically deformed the surface of a Marks decahedron and observed

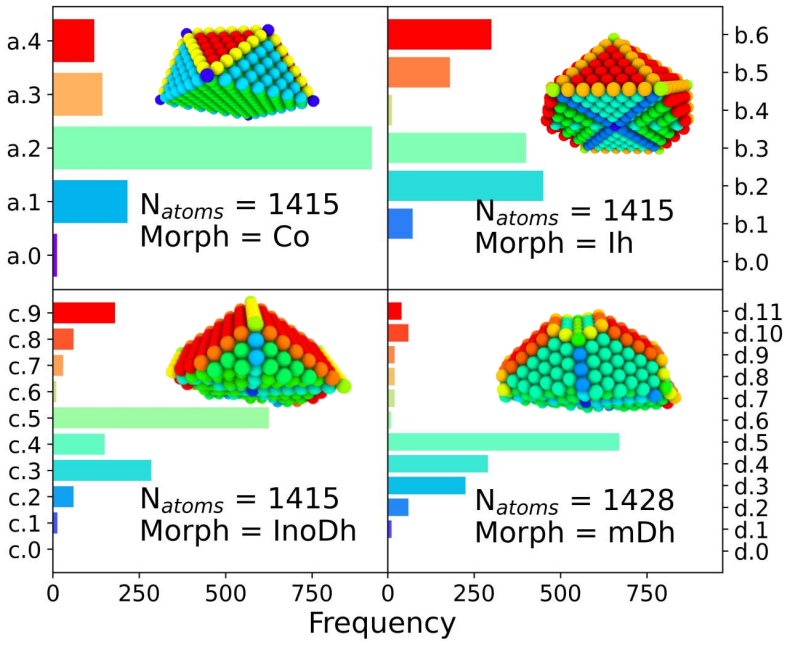

Fig. 5 Frequency chart of regularly occurring CNA Patterns in regular NA geometries. On the y axis are the subplot letters followed by an index for cross-referencing with tab. 1. Geometrical conformation responsible for each pattern fingerprint included in each insert. KR: here i would suggest the following: each of the four panel is composed of three elements: 1) the snapshot of the full nanoparticle with atoms all in gold colour. Number of atoms and morphology appears below. 2) the slice of the nanoparticle colour coded according to the CNA 3) the histogram plot .

Table 1 Definition of regularly observed CNA Patterns and descriptions of the local environment in which they are found. I need to rewrite some of these descriptions.RMJ

Pattern	Description	Found in Fig. 5	Surface
$[4(211),4(421)]$	(100) Facet.	a.1, c.9, d.11	Yes
$[3(211),2(311),2(421)]$	Boundary between (100) and (111) facets.	a.3, d.10	Yes
$[2(100)2(211)2(422)]$	Boundary between (100) and (111) facets.	c.8	Yes
$[1(200),2(211),2(311),1(421)]$	Corners of box created by the above patterns	c.7, d.9	Yes
$[(2,(2,0,0)),(4,(3,1,1)),(1,(4,2,1))]$	Vertex connecting re-entrance with five-fold symmetry axes.	d.8	Yes
$[2(300),4(311),2(421),2(422)]$	Re-entrance centre bounded on either side by (111) facets.	d.7	yes/no
$[(1,(1,0,0)),(2,(2,1,1)),(1,(3,2,2)),(1,(4,2,2))]$	Re-entrance centre bounded on either side by (111) facets.	c.6	yes
$[2(200),1(300),2(311),1(322),1(422)]$	Vertex connecting re-entrance with five-fold symmetry axes.	d.6	Yes
$[12(421)]$	Bulk face centred cubic (FCC) structure.	a.2, b.3, c.5, d.5	No
$[6(311),3(421)]$	(111) Facet.	a.4, b.6, c.4, d.4	Yes
$[6(421),6(422)]$	Bulk hexagonal close packed (HCP) structure.	b.2, c.3, d.3	No
$[4(311),2(322),2(422)]$	Five-fold symmetry axes separating (111) facets	b.5, c.2, d.2	Yes
$[10(422),2(555)]$	HCP twinning planes	b.1, c.1, d.1	No
$[5(322),1(555)]$	Tip of the five-fold symmetry axes	b.4, c.0, d.0	Yes
$[12(555)]$	Centre of an icosahedron	b.0	No

the resultant environment and emergent CNA patterns.

"classification" ???

to show different cases mono and bimet and also to see how the classification works. I mean do we have an output which said morphology belongs to? Use always the Mickey Au-Pt (initial) and the initial coalesce Pt onto Au and any AuRh shell from Mirko to show the tools.

5 Surface Identification

gets atoms at surfaces according to a solid angle criterion

Robert add any consideration on estimate and counting of surface atoms fran will add the bit on solid angle: yes looking forward to do it on Thursday Any figure you feel may be instructive is more than welcome.

Another way to recognise and categorise surface atoms is offered by a clustering approach based on local atomic environment descriptors. This approach, first employed in Ref.[?] to clas-

sify atoms in Au nanoparticles, and implemented in the RAFFY Python package[6], 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 expansion[7] (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.[?] 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. One also has the option to save a new form of the metadata created by considering the averages over multiple simulations.

KR: here you need to explain the physics and not just write down the commands. Things to be discussed:

"Average over multiple independent simulations" –> is this an arithmetic average ? a boltzmann weighted average ? are error bars calculated ? are they reported in the graph ? does it always make sense to look at averaged quantities for multiple simulation ?

"Average over time" –> same questions as above... averaging over a trajectory makes sense only if the ensemble we are sampling is constant over time and the trajectory is equilibrated.

KR please put a figure here !!!

7 From structure characterization to structure-performance relationship

The Sapphire library is focused on the characterization of MNPs and NAs architecture (i.e., morphology and chemical ordering). Sapphire can be used in synergy with available codes geared toward the prediction of chemophysical properties of MNPs and NAs. In this section we briefly discuss two examples for the case of MNPs with an application in the electrochemical conversion of small molecules, and for NAs with unique optical properties.

7.1 Electrocatalytic properties via microkinetic models

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 $j_{\text{nanoparticle}}$(t,T,U) produced by a nanoparticle for such an electrochemical process can be written as:

$$j_{\text{nanoparticle}}(t,T,U) \quad = \quad \sum_{\alpha} \mathscr{C} \xi(t,T) \alpha e^{\beta \Delta G(U,T,\alpha)} \,, \qquad (16)$$

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 $F_{nanoparticle}$ to the total number of sites available, N_{site}, and the sum runs over the collection of the non equivalent sites appearing in the NP under consideration. The latter are also characterized by their α values. Note that in 16 the effect of the potential is treated as an a posteriori correction, in line with the computational hydrogen electrode model.[?]

The atop generalized coordination number (See Sec **??**) of a surface site is an accurate descriptor to predict the activity of a metallic adsorption site (as demonstrated for Au, Cu, Pt) catalysing the electrochemical oxygen or carbon dioxide reduction.[4,8? ? ? ? ,9] Assuming a Volcano-plot relationship, the reaction free energy ΔG and the GCN are related by the following general expression:

$$\Delta G = \begin{cases} +a1 \ \text{aGCN}_n - b1 & \text{if } \ \text{aGCN}_n < aGCN^{volcano-peak} \\ -a2 \ \text{aGCN}_n + b2 & \text{if } \ \text{aGCN}_n > aGCN^{volcano-peak} \end{cases} . \qquad (17)$$

where the coefficients a1, b1, a2, b2, and the value of the $aGCN^{volcano-peak}$ can be found from, e.g., a small set of DFT calculations. For the case of ORR on Pt, for example:[? ?]

$$\Delta G = \begin{cases} +a1 \ \text{aGCN}_n - b1 & \text{if } \ \text{aGCN}_n < aGCN^{volcano-peak} \\ -a2 \ \text{aGCN}_n + b2 & \text{if } \ \text{aGCN}_n > aGCN^{volcano-peak} \end{cases} . \qquad (18)$$

One can thus utilize Equation 16 and Equation 18 to ...

KR please put a figure here !!!

KR: cite E. Gazzarrini github for the code

7.2 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),[10,11] as implemented in the pyGDM code.[12,13] 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

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

$$\sigma_{ext}(\omega) = \frac{8\pi^2}{n_{env}\lambda_0} \sum_{j}^{N_{cells}} \Im\left(\mathbf{E}^*_{0,j} \cdot \mathbf{P}_j\right), \tag{19}$$

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 field.[14] In the described method, the near field may be found by self-consistently solving

$$\mathbf{E}(\mathbf{r}_i,\omega) = \mathbf{E}_0(\mathbf{r}_i,\omega) + \sum_{j}^{N_{Cells}} G^{EE}_{tot}(\mathbf{r}_i,\mathbf{r}_j,\omega)\,\chi\mathbf{E}'(\mathbf{r}_j,\omega)\,V_{cell},$$

where $G^{EE}_{tot}(\mathbf{r}_i,\mathbf{r}_j,\omega)$ 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, $\mathbf{P}_j = V_{cell}\chi_j \cdot \mathbf{E}_j$, 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. 19. All metallic dielectric parameters used in these calculations are provided by Johnson and Christie.[15] The interested reader is referred to previous literature for further detail[10,12,13].

KR: Figure 6 reports... COMMENT ON THE FIGURE !!!! are the extinction spectra peak always at the same wavelength ? why is that so ? is their number constant ?

Because of the approximations in the proposed method, we note that correlated behaviour that may arise from quantum many-body effects is neglected. Further, when considering structures with size ~ 1 nm illuminated in the ultraviolet–visible–near-infrared, the incident field will only weakly couple to the structure and result in non-trivial internal field enhancement.

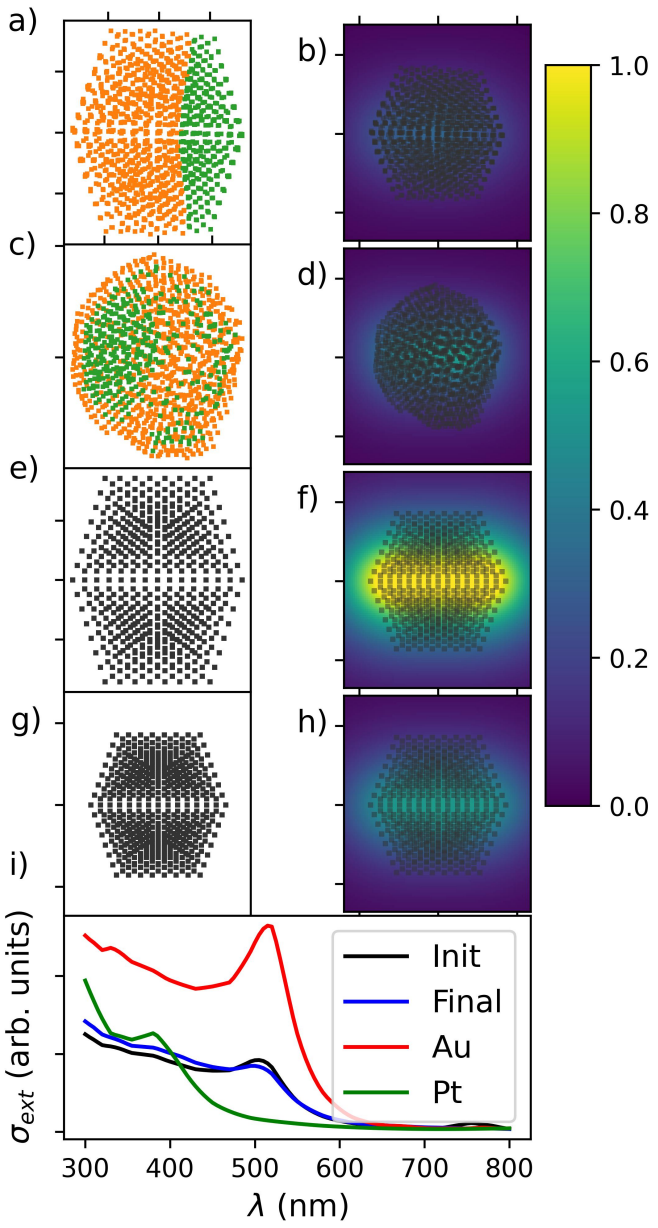

Fig. 6 An illustration of the interaction between Sapphire and pyGDM. a) The initial structure consisting of Au (orange), and Pt (green). b) Near field enhancement of the structure when illuminated by plane waves at 550 nm. c) Photo-extinction spectra as computed via eqn 19. Initial structure is identical to that in Fig. **??**.b, $Au^{Ih}_{147}Pt^{Ih}_{147}$ Janus NA.

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 a nanoparticle, not necessarily well-characterized by

broader scope commercial softwares, in so far. In the long term, we hope Sapphire serves to create a standardised, and 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

The Acknowledgements come at the end of an article after Conflicts of interest and before the Notes and references.

Notes and references

1 R. M. Jones, "Sapphire." https://github.com/kcl-tscm/Sapphire.git, 2022.

2 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.

3 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.

4 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.

5 K. Rossi, G. G. Asara, and F. Baletto, "A genomic characterisation of monometallic nanoparticles," *Physical Chemistry Chemical Physics*, vol. 21, no. 9, 2019.

6 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.

7 R. Drautz, "Atomic cluster expansion for accurate and transferable interatomic potentials," *Phys. Rev. B*, vol. 99, p. 014104, Jan 2019.

8 G. G. Asara, L. O. Paz-Borbón, and F. Baletto, ""Get in Touch and Keep in Contact": Interface Effect on the Oxygen Reduction Reaction (ORR) Activity for Supported PtNi Nanoparticles," *ACS Catalysis*, vol. 6, no. 7, 2016.

9 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.

10 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.

11 C. Girard, "Near fields in nanostructures," *Rep. Prog. Phys.*, vol. 68, pp. 1883–1933, jul 2005.

12 C. Girard, "pygdm." https://pypi.org/project/pyGDM2/, 2021.

13 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.

14 B. T. Draine, "The Discrete-Dipole Approximation and Its Application to Interstellar Graphite Grains," *American Physical Journal*, vol. 333, p. 848, Oct. 1988.

15 P. B. Johnson and R. W. Christy, "Optical constants of the noble metals," *Phys. Rev. B*, vol. 6, pp. 4370–4379, Dec 1972.