



**HAL**  
open science

# Parameter-free coordination numbers for solutions and interfaces

Ruben Staub, Stephan N. Steinmann

► **To cite this version:**

Ruben Staub, Stephan N. Steinmann. Parameter-free coordination numbers for solutions and interfaces. The Journal of Chemical Physics, 2020, 152 (2), pp.024124. 10.1063/1.5135696 . hal-02917716

**HAL Id: hal-02917716**

**<https://hal.science/hal-02917716v1>**

Submitted on 19 Aug 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Parameter-Free Coordination Numbers for Solutions and Interfaces

Ruben Staub<sup>1</sup> and Stephan N. Steinmann<sup>1, a)</sup>

*Univ Lyon, Ecole Normale Supérieure de Lyon, CNRS Université Lyon 1,  
Laboratoire de Chimie UMR 5182, 46 allée d'Italie, F-69364, LYON,  
France*

(Dated: 21 December 2019)

Coordination numbers are among the central quantities to describe the local environment of atoms and are thus used in various applications such as structure analysis, fingerprints and parameters. Yet, there is no consensus regarding a practical algorithm, and many proposed methods are designed for specific systems. In this work, we propose a scale-free and parameter-free algorithm for nearest neighbor identification. This algorithm extends the powerful Solid-Angle based Nearest-Neighbor (SANN) framework to explicitly include local anisotropy. As such, our Anisotropically corrected Solid-Angle based Nearest-Neighbor (ASANN) algorithm provides with a fast, robust and adaptive method for computing coordination numbers. The ASANN algorithm is applied to flat and corrugated metallic surfaces to demonstrate that the expected coordination numbers are retrieved without the need for any system-specific adjustments. The same applies to the description of the coordination numbers of metal atoms in AuCu nano-particles and we show that ASANN based coordination numbers are well adapted for automatically counting neighbors and the establishment of cluster expansions. Analysis of classical molecular dynamics simulations of an electrified graphite electrode reveals a strong link between the coordination number of Cs<sup>+</sup> ions and their position within the double layer, a relation that is absent for Na<sup>+</sup>, which keeps its first solvation shell even close to the electrode.

PACS numbers: 68.08.De

---

<sup>a)</sup>Electronic mail: [stephan.steinmann@ens-lyon.fr](mailto:stephan.steinmann@ens-lyon.fr)

## I. INTRODUCTION

Detailed analysis of chemical systems allows to gain insight into their working mechanisms, to rationalize their properties and sometimes even to predict the impact of chemical modifications such as functionalization of molecules or alloying of surfaces. Roughly two big categories can be distinguished: energetic and structural characterizations. While energy decomposition analysis provides detailed energetic insight in molecular chemistry, solid interfaces and in liquid phase,<sup>1-4</sup> it is the coordination environment that commonly defines the building block for local structural interpretations. The main applications in chemistry involve hydration numbers of ions<sup>5,6</sup>, which are of particular interest in the context of biochemistry,<sup>7</sup> the link between structure and activity in heterogeneous catalysis<sup>8,9</sup> and the characterization of alloys.<sup>10,11</sup>

Intuitively, the coordination number of a particle is defined as the number of nearest neighboring particles. However, this does not provide with a rigorous definition: what is neighboring after all? Chemists tend to link neighboring to the idea of bonding: two items are connected if there is a strong interaction between them. But again, “strong” is not defined and most of the time the strength of the interaction is not easily accessible. Therefore, it is far more common and convenient to think in terms of spatial proximity, echoing the initial terminology of “neighbors”.

In chemistry and physics, many different algorithms are used to define the connectivity between atoms, using more or less additional parameters (e.g., nature of the atoms or reference bond lengths). We propose to classify the algorithms based on their input, using only two criteria: (i) Usage of parameters (e.g., particles nature, reference bond lengths). Any algorithm that depends on more than a set of point coordinates falls into this category and is denoted with a  $P^1$ , while parameter-free algorithms are denoted by  $P^0$ . (ii) Local adaptivity. By this we mean that the connectivity between two particles depends on the presence of other surrounding particles and denote it by  $A^1$ , while the absence of local adaptivity corresponds to  $A^0$ . Since these two criteria are non-exclusive, Figure 1 provides relevant examples of algorithms for the four possible combinations.

The simplest coordination number algorithm is based on tabulated “typical” bond lengths: only if the atoms are closer, they are connected. Hence, it falls into the class  $P^1A^0$ . This simple and rapid algorithm is most popular in visualization softwares,<sup>12</sup> where

Criteria	Locally-adaptive (A <sup>1</sup> )	Locally-non-adaptive (A <sup>0</sup> )
Parameter-free (P <sup>0</sup> )	Voronoi SANN ASANN	Radial distribution analysis
Parameter-dependent (P <sup>1</sup> )	RAD	Reference cut-off

FIG. 1: Overview of available coordination number algorithms. We report only popular algorithms of relevance for the context of the ASANN algorithm development, with the references provided in the text.

bond length cut-offs are typically constructed from predefined covalent radii<sup>13</sup> of considered elements.

Instead of using predetermined typical/average bond lengths that might not be of relevance for the considered sample, coordination cut-offs can be estimated from the sample itself. This led to the development of multiple parameter-free methods<sup>7,10,11,14–19</sup> using a cut-off based on the radial distribution function  $g(r)$ . However, such algorithms require enough data to perform meaningful radial distribution analysis. Finally, these algorithms have an inherently mean-field description of coordination since they use averaged cut-offs. Hence, local fluctuations are not necessarily well handled as these algorithms do not take the remaining neighboring particles into account, i.e., it is a  $P^0A^0$  algorithm.

Constructing locally adaptive coordination numbers requires an analysis of the surroundings of each possible connection, which necessarily increases the computational cost. The best known algorithm to extract context-dependent coordination numbers uses a Voronoi tessellation,<sup>18,20</sup> where space is decomposed into regions (polyhedra for 3D structures), based on which particle is the closest. Two particles are connected if their corresponding regions share a face. This algorithm is parameter-free and thus falls into the  $P^0A^1$  category, which means it can be applied without any additional knowledge on the set of particles considered. However, it has been shown that Voronoi-based coordination numbers have a tendency to be overestimated,<sup>18–21</sup> which can be alleviated by re-weighting.<sup>11</sup> Furthermore, 3D Voronoi

tesselations are computationally expensive<sup>22</sup>. In the same  $P^0A^1$  category, the Solid-Angle based Nearest-Neighbor algorithm (SANN) algorithm<sup>21</sup> was developed as an efficient alternative to Voronoi tessellations. SANN attributes a solid angle to each possible neighbor and iteratively increases the cut-off radius to reach a sum of  $4\pi$  for the solid angles, corresponding to a fully surrounded coordination environment. This elegant method is faster than the Voronoi tessellation and, furthermore, reduces the overestimation of the coordination numbers, without the introduction of any parameters<sup>21</sup>. However, SANN suffers from coordination number overestimation at interfaces<sup>23</sup>.

Recently, the Relative Angular Distance (RAD) algorithm<sup>23</sup> was developed to amend the SANN overestimation at interfaces. The reduction of the coordination number is achieved by accounting for blocking: the coordination of two particles can be obstructed by particles between the considered pair. Compared to other available blocking-based methods<sup>24,25</sup> RAD does not require predefined cut-off distances, but still requires additional parameters, so that the algorithm falls in the  $P^1A^1$  class. An earlier  $P^0A^1$  algorithm<sup>26</sup> accounted for blocking by considering only bonds that are not intersecting a triangle formed by shortest bonds. However, this idea was not further explored as its computational cost is expected to be large, and we found the method not suitable to describe surfaces.

Herein we propose to augment the SANN algorithm with a correction term that takes the local anisotropy into account, while preserving all properties of SANN, i.e., low computational cost, no parameters, being locally adaptive ( $P^0A^1$  class) and being scale free, i.e., an isotropic scaling of all coordinates does not change the coordination number. This scheme is called Anisotropically corrected Solid-Angle based Nearest-Neighbors (ASANN).

To set the stage, we introduce the SANN algorithm in section II A. Then, in section II B, we identify the SANN isotropic description as source of coordination number overestimation and propose a natural solution to tackle local anisotropy through ASANN. After the computational details in section III, section IV contains tests and applications of ASANN for characterizing metal surfaces, alkali metal ion coordination at an electrified interface and for the construction of model Hamiltonians for alloy nanoparticles.

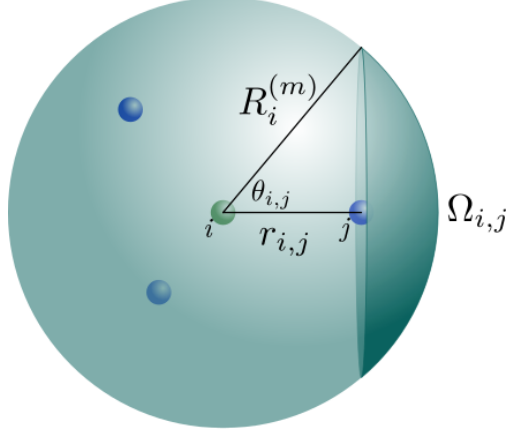


FIG. 2: Schematic representation of the solid angle definition  $\Omega_{i,j} = 2\pi(1 - \cos(\theta_{i,j}))$  associated with neighbor  $j$  at distance  $r_{i,j}$  of central particle  $i$ , and using a coordination radius of  $R_i^{(m)}$ .

## II. ALGORITHMS

### A. The SANN algorithm

The reader is strongly encouraged to read the original SANN article<sup>21</sup> for details. Nevertheless, the main points are outlined below to make the ASANN algorithm understandable.

SANN defines a coordination number  $\text{CN}_{\text{SANN}}(i)$  of particle  $i$  to be the number of the  $m$  nearest neighbors found within the individual coordination sphere of radius  $R_i^{(m)}$  (also called shell radius). SANN elegantly provides a self-consistent definition of  $m$  and the associated coordination sphere radius  $R_i^{(m)}$  as detailed below.

Let  $r_{i,j}$  be the distance between particle  $i$  and its  $j$ -th nearest neighbor (denoted simply as  $j$  in the following). By definition of  $R_i^{(m)}$  and  $m$ , one can already write the inequality:

$$r_{i,m} \leq R_i^{(m)} < r_{i,m+1} \quad (1)$$

Within SANN, the  $m$  nearest neighbors are defined such that the sum of their solid angles (with respect to  $i$ ) is equal to  $4\pi$ . This requirement is related to the intuitive idea that the nearest neighbors of  $i$  are the smallest set of particles fully covering its field of view. Indeed, each nearest neighbor  $j$  of particle  $i$  is associated with a solid angle  $\Omega_{i,j}$ , which is defined by  $\Omega_{i,j} = 2\pi(1 - \cos(\theta_{i,j}))$ , where the angle  $\theta_{i,j}$  is depicted on Figure 2.

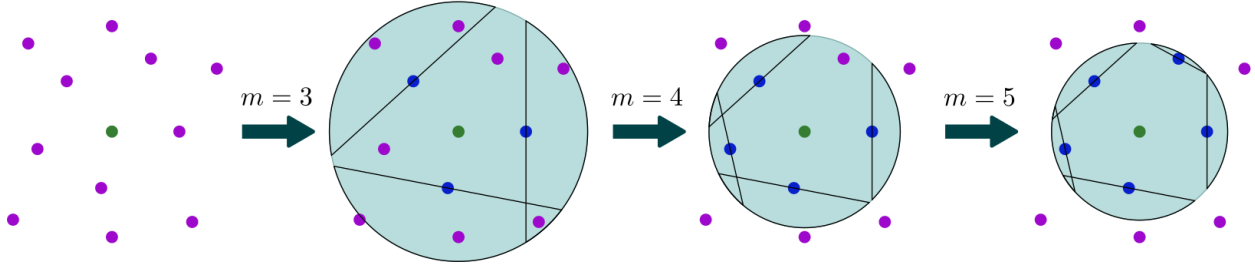


FIG. 3: Illustration of the SANN iterative algorithm in 2D. Blue dots represents neighbors of the considered particle (green dot), while other particles are displayed in purple. For each increasing number of neighbors  $m$  (starting from 3), a coordination sphere radius is computed such that the SANN equality is respected (sum of neighboring solid angle requirement). The algorithm continues until the coordination radius is well-defined (i.e. surrounding particles are considered neighbors if and only if they are within the coordination radius).

The above definitions are summarized in the following equation:

$$4\pi = \sum_{j=1}^m \Omega_{i,j} = \sum_{j=1}^m 2\pi(1 - \cos(\theta_{i,j})) = \sum_{j=1}^m 2\pi \left( 1 - \frac{r_{i,j}}{R_i^{(m)}} \right) \quad (2)$$

One can rewrite the two fundamental Eqs 1 and 2 into a much simpler form:

$$R_i^{(m)} = \frac{\sum_{j=1}^m r_{i,j}}{m-2} < r_{i,m+1} \quad (3)$$

The inequality of Eq. 3 allows to easily compute the  $\text{CN}_{\text{SANN}}(i) = m$  by iteratively increasing  $R_i^{(m)}$  from  $m = 3$  until the inequality is respected (see Figure 3). The convergence of the algorithm was mathematically proven<sup>21</sup>.

SANN is scale free, i.e., a uniform scaling of all coordinates will not affect the coordination number. This property can have counter-intuitive consequences when the analyzed phase features significant density fluctuations. This is particularly problematic when applied to the gas-liquid interface, where thermal fluctuations might create nearly isolated particles. The scale-free algorithm will assign a significantly larger coordination sphere to such an isolated particle, which potentially leads to a higher coordination number compared to particles in dense regions. However, such problematic cases can be spotted by monitoring the coordination sphere radius.

Furthermore, in situations where the environment of  $j$  is more dense than the one of  $i$ , it is possible that  $j$  is considered a neighbor of  $i$ , even though  $i$  is not considered a neighbor of  $j$ . Further along this line, we note that the solid-angle based approach behind SANN is inherently linked to the assumption of locally identical coordination radii. This hypothesis is the only meaningful approach within a parameter-free approach and represents a first order expansion for the coordination sphere. As a consequence, directional bonding (e.g., in organic molecules) which leads to locally more compact geometries is can lead to an overestimation of the coordination number compared to conventional views.

## B. Anisotropically corrected Solid-Angle based Nearest-Neighbors, ASANN algorithm

The SANN algorithm implicitly uses a locally isotropic description of the coordination shell, i.e., the angular distribution of neighboring particles is not taken into account. At surfaces or tips there are, however, simply no neighbors in some directions, leading to empty sections in the coordination sphere. Therefore, the SANN radius must expand to reach more neighbors in the non-empty directions in order to compensate and reach a total sum of  $4\pi$ . Hence, one observes an over-estimation of coordination numbers when neighboring particles are not isotropically distributed.<sup>23</sup> This issue was already reported in the literature, and led to the development of the relative angular distance (RAD) algorithm. However, RAD requires the particles sizes as input parameters, and is fundamentally different from SANN. Here, we propose an Anisotropically corrected Solid-Angle based Nearest neighbors (ASANN) algorithm, which is based on the SANN framework, but includes a parameter-free correction term to correctly describe intrinsically anisotropic coordination shells.

The basic idea of the ASANN algorithm is that some directions are more relevant for coordination than others. Therefore, instead of requiring the neighbors to fill up the whole field of view, ASANN only requires to fill up the relevant neighboring region of space (see Figure 4).

An estimator of the first order local anisotropy is obtained based on the barycentre. Specifically, we define the SANN-based barycentre  $\vec{G}_i^{(m)}$  as the position barycentre over all SANN-coordinated neighbors of  $i$ , weighted by their SANN-based solid angle  $\Omega_{i,j}$  (with respect to  $i$ ). A similar solid-angle based neighbors weighting has been proposed in the



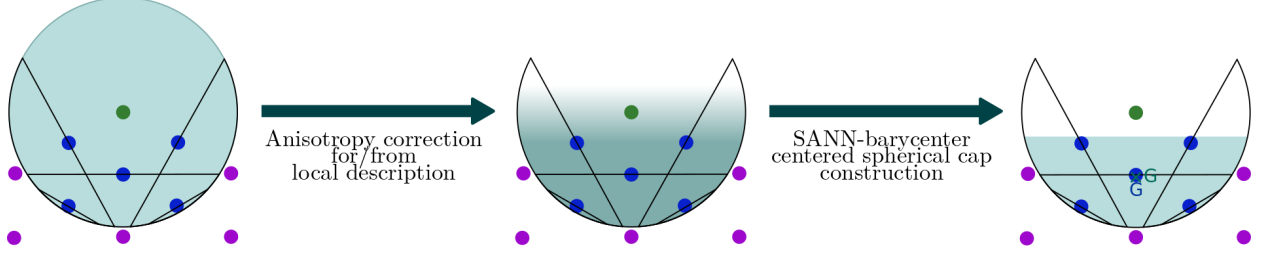


FIG. 4: Illustration of the SANN coordination number overestimation, and founding idea of the ASANN algorithm: tackle local anisotropy by focusing on relevant regions for coordination. This leads to the consideration of a coordination spherical cap instead of a sphere. The chosen spherical cap reproduces the local anisotropy, through our first order local anisotropy estimator (using a SANN-based barycentre with solid-angle weighting).

context of the Voronoi tessellation for its canonicity.<sup>11</sup> Mathematically, this can be written:

$$\vec{G}_i^{(m)} = \frac{\sum_{j=1}^m \Omega_{i,j} \vec{r}_{i,j}}{\sum_{j=1}^m \Omega_{i,j}} \quad (4)$$

This SANN-based barycentre is in agreement with the properties expected for retrieving the first order local anisotropy. Indeed, it identifies a global privileged direction in the distribution of coordinating neighbors and, more importantly, it provides with a quantitative estimation of the first order local anisotropy magnitude  $\alpha_i^{(m)}$  defined by:

$$\alpha_i^{(m)} = \frac{\|\vec{G}_i^{(m)}\|}{R_i^{(m)}} \quad (5)$$

Instead of considering a coordination sphere, the ASANN algorithm uses a coordination spherical cap  $\mathcal{C}$  which focuses on the locally relevant region of space for coordination.  $\mathcal{C}_i$  is the spherical cap centered on  $i$  having the same anisotropic properties as the neighboring particles through our first order local anisotropy estimator. This is achieved by setting the radius  $R_{\mathcal{C}_i} = R_i^{(m)}$  and requiring that its barycentre  $\vec{G}_{\mathcal{C}_i}$  is confounded with the SANN-based barycentre  $\vec{G}_i^{(m)}$ :

$$\begin{aligned} \mathcal{C}_i \text{ is centred on } i \\ R_{\mathcal{C}_i} &= R_i^{(m)} \\ \vec{G}_{\mathcal{C}_i} &= \vec{G}_i^{(m)} \end{aligned} \quad (6)$$

The solid angle  $\Omega_{\mathcal{C}_i}$  associated with this spherical cap (with respect to  $i$ ) can be written as:

$$\Omega_{\mathcal{C}_i} = 4\pi(1 - \gamma_{\mathcal{C}_i}) \quad (7)$$

where  $\gamma_{\mathcal{C}_i} \in [0, 1[$  corresponds to the angular correction coefficient, which only depends on the first order local anisotropy magnitude  $\alpha_i^{(m)}$  (see section S1 in the SI for a derivation):

$$\gamma_{\mathcal{C}_i} = \frac{\alpha_i^{(m)} + \sqrt{\alpha_i^{(m)2} + 3\alpha_i^{(m)}}}{3} \quad (8)$$

Therefore, instead of filling up the whole field of view ( $\sum_{j=1}^m \Omega_{i,j} = 4\pi$ ), ASANN neighbors only cover the spherical cap. This leads to the fundamental equation of ASANN:

$$\sum_{j=1}^{m'} \Omega'_{i,j} = \sum_{j=1}^{m'} 2\pi \left( 1 - \frac{r_{i,j}}{R_i^{(m')}} \right) = \Omega_{\mathcal{C}_i} = 4\pi(1 - \gamma_{\mathcal{C}_i}) \quad (9)$$

where the ASANN coordination number  $m'$  and the coordination cap radius  $R_i^{(m')}$  are defined by extending the SANN coordination radius definition to include the angular correction coefficient  $\gamma_{\mathcal{C}_i}$ :

$$R_i^{(m')} = \frac{\sum_{j=1}^{m'} r_{i,j}}{m' - 2 \times (1 - \gamma_{\mathcal{C}_i})} < r_{i,m'+1} \quad (10)$$

Note that for isotropic distributions ( $\gamma_{\mathcal{C}_i} = 0$ ), Eq. 10 goes back to Eq. 3, which means that SANN and ASANN coincide for this case.

Using the inequality of Eq. 10, the ASANN coordination number  $m'$  can be easily computed by iteratively computing  $R_i^{(m')}$ , starting with  $m' = \lfloor 2(1 - \gamma_{\mathcal{C}_i}) \rfloor + 1$ , and increasing  $m'$  until the inequality is respected.

In principle, this angular correction could be computed self-consistently, i.e., the angular correction term is recomputed at each iteration along with the ASANN coordination radius. This self-consistency breaks the analytic, direct link between the fundamental equation (Eq. 9) and the coordination radius definition (Eq. 10). Therefore, we only explore the “perturbative” angular correction, where the SANN-based barycentre  $\vec{G}_i^{(m)}$  and the associated angular correction term  $\gamma_{\mathcal{C}_i}$  are computed only once, using the SANN-based coordination radius and coordinating neighbors. Then, the iterative ASANN algorithm (Eq. 10) is performed with a fixed  $\gamma_{\mathcal{C}_i}$ . As demonstrated in the supporting information, this algorithm converges to a unique solution. The overall workflow for ASANN is depicted in Figure 5,

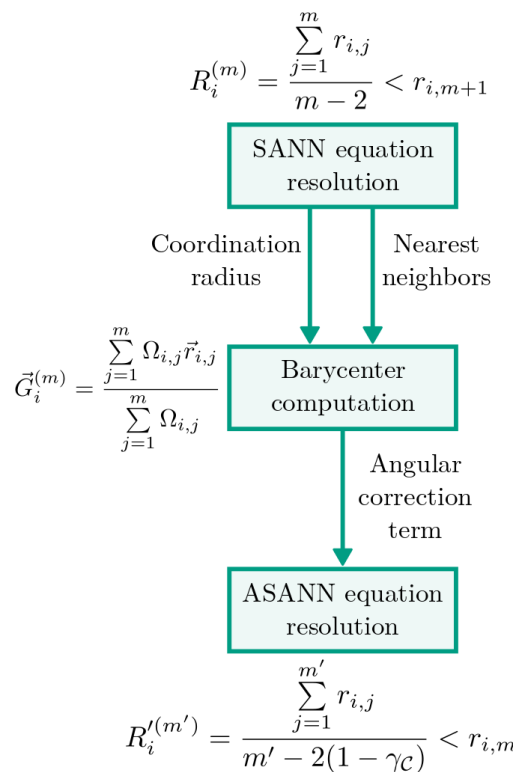


FIG. 5: Schematic representation of the ASANN workflow: Based on the SANN coordination sphere, the barycenter and the corresponding angular correction is computed, leading to the ASANN coordination spherical cap.

which also provides information regarding the computational cost of ASANN compared to SANN: Essentially, there is a factor of two between them, since the iterations have to be performed twice. Knowing that ASANN coordination number is lower or equal to the one from SANN (see SI for a proof), there are, in general, less ASANN than SANN iterations.

Based on general properties of ASANN, partially inherited from SANN, we show the expected applicability domain of the two algorithms in Figure 6. The first major limitation is related to low coordination numbers (e.g., four) in combination with isotropic distributions (e.g., a tetrahedron). In this case, the coordination radius  $R_i^{(4)}$  of Eq. 3 is twice the distance between atom  $i$  and its nearest neighbors  $j$ . If there are atoms closer than  $2r_{ij}$ , the inequality is not satisfied and the coordination shell is expanded to include next nearest neighbors. This is the case for most organic molecules (see Figs. S2-S5 and the associated discussion), but

also in ionic compounds where small cations are located in tetrahedral sites as shown in the SI (Tables S4 to S15), where we have analyzed a variety of salts, oxides and perovskites. The particularity of these systems is most clearly seen for the case of a graphene sheet (Fig. S5): the barycenter of neighboring atoms is confounded with the central atom, so there is no first order anisotropy and ASANN is equivalent to SANN. There are three nearest neighbor atoms at a distance  $r_{C-C}$  and the next nearest neighbors are already at  $\sqrt{3}r_{C-C}$ , again due to the directional bonding, but which is well below the  $R_i^{(3)} = 3r_{C-C}$  (Eq. 3 for the perfect symmetry of this system). If there would not be an anisotropic preference due to bond formation characteristic for an  $sp^2$  carbon atom, there would also be “space” for at least two more particles one above and one below the plane within such a coordination sphere. But this is not the case and, therefore, for geometrical reasons, ASANN significantly overestimates the conventional coordination number of 3, by adding the 6 next nearest neighbors, reaching a  $CN_{ASANN}=9$  for carbon in graphene. The same geometrical reasons and deviations from homogeneous distributions apply to the ionic compounds referred to above. Since ASANN is built on a first order anisotropy, the application to nanopores or slits is also at the border of its applicability: if the SANN coordination sphere spans two surfaces, the first order anisotropy vanishes and ASANN does not improve upon SANN. Higher order anisotropy corrections might correct SANN in this situation as well as for the isotropic, low-coordination environments mentioned above. Second, density fluctuations are problematic for the scale free algorithms. Therefore, cross-checks with the coordination radii are required for biphasic systems and the liquid-gas interface. As we will show by numerical results below, ASANN extends, however, the applicability of SANN to (inhomogeneous) solutions, the solid-liquid interface and rough surfaces including nano-particles.

The time complexity of (A)SANN is in the worst-case  $O(nk)$  for pre-treated input, where  $n$  is the number of particles, and  $k$  is the maximum coordination number. In fact, the most time consuming step is due to the input pre-processing, required for both SANN and ASANN algorithms. This pre-processing comprises the computation of pair-wise distances (and vectors for ASANN) in  $O(n^2)$ , and sorting those distances in  $O(n^2 \log(n))$ . It is, therefore, this step that is also responsible for the space complexity of (A)SANN in  $O(n^2)$ . In practice, this pre-processing quickly becomes the computational bottleneck for large systems. As a consequence, particular care was taken for optimizing these steps in our Python 3 implementation, heavily relying on the Numpy library<sup>27,28</sup>. We thus combine low-level

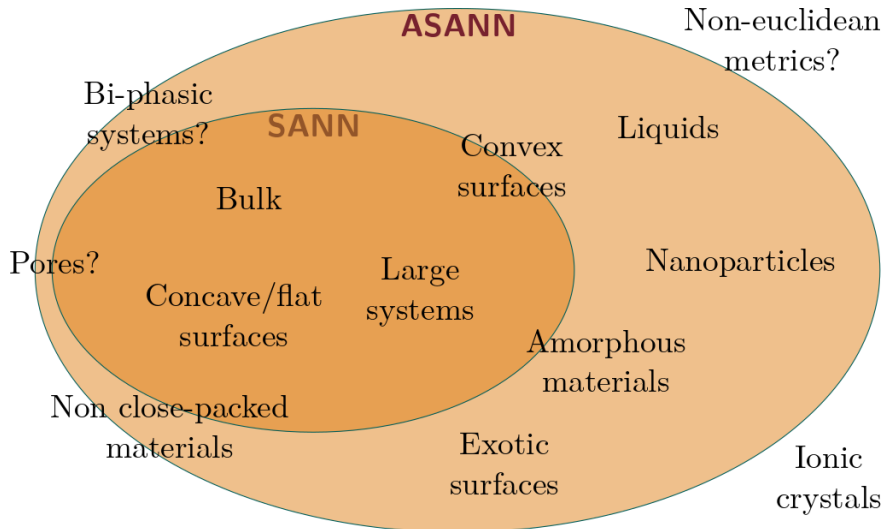


FIG. 6: Venn diagram depicting the applicability domains of both SANN and ASANN. By design, ASANN extends the applicability domain of SANN to systems presenting local anisotropy.

language performance for large systems, with convenient Python 3 syntax and modularity. As an example, the computation of the Cs coordination numbers in a 742 atom system (vide infra), took around 0.14 seconds and the computation of all coordination numbers in a 2942 atom system took about 2 seconds (93% is spent on input pre-processing) on a personal laptop (Core i5-6300HQ CPU).

### III. COMPUTATIONAL DETAILS

The presented algorithms have been implemented using Python3<sup>29</sup> and are available in the supporting information and also on the webpage of the authors. The implementation relies on the Numpy library<sup>27,28</sup> whenever possible and ASE<sup>12</sup> for reading common coordinate formats. A particular effort was made to provide with extensively commented code, including documentation and illustrative examples.

Figures have been generated with a custom ASE<sup>12</sup> fork to visualize the coordination numbers.

All DFT computations are performed using VASP<sup>30,31</sup>. The PBE exchange-correlation functional<sup>32</sup> was applied and a 400 eV energy cut-off was chosen for the plane-wave basis set. The PAW pseudo-potentials were used for describing the electron-ion interactions.<sup>33,34</sup>

Some of the optimized metallic surfaces and metal clusters have been previously published by our group.<sup>35-37</sup> The graphite electrode was set up as follows: A symmetric six layer p(2×2) graphite slab has been built based on the optimized bulk geometry using the PBE-dDsC<sup>38</sup> level of theory. A  $9 \times 9 \times 1$  Monkhorst-Pack K-point grid was used to integrate the Brillouin zone. To assess the charge distribution at negative potentials, the number of electrons was varied and the linearized Poisson-Boltzmann equation was utilized to neutralize the unit cell as implemented in the VASPsol package.<sup>39,40</sup> Similarly to our previous study on electroreduction in aprotic medium,<sup>41</sup> the isodensity value defining the cavity has been lowered to  $5 \cdot 10^{-4} e/\text{\AA}^3$ . In the case of graphite, this value ensures an absence of implicit solvent between the dispersion bound graphene layers.

The potential of zero charge of graphite is, according to our setup, -0.2 V vs SHE, which compares reasonably well with experiments (0.1 to 0.3 V depending on the nature of the carbon electrode<sup>42-44</sup>) considering the idealized morphology of our electrode and the expected systematic error due to the simplified electrolyte model as discussed in the context of the potential dependent adsorption of pyridine on a gold single crystal surface<sup>45</sup>. A negative charge of -0.25 electrons in the unit cell (-0.125 electrons per surface) led to a potential drop of 2.4 V compared to the potential of zero charge and thus corresponds to a significant negative polarization. Based on the DFT computations, the unit cell was replicated four times in each in-plane direction, resulting in 128 carbon atoms per layer and a total charge of 4 e-. Following a setup very similar to ref 46, the Lennard-Jones (LJ) parameters for graphite are taken from UFF.<sup>47</sup> The initial solvent distribution is obtained from the predefined TIP3P<sup>48</sup> box with about 35 Å of water surrounding the system, resulting in a box of 716 water molecules. To simulate a 1 M electrolyte, 15 Na<sup>+</sup> (or Cs<sup>+</sup>) and 11 Cl<sup>-</sup> ions, described by their standard AMBER force field parameters, are added using tleap of the AmberTools. MM computations are performed with NAMD 2.9<sup>49</sup> for 5 ns (time step of 2 fs), applying a Langevin thermostat (300 K) with a damping coefficient of 1 ps<sup>-1</sup>. The Langevin barostat (1 bar) is applied with a piston period of 100 fs and a decay of 150 fs.

## IV. RESULTS AND DISCUSSION

### A. Metallic surfaces

As a first test, we apply ASANN to various metallic surfaces. The selection of surfaces is intended to span a wide range of typical cases to assess the applicability domain of SANN and ASANN.

For face centered cubic (fcc) metallic surfaces, the metal-metal contact distance in the bulk is a reliable measure for identifying nearest neighbors. Reassuringly, closed packed, flat surfaces are well described by SANN and ASANN, giving identical results that are in perfect agreement with cutoff based methods (see SI). In other words, flat surfaces are not anisotropic enough to induce an expansion of the SANN coordination sphere to the next nearest neighbors.

An increased local anisotropy can be obtained for nano-rods. The large rod, constructed from the corresponding bulk without any geometry optimization (Figure 7), is still well and identically described by SANN and ASANN.

However, stepped surfaces (see Figures 8 and 9) are more challenging for SANN, while ASANN consistently retrieves the cutoff based coordination number. In particular, SANN is very sensitive to small local surface induced contractions, while ASANN is more robust, avoiding coordination number overestimations seen in Figure 9 for SANN.

To simulate low-coordinated atoms,<sup>9</sup> add-atoms have also been analyzed. Based on visual inspection (and confirmed by cutoff based methods), the square of addatoms on a (100) surface is characterized by two nearest neighbors from the square (the third is not “touching”, i.e., it is a next-nearest neighbor) and four from the (100) plane below, yielding a coordination number of six. Similarly, an addatom of the triangle on a (111) surface possesses two nearest neighbors from the triangle and three from the supporting (111) plane, resulting in a coordination number of five. Recalling that the flat (100) surface of fcc closed packed metals has a coordination number of 8 for its surface atoms, the  $CN_{SANN}=8$  for surface addatoms reveal a clear overestimation (see Figures 10 and 11), qualitatively failing to identify the low-coordination environment. This contrasts with  $CN_{ASANN}$  that is in full agreement with visual inspection and cutoff based methods.

To challenge the algorithms even further, we apply them to an optimized, amorphous gold

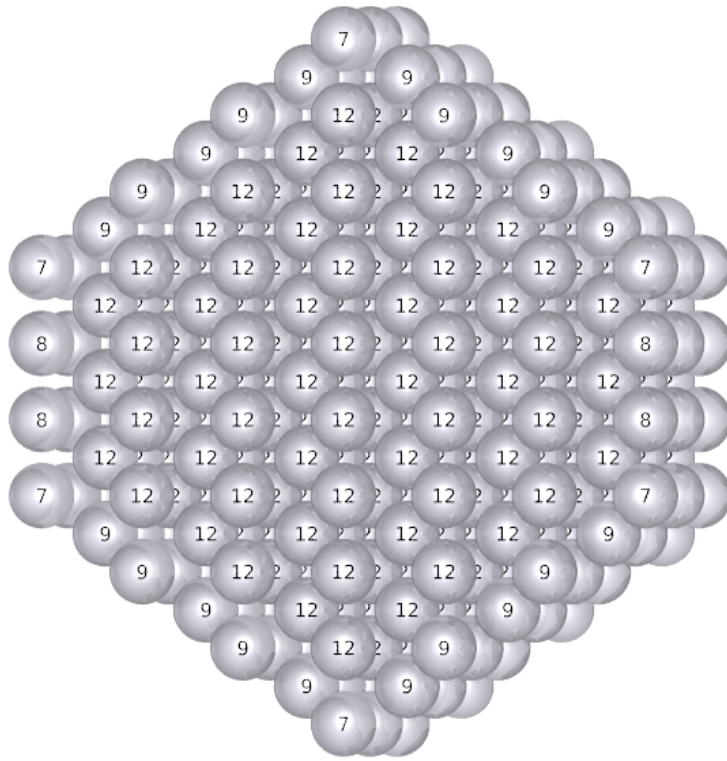


FIG. 7: Transversal cut of a periodic nano-rod of an fcc metal, with superimposed coordination numbers. SANN and ASANN coordination numbers are identical.

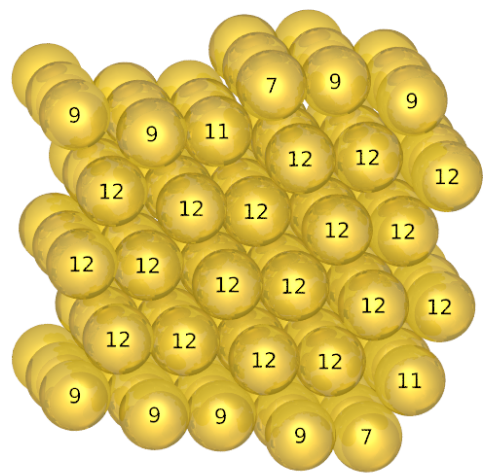


FIG. 8: Side view of a stepped gold surface (664), with superimposed coordination numbers. SANN and ASANN coordination numbers are identical.



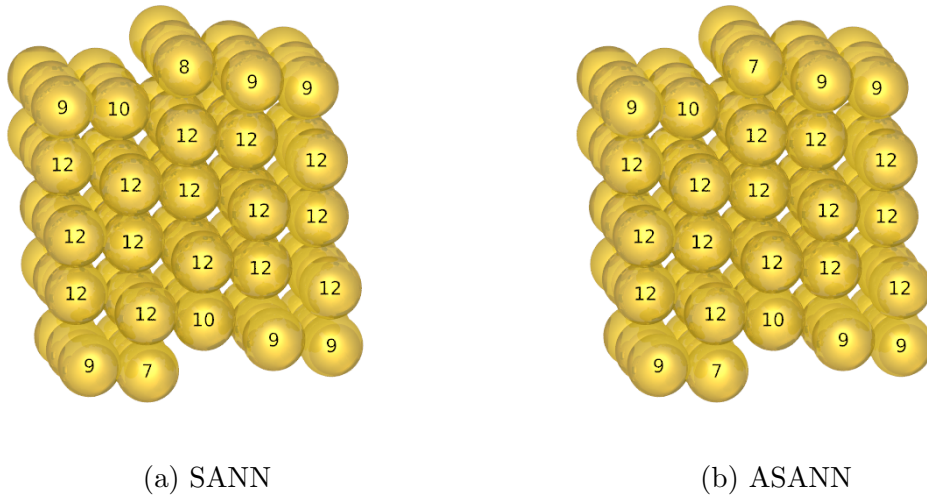


FIG. 9: Side view of a stepped gold surface (644), with superimposed SANN (a) and ASANN (b) coordination numbers.

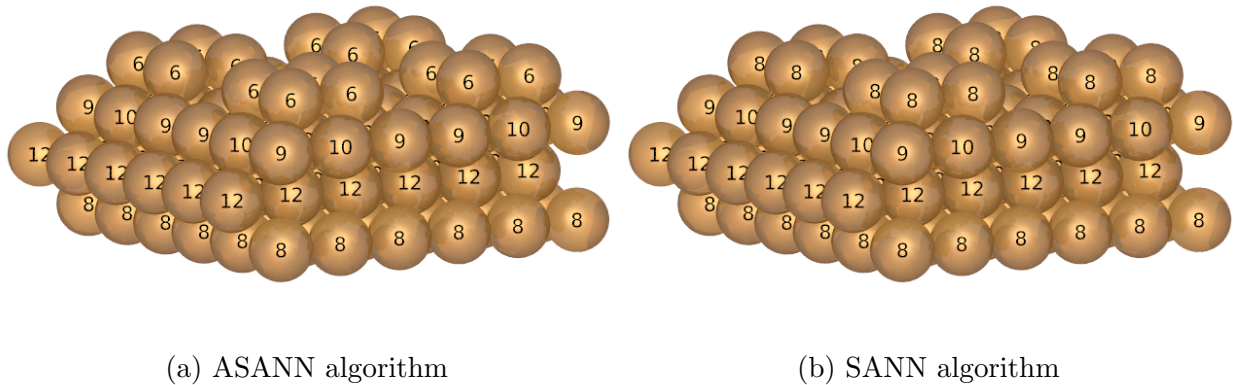
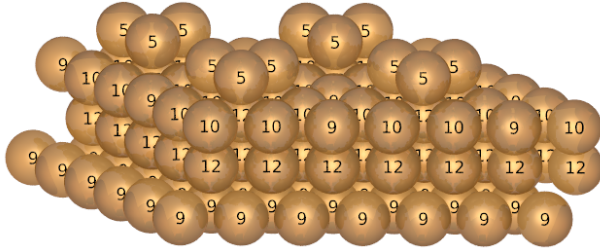
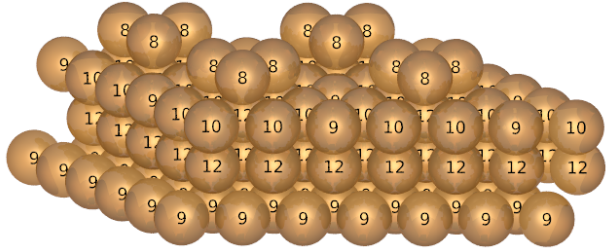


FIG. 10: Representation of a  $p(3 \times 3)$  Cu(100) surface with a square of 4 Cu adatoms. Coordination numbers from SANN (a) and ASANN (b) are superimposed.

nanoparticle (see Figure 12). As can be seen, the particle features atoms that clearly form tips and, therefore, we expect low CNs for them. However, just like for the adatoms on the planar surfaces, SANN fails to correctly identify these low-coordinated atoms. For these highly anisotropic distributions, the SANN coordination radius has to expand significantly to

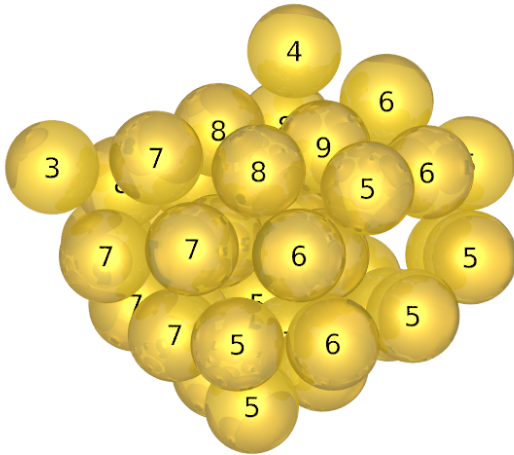


(a) ASANN algorithm

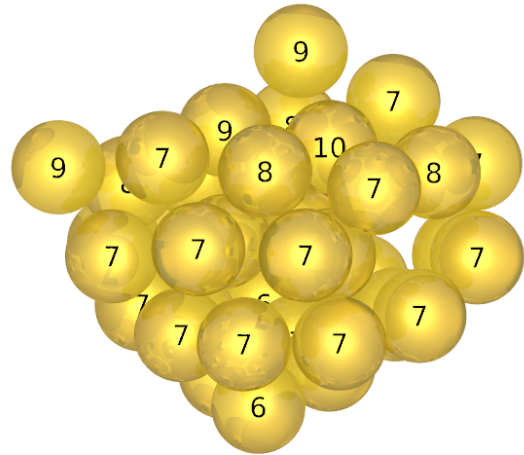


(b) SANN algorithm

FIG. 11: Representation of a  $p(3\times 3)$  Cu(111) surface with a triangle of 3 Cu addatoms. Coordination numbers from SANN (a) and ASANN (b) are superimposed.



(a) ASANN algorithm



(b) SANN algorithm

FIG. 12: Low energy, amorphous  $\text{Au}_{38}$  nanoparticle. Coordination numbers from SANN (a) and ASANN (b) are superimposed. SANN overestimation is particularly visible at the tips.

reach the entire  $4\pi$  for the sum of solid angles. In contrast, ASANN exploits the anisotropy of the SANN coordination sphere to construct a spherical cap and, therefore, succeeds in identifying intuitive nearest neighbors in all cases.

As a rule of thumb, we empirically found that SANN rarely finds coordination numbers below 7 in these systems, illustrating the surface-related overestimation reported earlier.<sup>23</sup> ASANN, however, performs very well in all tested systems, correctly identifying low-coordination environments. As shown in the supporting information, this robust, parameter-free performance is also transmitted to generalized coordination numbers<sup>9</sup> of various surfaces, dispensing with the need to define a characteristic distance or manual counting.

## B. Coordination numbers at charged interfaces

The interaction of cations with negatively charged electrodes is of growing interest in the electro-reduction of CO and CO<sub>2</sub>.<sup>50-52</sup> In a different context, electrolyte solutions at charged silica interfaces are widely studied to understand the behavior of confined electrolytes.<sup>53</sup> Finally, the advent of carbon based supercapacitors<sup>54,55</sup> has spurred simulations of electrolyte solutions at charged carbon interfaces.<sup>56,57</sup> Such simulations are also of particular interest to understand the intercalation mechanism of alkali metal ions into graphitic surfaces.<sup>58</sup> Here, we consider a simple model system: aqueous salt solutions in contact with a perfect graphite electrode. On the one hand, the experimental capacitance of a graphitic electrode is not strongly dependent on the alkali cation in the electrolyte,<sup>59</sup> which might suggest that the double layer structure is similar for all alkali metal cations. On the other hand, Cs<sup>+</sup> ions are, in contrast to Na<sup>+</sup> ions, known to be soft and large, implying a low charge density at their surface. As a consequence, Cs<sup>+</sup> ions are expected to partially loose their solvation shell to approach the negatively charged electrode surface more closely, which increases both the VdW interactions and the Coulomb attraction.

We probe the difference between Cs<sup>+</sup> and Na<sup>+</sup> based on classical molecular dynamics simulations of a graphite carbon electrode at a potential of about -2.5 V (see Computational Details) in ~1 M aqueous CsCl and NaCl solutions. To analyze the double layer, we investigate the existence of a link between coordination number of ions in water (determined by ASANN, SANN and a  $g(r)$  based cut-off algorithm) and their adsorption state.

Figure 13 displays the structuring of the electrolyte at the electrode surface. As commonly observed<sup>60</sup>, for both electrolyte solutions the distribution curve of the excess cation density presents two well-defined peaks, corresponding to two distinct layers of preferred

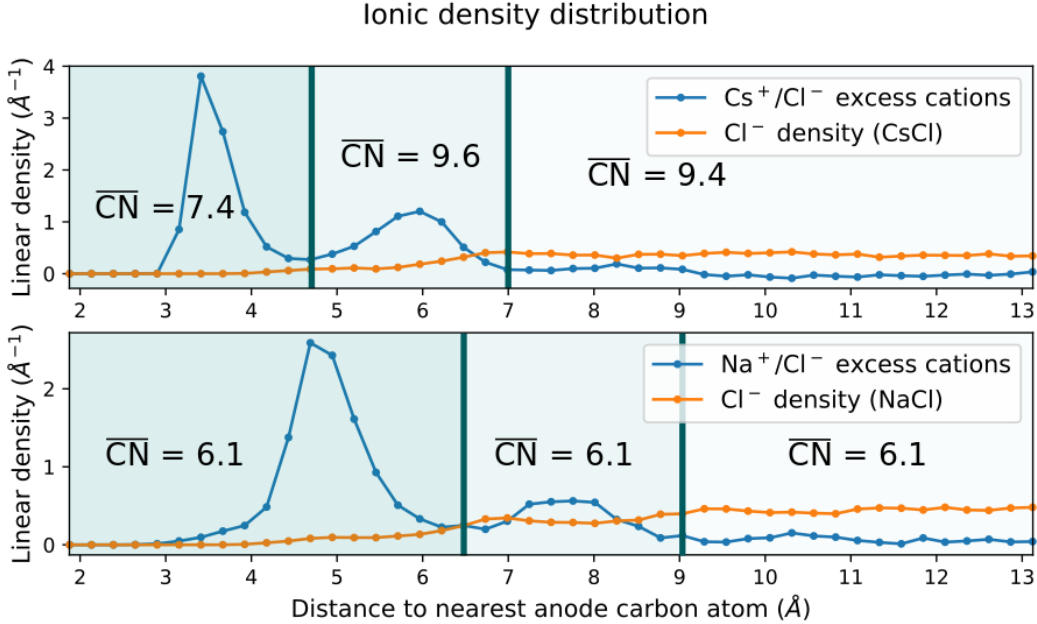


FIG. 13: Ionic density of  $\sim 1$  M CsCl (top panel) and NaCl (bottom panel) aqueous solutions near a graphite anode ( $-2.5$  V vs SHE). The excess counter-ions (compared with the co-ions, equivalent here to the charge density) and the co-ions density are represented for both systems. Three distinct zones are identified, corresponding to different adsorption/trapping states for the counter-ions. The average ASANN coordination number for the counter-ions (with water molecules and co-ions) is superimposed for each zone.

presence near the anode. The strong first cation excess peak is followed by a nearly charge compensated zone before a second, although lower, cation excess peak occurs which then decays to zero towards the bulk solution. Besides, the anion ( $\text{Cl}^-$ ) distribution is barely structured and independent on the nature of the cation. The details of the charge-excess profiles are quite different:  $\text{Cs}^+$  excess density has the first peak at about  $3.5$  Å, while the corresponding one for  $\text{Na}^+$  is at  $4.8$  Å. This difference indicates that  $\text{Cs}^+$  can indeed approach the electrode more closely. The second peak is, accordingly, shifted from  $6$  Å to  $7.5$  Å when going from caesium to sodium.

Table I and II report the ASANN and SANN coordination numbers of  $\text{Cs}^+$  and  $\text{Na}^+$ , respectively. For reference purposes, the radial distribution function of the O- $\text{Cs}^+$  was analysed and a first coordination sphere radius cutoff of  $4.0$  Å was identified. From this radius, cutoff-based coordination numbers have been computed. All three algorithms agree that on average at least one water molecule is lost when  $\text{Cs}^+$  is located in the first layer,

Average coordination numbers of Cs <sup>+</sup> ions			
Method	Desolvated trapping layer	Solvated trapping layer	Bulk-like layer
	$h < 4.7 \text{ \AA}$	$4.7 < h < 7 \text{ \AA}$	$h > 7 \text{ \AA}$
ASANN	7.4 (1.1)	9.6 (1.2)	9.4 (1.2)
SANN	9.1 (1.1)	10.6 (1.2)	10.5 (1.2)
fixed cutoff	7.1 (1.1)	9.2 (1.3)	9.0 (1.3)

TABLE I: Comparison of average Cs<sup>+</sup> ions coordination numbers determined by ASANN, SANN and a fixed cutoff (4.0 Å) for each defined region. The data is displayed in the form: mean (standard deviation).

while the second layer is indistinguishable from the bulk solution in terms of the coordination numbers. This can be rationalized based on Van der Waals (VdW) radii: a fully solvated Cs ion cannot approach the graphitic electrode closer than roughly 7 Å ( $r_C + r_{Cs^+} + 3 \text{ \AA}$ , where 3 Å is a rough estimate for the size of a water molecule).

For Na<sup>+</sup> ions, however, the coordination numbers are the same in the two excess regions as in the bulk, despite the qualitative similarity between the two cation excess profiles. Even the nearest Na<sup>+</sup> layer ( $\sim 4.8 \text{ \AA}$ ) is not close enough to the surface to induce a partial desolvation (see Table II). In other words, in contrast to Cs<sup>+</sup>, the solvation shell of Na<sup>+</sup> is strong and stable across the double layer and the three methods agree on an approximate average coordination number of six, indicating minor fluctuations and a rather isotropic distribution, in agreement with spectroscopic data.<sup>6</sup>

To quantitatively determine if ASANN provides with “good” coordination numbers, we assess their ability to discriminate between “close” and “bulk” Cs ions. Specifically, we want to quantify if  $CN_{ASANN}$  is statistically lower (i.e., below  $CN_{\text{thresh}}$ ) at the interface than in the bulk. For the sake of comparison, we apply the same analysis also to SANN and to the cutoff based algorithm.

To have a quantifiable measure whether or not  $CN_{ASANN}$  is lower at the interface than in the bulk, we explore the relation between closeness and loss of coordination via a Receiver Operating Characteristic (ROC) curve analysis. The ROC curve is one of the most rigorous and well-established tools to evaluate the discriminating power for a binary classifier (here

Average coordination numbers of Na <sup>+</sup> ions			
Method	First trapping layer	Second trapping layer	Bulk-like layer
	$h < 6.5 \text{ \AA}$	$6.5 < h < 9 \text{ \AA}$	$h > 9 \text{ \AA}$
ASANN	6.1 (0.6)	6.1 (0.6)	6.1 (0.5)
SANN	6.5 (0.7)	6.6 (0.8)	6.5 (0.8)
fixed-cutoff	5.8 (0.5)	5.8 (0.5)	5.8 (0.5)

TABLE II: Comparison of average Na<sup>+</sup> ions coordination numbers determined by ASANN, SANN and a fixed cutoff (3.2 Å) for each defined region. The data is displayed in the form: mean (standard deviation).

close/not close) depending on a threshold value (here  $CN_{ASANN}$ ).<sup>61,62</sup> The ROC curve of a class of binary classifiers is constructed by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) computed on a data set, for different threshold values.

Here, the TPR is the ratio of  $CN^i < CN_{\text{thresh}}$  for the  $i$  Cs ions “close” to the electrode to the number of Cs ions “close” to the electrode overall, while the False Positive Rate is the ratio of  $CN^i < CN_{\text{thresh}}$  for Cs ions *not* “close” to the electrode to the number of Cs ions *not* “close” to the electrode overall. The threshold value can be varied between the minimum and maximum observed with a given algorithm. Here, we define “close” as the Cs – surface distance  $\leq 4.7 \text{ \AA}$ .

Such an analysis is reported on Figure 14. Let us first look at the ASANN results to explain how to read a ROC curve. Since the CNs are reasonable descriptors, for low  $CN_{\text{thresh}}$  (e.g., 7), only “close” Cs ions are identified (FPR=0%), but not all of them (TPR $\approx$ 20%), as shown in Table S16a. Similarly, at high CN (e.g., 10), the TPR reaches  $\approx$ 100%, i.e., all “close” Cs ions have CNs below this value, but about 50% (FPR) of the ions *not* close to the surface have CNs below this value. Hence, increasing the threshold further does not change the TPR, but the FPR reaches 100%.

A major ROC curve analysis is performed through the Area Under the Curve (AUC), as it provides an aggregate measure of classification performance<sup>63</sup>. In this particular example, the AUC is associated with the probability that a random interfacial Cs ion has a lower coordination number than a random bulk Cs ion.

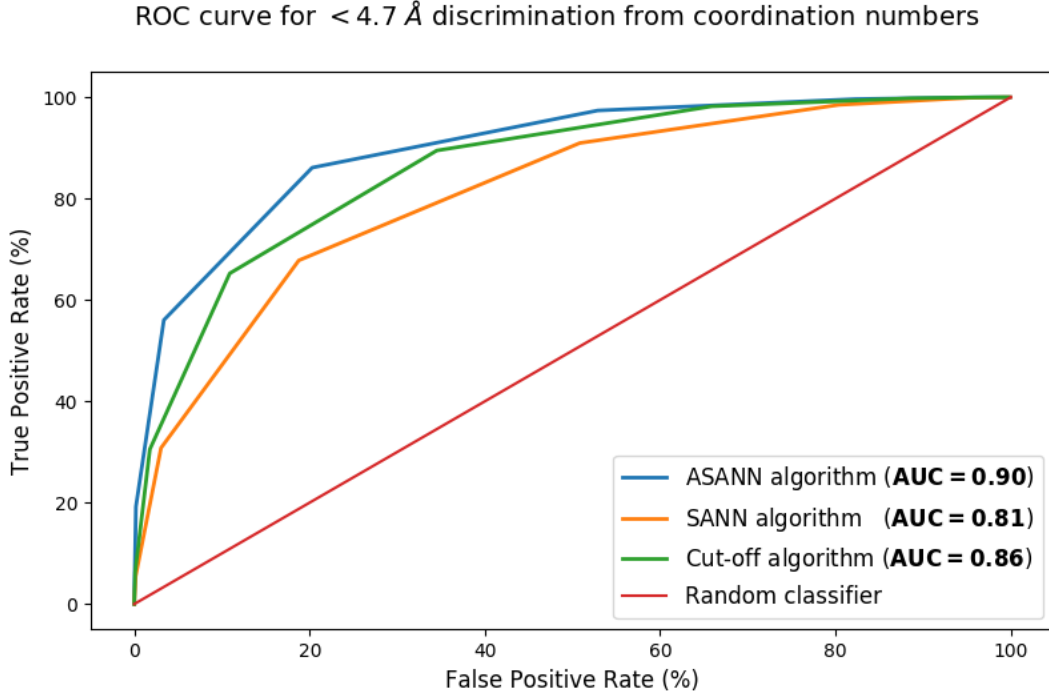


FIG. 14: ROC curve for classifying Cs ions for being “close” to the carbon anode based on the coordination number being below a given threshold. When moving from TPR=20% to TPR=55% with ASANN (blue curve),  $CN_{\text{thresh}}$  is increased from 7 to 8 (see Table S16a in the SI for the raw data). This exploits the geometrical constraint responsible for a partial desolvation when close enough to the surface. The distance criterium for being close enough to the surface is chosen to be  $4.7 \text{ \AA}$ , since it corresponds to the empirical frontier between the partially desolvated adsorption zone and the fully solvated adsorption zone for Cs ions near the graphite anode (see Fig. 13).

With an AUC of 0.90, we find a strong correlation between  $CN_{\text{ASANN}}$  and the proximity of  $\text{Cs}^+$  and the surface. This finding is coherent with ASANN providing reasonable coordination numbers in this system. The same analysis performed with SANN only yields an AUC of 0.81. The cutoff-based algorithm yields an AUC of 0.86, outperforming SANN, but not reaching the performance of ASANN. This is rather remarkable, given that the cutoff algorithm is tuned to the system by analyzing the radial distribution function, while ASANN does not know anything about the specificities of the system. On a physical level, the major difference between the cutoff algorithm and ASANN is that the ASANN coordination sphere adapts itself to the local environment, while the cutoff does not. This means that the

cutoff based CNs are more sensitive to thermal fluctuations. Indeed, we found cutoff-based coordination numbers to be slightly noisier ( $\approx +8\%$ ) than their ASANN/SANN counterparts within each region of the studied system (see Table I), rationalizing the poorer AUC performance.

From this comparison it can be concluded that fixed-cutoff based algorithms are somewhat challenged when dealing with liquids, while SANN is not suitable for interfaces. In contrast, ASANN combines the best features of these algorithms and performs well. Therefore, ASANN is most relevant for describing solid/liquid interfaces.

### C. Application to Au NP energy fitting

Exploring the energy landscape of alloy nano-particles from first principles is computationally very expensive. Therefore, cheap energy evaluations based on simplified Hamiltonians are often applied.<sup>64-66</sup> Furthermore, if the model Hamiltonian captures the essential physics, it can also be used to gain insight into the driving force of the alloy formation. Inspired from cluster expansions, we here test the use of a simple model Hamiltonian  $H_m$  for bimetallic materials.  $H_m$  depends only on the number of atoms of each element and the number of bonds between each possible elements pairs:

$$H_m = \sum_x \Gamma_x N_x + \sum_{x,x' \geq x} \Gamma_{x,x'} N_{x,x'} \quad (11)$$

In other words, it is a two-body model Hamiltonian. For alloys of elements with quite different atomic radii, the alloy nano-particles are not necessarily very regular and selecting characteristic distances is not straight forward. Therefore, an automatic, parameter free algorithm to identify bonds is particularly beneficial for these systems. In the following, we apply this simplified model to bimetallic Au-Cu materials, where  $H_m$  becomes

$$\begin{aligned} H_m &= \Gamma_{Au} N_{Au} + \Gamma_{Cu} N_{Cu} + \Gamma_{Au,Au} N_{Au,Au} + \Gamma_{Au,Cu} N_{Au,Cu} + \Gamma_{Cu,Cu} N_{Cu,Cu} \\ &= \mathbf{\Gamma} \cdot \mathbf{N} \end{aligned} \quad (12)$$

with  $\mathbf{\Gamma}$  the set of parameters to fit.  $\mathbf{N}$  is constructed by enumeration of atoms and bonds of each type. The automatic bond enumeration was performed using ASANN and SANN for comparison purposes. In order to have consistent bond description, a bond between atoms A and B is only counted if and only if B is found as nearest neighbor of A and reciprocally.



DFT energy vs. predicted energy using a model Hamiltonian based on (A)SANN  
(training set: NP + Bulk; test set: NP)

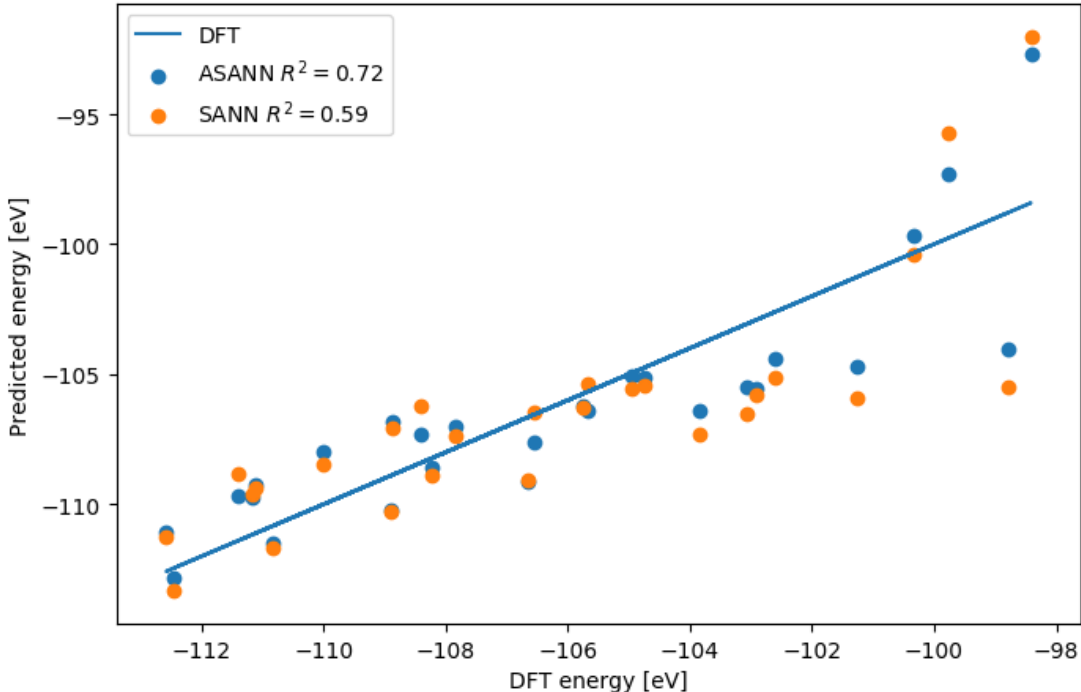


FIG. 15: Parity plot for the total energy predicted by a 2-body model Hamiltonian fitted using the SANN vs. ASANN coordination algorithm. The whole set of 27 nanoparticles + 6 Bulk structures was used as training set, and the test set was composed of the 27 Au-Cu bimetallic nanoparticles only for graphical considerations.

Such filtering is particularly beneficial to SANN, where buried atoms are mostly spared from coordination overestimation. Hence, the bidirectional bond definition improves the soundness of bonds between bulk and surface atoms.

For fitting the model Hamiltonian, we use a structurally diverse set of 27 Au-Cu bimetallic nanoparticles composed of 38 atoms with various compositions and morphologies and 6 Au-Cu bulk structured alloys with various Au-Cu ratios. The coordinates are provided in the supplementary information.

When fitting only the nanoparticle set, the algorithms lead to model Hamiltonians with mean absolute deviation (MAD) of 0.8 eV for ASANN and 0.7 eV for SANN, see Figure S1. In other words, both approaches are very similar, with SANN performing slightly better than ASANN. However, the ASANN-derived parameters are found to be more robust than their SANN-based counterparts when adding the bulk structured alloys to the training

set (i.e. diversifying the training set). The fitted parameters are not only slightly less impacted (see Tables S1 and S2 in the SI), but the overall model performance is better with ASANN-based bonds (MAD near 1.5 eV for ASANN, versus 1.8 eV for SANN). This is especially true when using only the nanoparticles as test set. The predicted energy for the nanoparticles with respect to the DFT reference energies yields a coefficient of determination  $R^2 = 0.72$  and  $R^2 = 0.59$  with ASANN and SANN based 2-body patterns, respectively (see Figure 15). Similarly, the pure bulks are better described by ASANN (MAD=0.37 eV), than SANN (MAD=0.60 eV). Hence, the better performance of SANN for exclusively fitting the nanoparticles set can be traced back to error cancellation between coordination number overestimation and missing higher-order many-body terms at the surface. In summary, we conclude that ASANN provides a balanced and physical sound description between bulk and surface-rich nanoparticles. Therefore, ASANN can also be applied to alloys of elements with quite different atomic radii. Similarly, ASANN could be used to provide a simple estimation of the surface of nanoparticles, by considering the union (or surface sum) of the planar edge of each coordination cap. A more rigorous approach using ASANN to define the surface of nanoparticles would be through the definition of a locally adaptative coordination radius, defining the local coarseness of the surface.

## V. CONCLUSION

In this work we have extended the solid-angle based nearest-neighbor (SANN) algorithm to account for the local anisotropy, defining the anisotropically corrected solid-angle based nearest-neighbor (ASANN) method. ASANN allows to efficiently compute parameter-free coordination numbers and significantly improves over SANN at interfaces, where the coordination environment is highly anisotropic, such as nano-particles and partially desolvated ions at solid/liquid interfaces, where SANN leads to overestimations of the coordination number. In all situations tested, the ASANN coordination numbers are either identical or physically more sound than the ones provided by SANN and on par with system-specific, cut-off based coordination numbers. Therefore, ASANN provides a robust method to compute coordination numbers for describing alloy nano-particles, extended surfaces and (electrified) solid/liquid interfaces.

## ACKNOWLEDGMENTS

We thank Paul Clabaut, Mathilde Iachella, David Loffreda and Kamila Kazmierczak for providing us the geometries of nanorods and nanoparticles, bulk alloys and cobalt surfaces, respectively. The authors thank the SYSPROD project and AXELERA Pôle de Compétitivité for financial support (PSMN Data Center).

## VI. SUPPLEMENTARY MATERIAL

Additional Theorems, Figures and Tables are available in the supplementary material.

## REFERENCES

- <sup>1</sup>A. E. Reed, L. A. Curtiss, and F. Weinhold, *Chem. Rev.* **88**, 899 (1988).
- <sup>2</sup>H. Elgabarty, R. Z. Khaliullin, and T. D. Kuhne, *Nat. Commun.* **6** (2015).
- <sup>3</sup>L. Pecher, S. Laref, M. Raupach, and R. Tonner, *Angew. Chem., Int. Ed.* **56**, 15150 (2017).
- <sup>4</sup>R. Staub, M. Iannuzzi, R. Z. Khaliullin, and S. N. Steinmann, *Journal of Chemical Theory and Computation* **15**, 265 (2019).
- <sup>5</sup>J. Marcalo and A. P. De Matos, *Polyhedron* **8**, 2431 (1989).
- <sup>6</sup>J. Mahler and I. Persson, *Inorg. Chem.* **51**, 425 (2012).
- <sup>7</sup>S. Varma and S. B. Rempe, *Biophysical Chemistry* **124**, 192 (2006).
- <sup>8</sup>L. M. Falicov and G. A. Somorjai, *Proceedings of the National Academy of Sciences* **82**, 2207 (1985).
- <sup>9</sup>F. Calle-Vallejo, J. I. Martínez, J. M. García-Lastra, P. Sautet, and D. Loffreda, *Angewandte Chemie International Edition* **53**, 8316 (2014).
- <sup>10</sup>F. C. Frank and J. S. Kasper, *Acta Cryst* **11**, 184 (1958).
- <sup>11</sup>M. O’Keeffe, *Acta Crystallogr A Cryst Phys Diffr Theor Gen Crystallogr* **35**, 772 (1979).
- <sup>12</sup>A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Duak, 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. Schitz,

- O. Schtt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, *Journal of Physics: Condensed Matter* **29**, 273002 (2017).
- <sup>13</sup>B. Cordero, V. Gmez, A. E. Platero-Prats, M. Revs, J. Echeverra, E. Cremades, F. Barragn, and S. Alvarez, *Dalton Trans.* , 2832 (2008).
- <sup>14</sup>C. A. Coulson and G. S. Rushbrooke, *Phys. Rev.* **56**, 1216 (1939).
- <sup>15</sup>J. G. Kirkwood and E. M. Boggs, *The Journal of Chemical Physics* **10**, 394 (1942).
- <sup>16</sup>W. Brostow, *Chemical Physics Letters* **49**, 285 (1977).
- <sup>17</sup>P. G. Mikolaj and C. J. Pings, *Physics and Chemistry of Liquids* **1**, 93 (1968).
- <sup>18</sup>J. D. Bernal, *Trans. Faraday Soc.* **33**, 27 (1937).
- <sup>19</sup>A. Rahman, *The Journal of Chemical Physics* **45**, 2585 (1966).
- <sup>20</sup>J. L. Finney and J. D. Bernal, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **319**, 479 (1970).
- <sup>21</sup>J. A. van Meel, L. Filion, C. Valeriani, and D. Frenkel, *The Journal of Chemical Physics* **136**, 234107 (2012).
- <sup>22</sup>H. Ledoux, in *4th International Symposium on Voronoi Diagrams in Science and Engineering (ISVD 2007)* (2007) pp. 117–129.
- <sup>23</sup>J. Higham and R. H. Henchman, *The Journal of Chemical Physics* **145**, 084108 (2016).
- <sup>24</sup>J. K. Noel, P. C. Whitford, and J. N. Onuchic, *The Journal of Physical Chemistry B* **116**, 8692 (2012).
- <sup>25</sup>R. Taylor, *CrystEngComm* **16**, 6852 (2014).
- <sup>26</sup>R. Collins, *Proceedings of the Physical Society* **86**, 199 (1965).
- <sup>27</sup>T. E. Oliphant, *Guide to NumPy*, 2nd ed. (CreateSpace Independent Publishing Platform, USA, 2015).
- <sup>28</sup>S. v. d. Walt, S. C. Colbert, and G. Varoquaux, *Computing in Science & Engineering* **13**, 22 (2011).
- <sup>29</sup>G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Paramount, CA, 2009).
- <sup>30</sup>G. Kresse and J. Hafner, *Phys Rev B* **47**, 558 (1993).
- <sup>31</sup>G. Kresse and J. Furthmuller, *Phys Rev B* **54**, 11169 (1996).
- <sup>32</sup>J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys Rev Lett* **77**, 3865 (1996).
- <sup>33</sup>P. E. Blochl, *Phys Rev B* **50**, 17953 (1994).
- <sup>34</sup>G. Kresse and D. Joubert, *Phys Rev B* **59**, 1758 (1999).

- <sup>35</sup>A. Viola, J. Peron, K. Kazmierczak, M. Giraud, C. Michel, L. Sicard, N. Perret, P. Beau-  
nier, M. Sicard, M. Besson, and J.-Y. Piquemal, *Catal. Sci. Technol.* **8**, 562 (2018).
- <sup>36</sup>M. Dhifallah, M. Iachella, A. Dhouib, F. Di Renzo, D. Loffreda, and H. Guesmi, *J. Phys.  
Chem. C* **123**, 4892 (2019).
- <sup>37</sup>R. V. Mom, S. T. A. G. Melissen, P. Sautet, J. W. M. Frenken, S. N. Steinmann, and  
I. M. N. Groot, *J. Phys. Chem. C* **123**, 12382 (2019).
- <sup>38</sup>S. N. Steinmann and C. Corminboeuf, *J Chem Theory Comput* **7**, 3567 (2011).
- <sup>39</sup>K. Mathew, R. Sundararaman, K. Letchworth-Weaver, T. A. Arias, and R. G. Hennig, *J  
Chem Phys* **140**, 084106 (2014).
- <sup>40</sup>K. Mathew, V. S. C. Kolluru, S. Mula, S. N. Steinmann, and R. G. Hennig, *J. Phys.  
Chem.*, DOI: 10.1063/1.5132354 (2019).
- <sup>41</sup>S. N. Steinmann, C. Michel, R. Schwiedernoch, and P. Sautet, *Phys. Chem. Chem. Phys.*  
**17**, 13949 (2015).
- <sup>42</sup>B. Kastening and S. Spinzig, *Journal of Electroanalytical Chemistry and Interfacial Elec-  
trochemistry* **214**, 295 (1986).
- <sup>43</sup>H. R. Zebardast, S. Rogak, and E. Asselin, *Journal of Electroanalytical Chemistry* **724**,  
36 (2014).
- <sup>44</sup>J. Poon, C. Batchelor-McAuley, K. Tschulik, and R. G. Compton, *Chemical Science*  
(2015).
- <sup>45</sup>S. N. Steinmann and P. Sautet, *J. Phys. Chem. C* **120**, 5619 (2016).
- <sup>46</sup>S. N. Steinmann, P. Sautet, and C. Michel, *Phys. Chem. Chem. Phys.* **18**, 31850 (2016).
- <sup>47</sup>A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, *J Am Chem  
Soc* **114**, 10024 (1992).
- <sup>48</sup>W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J  
Chem Phys* **79**, 926 (1983).
- <sup>49</sup>J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D.  
Skeel, L. Kalé, and K. Schulten, *J Comput Chem* **26**, 1781 (2005).
- <sup>50</sup>J. Resasco, L. D. Chen, E. Clark, C. Tsai, C. Hahn, T. F. Jaramillo, K. Chan, and A. T.  
Bell, *J. Am. Chem. Soc.* **139**, 11277 (2017).
- <sup>51</sup>E. Pérez-Gallent, G. Marcandalli, M. C. Figueiredo, F. Calle-Vallejo, and M. T. M. Koper,  
*J. Am. Chem. Soc.* **139**, 16412 (2017).

- <sup>52</sup>S. Ringe, E. L. Clark, J. Resasco, A. Walton, B. Seger, A. T. Bell, and K. Chan, *Energy Environ. Sci.*, 10.1039.C9EE01341E (2019).
- <sup>53</sup>T. A. Ho, D. Argyris, D. R. Cole, and A. Striolo, *Langmuir* **28**, 1256 (2012).
- <sup>54</sup>S. Bose, T. Kuila, A. K. Mishra, R. Rajasekar, N. H. Kim, and J. H. Lee, *J. Mater. Chem.* **22**, 767 (2011).
- <sup>55</sup>X. Chen, R. Paul, and L. Dai, *Natl Sci Rev* **4**, 453 (2017).
- <sup>56</sup>C. Merlet, C. Péan, B. Rotenberg, P. A. Madden, P. Simon, and M. Salanne, *J. Phys. Chem. Lett.* **4**, 264 (2013).
- <sup>57</sup>A. C. Forse, C. Merlet, J. M. Griffin, and C. P. Grey, *J. Am. Chem. Soc.* **138**, 5731 (2016).
- <sup>58</sup>M. Petrović, I. Šrut Rakić, S. Runte, C. Busse, J. T. Sadowski, P. Lazić, I. Pletikosić, Z.-H. Pan, M. Milun, P. Pervan, N. Atodiresei, R. Brako, D. Šokčević, T. Valla, T. Michely, and M. Kralj, *Nat Commun* **4**, 2772 (2013).
- <sup>59</sup>D. Golub, A. Soffer, and Y. Oren, *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry* **260**, 383 (1989).
- <sup>60</sup>D. A. Welch, B. L. Mehdi, H. J. Hatchell, R. Faller, J. E. Evans, and N. D. Browning, *Advanced Structural and Chemical Imaging* **1**, 1 (2015).
- <sup>61</sup>K. A. Spackman, in *Proceedings of the Sixth International Workshop on Machine Learning*, edited by A. M. Segre (Morgan Kaufmann, San Francisco (CA), 1989) pp. 160–163.
- <sup>62</sup>T. Fawcett, *Pattern Recognition Letters ROC Analysis in Pattern Recognition*, **27**, 861 (2006).
- <sup>63</sup>C. Ferri, J. Hernández-Orallo, and P. A. Flach, in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* (2011) pp. 657–664.
- <sup>64</sup>W. Chen, D. Schmidt, W. F. Schneider, and C. Wolverton, *J. Phys. Chem. C* **115**, 17915 (2011).
- <sup>65</sup>B. Zhu, J. Creuze, C. Mottet, B. Legrand, and H. Guesmi, *J. Phys. Chem. C* **120**, 350 (2016).
- <sup>66</sup>E. Vignola, S. N. Steinmann, K. Le Mapihan, B. D. Vandegehuchte, D. Curulla, and P. Sautet, *J. Phys. Chem. C* **122**, 15456 (2018).