



This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Aarva, Anja; Deringer, Volker L.; Sainio, Sami; Laurila, Tomi; Caro, Miguel A.

# Understanding X-ray Spectroscopy of Carbonaceous Materials by Combining Experiments, Density Functional Theory, and Machine Learning. Part II

Published in: Chemistry of Materials

DOI: 10.1021/acs.chemmater.9b02050

Published: 26/11/2019

*Document Version* Publisher's PDF, also known as Version of record

Published under the following license: CC BY-NC-ND

Please cite the original version:

Aarva, A., Deringer, V. L., Sainio, S., Laurila, T., & Caro, M. A. (2019). Understanding X-ray Spectroscopy of Carbonaceous Materials by Combining Experiments, Density Functional Theory, and Machine Learning. Part II: Quantitative Fitting of Spectra. *Chemistry of Materials*, *31*(22), 9256-9267. https://doi.org/10.1021/acs.chemmater.9b02050

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.





pubs.acs.org/cm

# Understanding X-ray Spectroscopy of Carbonaceous Materials by Combining Experiments, Density Functional Theory, and Machine Learning. Part II: Quantitative Fitting of Spectra

Anja Aarva,\*<sup>,†</sup> Volker L. Deringer,<sup>‡,§</sup> Sami Sainio,<sup>||</sup> Tomi Laurila,<sup>†</sup> and Miguel A. Caro<sup>†,⊥</sup>

<sup>†</sup>Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University, 02150 Espoo, Finland <sup>‡</sup>Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, United Kingdom

<sup>II</sup>Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025, United States <sup>L</sup>Department of Applied Physics, Aalto University, 02150 Espoo, Finland

**ABSTRACT:** Carbon-based nanomaterials are a promising platform for diverse technologies, but their rational design requires a more detailed chemical control over their structure and properties than is currently available. A long-standing challenge for the field has been in the interpretation and use of experimental X-ray spectra, especially for the amorphous and disordered forms of carbon. Here, we outline a unified approach to simultaneously and quantitatively analyze experimental X-ray absorption spectroscopy (XAS) and X-ray photoelectron spectroscopy (XPS) spectra of carbonaceous materials. We employ unsupervised machine learning to identify the most representative chemical environments and deconvolute experimental data according to these



spectral contributions. To fit experimental spectra we rely on *ab initio* references and use all the information available: to fit experimental XAS spectra, the whole XAS fingerprint (reference) spectra of certain sites are taken into account, rather than just peak positions, as is currently the standard procedure. We argue that, even for predominantly pure-carbon materials, carbon K-edge and oxygen K-edge spectra should not be interpreted separately, since the presence of even small amounts of functional groups at the surface manifests itself on the X-ray spectroscopic signatures of both elements in an interlinked manner. Finally, we introduce the idea of carrying out simultaneous fits of XAS and XPS spectra, to reduce the number of degrees of freedom and arbitrariness of the fits. This work opens up a new direction, tightly integrating experiment and simulation, for understanding and ultimately controlling the functionalization of carbon nanomaterials at the atomic level.

# I. INTRODUCTION

Careful material characterization is needed to understand the connection between the structure and the properties of materials. This understanding is vital in development of several promising applications, such as mechanical and biomedical coatings,<sup>1,2</sup> and, for instance, in electrochemical detection of biomolecules.<sup>3</sup> X-ray spectroscopy offers powerful means for detailed characterization but, especially with disordered materials, interpretation of the experimental data is heavily convoluted. We have previously introduced the concept of computer-simulated fingerprint spectra and showed how they can be used in deconvolution and *qualitative* understanding of the experimental X-ray absorption spectroscopy (XAS) and X-ray photoelectron spectroscopy (XPS) spectra of carbonaceous materials.<sup>4</sup>

It has been shown in earlier work that computational references can indeed be promising tools in interpretation of X-ray spectroscopy. In the case of XPS, several studies have presented and utilized computational references in the interpretation of experimental work.<sup>5–14</sup> Computational references have also taken a foothold in the field of XAS characterization.<sup>15–21</sup> For instance, the structure of water in different forms has been successfully studied with the aid of

density functional theory (DFT) calculations combined with experimental work.<sup>16,20–22</sup> Computational references are also applied with well-defined crystal structures.<sup>18,19</sup> Some of these studies approach also quantitative analysis of XAS by weighting linear combinations of the computed spectra<sup>16,20,21</sup> or by utilizing machine learning (ML) based methods.<sup>18,19</sup> Efforts to combine different analytical methods in parallel have also been made.<sup>20–22</sup>

In this work we apply a new approach with a similar philosophy in the characterization of carbonaceous materials. Carbon-based materials work as an ideal test case for our method since we can make a comparison between ordered susbstances, i.e., graphene or diamond, and completely disordered amorphous carbon (a-C). We solve the problem, i.e., systematic characterization of the structure of amorphous carbon and its functionalization, by utilizing ML based clustering<sup>4,23</sup> and a large data set of structures.<sup>24,25</sup> Here, the simulation outcomes are compared with experimental results to provide a more in-depth knowledge about the contribution

ACS Publications © 2019 American Chemical Society

Received:May 24, 2019Revised:October 26, 2019Published:October 28, 2019

of the different functional groups to the X-ray spectra of these materials. Our goal is to provide a quantitative estimate of the presence of functional groups on a-C surfaces based on fitting an experimental spectrum to a weighted combination of computational fingerprint spectra. In other words, when the fitting is carried out, we utilize the whole shape of the spectrum instead of certain peak positions as is traditionally done.<sup>26</sup> In addition, we explore the possibility of parallel fitting of XAS and XPS and discuss the advantages and disadvantages of the approach.

#### **II. SIMULATION DETAILS**

II.A. Functionalized Carbonaceous Surfaces. As explained in more detail in Part I of this study,<sup>4</sup> we use three different types of a-C models: (1) bulk samples constructed by geometry optimization;<sup>3,27</sup> (2) reconstructed surface samples that are cleaved from the bulk sample; and, in order to achieve a large enough sampling, (3) a library of a-C surfaces created by the melt-quench technique with a novel machine-learning (ML)-based interatomic potential.<sup>25</sup>. Ideal graphene, defective graphene, and reconstructed diamond surfaces are also used. In addition, all surfaces were functionalized with hydrogen (C-H), oxygen (C=O/C-O-C), hydroxyl (C-OH), and carboxylic acid (C-COOH) groups. Subsequently, we obtained the characteristic fingerprint XAS and XPS signals of pure carbon, carbon bonded to one of these groups, and corresponding oxygen K-edge spectra. In previous work,<sup>23</sup> the atomic motifs in pure a-C samples were clustered using a structural distance metric, employing the so-called "smooth overlap of atomic positions" (SOAP) kernel.4,23,28,29

In the present work, SOAP-based clustering is employed again, but this time we use multispecies descriptors (distinguishing between different atomic species, here, C, H, and O) to separate between different functionalities. Furthermore, we augment the structural metric, on which the clustering is based, with electronic structure information in order to reduce complexity, to identify overlapping features, and to aid in reconstruction and fitting of the experimental spectra. This will be discussed thoroughly in the next section.

II.B. Reclassifying Atomic Motifs for Quantitative Spectral Analysis. The data clustering technique used in our previous work was based solely on classification derived from structural features by utilizing a SOAP-based scheme.<sup>23</sup> However, we note that this may not be an optimal classification scheme in terms of grouping together sites with shared spectroscopic signatures. This is unsurprising since the information regarding XAS or XPS is not explicitly taken into account when classifying the atomic sites. Here we explore the possibility to improve this classification scheme for the specific task of fitting experimental spectra by incorporating some of this information, namely, site-resolved "delta-Kohn-Sham"  $(\Delta KS)$  values, which can be effectively used to computationally estimate core-electron binding energies.<sup>7,10,30</sup> For fitting purposes, we need to use an improved clustering scheme that satisfies three conditions: (1) it should recognize the structure/ chemistry of the site so that also hydrogen and oxygen are taken into account (i.e., we need to use a multispecies SOAP); (2) individual fingerprint spectra within the same cluster should have similar shapes; and (3)  $\Delta KS$  values of the sites in the same cluster should be similar. In addition, we need to consider oxygen K-edge spectra and carbon K-edge spectra separately.

SOAP descriptors characterize local atomic environments in a rotationally invariant way.<sup>28,29</sup> SOAP is therefore optimally suited to measure the similarity between two atomic sites. The atomic density of site *i* within a cutoff sphere is encoded into a SOAP vector  $q_i$ . The cutoff radius determines how far the SOAP algorithm can "see" around the site. A smearing parameter,  $\sigma_{\text{atom}}$ , defines how sharply the atomic densities of the neighboring atoms are resolved. Once the SOAP vectors for each site have been computed, the measure of similarity is given by the SOAP kernel, which is the dot product of two SOAP vectors, typically raised to a small positive number:

$$k^{\text{SOAP}}(i,j) = (\mathbf{q}_i \cdot \mathbf{q}_j)^{\zeta}$$
(1)

Typical values for  $\zeta$  are small integers, e.g.,  $\zeta = 4$ .<sup>24,31</sup> Increasing  $\zeta$  makes the kernel sharper (it emphasizes the differences between sites). The kernel is bound between 0 (totally dissimilar sites) and 1 (equivalent sites). From the kernel, a measure of distance can be defined as follows:

$$D^{\text{SOAP}}(i,j) = \sqrt{1 - k^{\text{SOAP}}(i,j)}$$
(2)

where D(i, j) varies inversely to the kernel, from 1 (totally dissimilar sites) to 0 (equivalent sites). This distance can be fed into machine-learning data clustering algorithms. More details of this approach are given in ref 23. Structural similarity measures can be complemented by augmenting the SOAP kernel by adding a Gaussian kernel based on  $\Delta KS$  values. This ensures that similar spectra are clustered together. A similar approach was employed in our previous work to predict adsorption energies.<sup>23</sup> Here, this augmented kernel achieves a compromise between the characteristics of the spectra and the chemistry (structure) of the sites: when similar (or overlapping) spectra are clustered together we lose information about the chemistry of the site. For instance, by ignoring atomic structure, carbon sites that are bonded to hydrogen would be clustered together with pure carbon sites, since the respective X-ray spectra cannot be told apart. On the other hand, without any connection to chemistry, fitting experimental spectra would be unavailing from the point of view of resolving the material nanostructure. Hence the need to retain structural information to be able to separate between functionalities arises. The combined kernel that we use,  $k^{\text{tot}}(i, j)$ *i*), is a linear combination of SOAP and  $\Delta$ KS kernels:

$$k^{\text{tot}}(i, j; x) = x k^{\text{SOAP}}(i, j) + (1 - x) k^{\Delta \text{KS}}(i, j)$$
(3)

where the  $\Delta KS$  kernel is defined as

$$k^{\Delta \mathrm{KS}}(i,j) = \exp\left(-\frac{(\Delta \mathrm{KS}_i - \Delta \mathrm{KS}_j)^2}{2\sigma_{\Delta \mathrm{KS}}^2}\right) \tag{4}$$

 $\Delta KS_i$  and  $\Delta KS_j$  are the  $\Delta KS$  values of sites *i* and *j*, respectively. From this kernel, distances are defined as follows:

$$D^{\text{tot}}(i,j) = \sqrt{1 - k^{\text{tot}}(i,j)}$$
(5)

All of the kernel hyperparameters, as well as x for the linear combination of kernels, eq 3, were optimized to achieve the best possible clustering with respect to the chemistry of the site and its spectral features. This was controlled by ensuring that similar spectra are clustered together and that the contents of the clusters agree with chemical intuition. This distance matrix is fed into the *k*-medoids<sup>32,33</sup> clustering algorithm, which then provides the data classification.



**Figure 1.** Workflow of the current approach: (1) we collect data (the structure,  $\Delta KS$ , and individual XAS spectra) from DFT calculations; (2) the data are clustered according to a multispecies SOAP kernel and a  $\Delta KS$  kernel; (3) we then obtain the average spectrum (XPS and XAS) of each cluster as the fingerprint of the cluster; (4) and finally, the experimental spectrum is fitted with these spectra.

#### **III. EXPERIMENTAL DETAILS**

In this work two different experimental a-C/ta-C samples are studied. Detailed descriptions of the film fabrication processes are given elsewhere.<sup>34,35</sup> Briefly, in the case of sample 1,<sup>34</sup> films were deposited on boron-doped (100) Si wafers (Ultrasil) using closed-field unbalanced magnetron sputtering in DC mode. The deposition rate and time were controlled to produce 100 nm thick a-C films. In the case of sample 2,<sup>35</sup> the films were deposited on boron-doped (100) Si wafers (Siegert). Direct current magnetron sputtering was used to deposit a 20 nm Ti adhesion layer on the Si wafer, followed by a 7 nm ta-C top layer.

NEXAFS data from both samples were acquired at a 55° incidence angle (magic angle) of X-ray incidence using the bending magnet beamline 8-2 at the Stanford Synchrotron Radiation Lightsource (SSRL). Beamline 8-2 is equipped with a spherical grating monochromator, operated using 40  $\times$  40  $\mu$ m<sup>2</sup> slits corresponding to a resolution of around 0.2 eV. The spot size at the interaction point was around  $1 \times 1 \text{ mm}^2$ , and a flux of  $10^{10}$  photons/s was used, at which beam damage is not noticeable even for extended exposure. The X-ray energies for the carbon 1s and oxygen 1s edges were scanned from 260 to 350 eV and 520 to 580 eV, respectively. The data were collected both in total electron yield (TEY) and Auger electron yield (AEY) modes using the drain current (amplified by a Keithley picoampmeter) and a cylindrical mirror analyzer (CMA) operated with a pass energy of 200 eV and set to record the main Auger line for the various edges, respectively. The incoming flux was recorded using a nickel grid with a Au sputtered film.

#### IV. RESULTS AND DISCUSSION

**IV.A. Reclassified Atomic Motifs.** Because oxygen and carbon sites differ significantly both in structure and in their spectroscopic signatures, clustering of carbon and oxygen sites is done following different recipes. An overview of the process is depicted in Figure 1. The work flow of the approach follows this logic: (1) the reference data are collected from DFT calculations (the structure,  $\Delta KS$ , and individual XAS spectra); (2) these data are clustered according to either a multispecies SOAP kernel (for the oxygen K-edge) or according to a multispecies SOAP kernel combined with a  $\Delta KS$  kernel (for the carbon K-edge); (3) the average, characteristic spectrum (XPS and XAS) of each cluster is obtained as the "fingerprint"

of the cluster; and (4) finally, the experimental spectrum is fitted with a weighted linear combination of these spectra.

To understand oxygen K-edge spectra, six clusters are sufficient to classify the sites and the corresponding fingerprint spectra (Table 1). Optimal oxygen clustering was achieved

Table 1. Representative Atomic Environments, As Obtained
from Clustering Analysis, in the Simulated a-C Samples with
Oxygen Functionalization, i.e., O K-Edge Clustering <sup>a</sup>

cluster	group	core
O <sub>(C-O-C)</sub>	ether/epoxide	С-О-С
O <sub>(C-COOH)</sub>	carboxylic acid	С-СООН
O <sub>(spC=O)</sub>	ketone ( <i>sp</i> site)	C=0
O <sub>(C-OH)</sub>	hydroxyl group	С-ОН
O <sub>(C-COOH)</sub>	carboxylic acid	C-COOH
$O_{(sp^2C=0)}$	ketone $(sp^2 \text{ site})$	C=0
<sup>a</sup> See Figure 2a for vis	ualization.	

with a multispecies SOAP kernel with a cutoff of 3 Å,  $\sigma_{\rm atom}$  = 0.3 Å and  $\zeta = 4$ , that is, without the need to incorporate the  $\Delta$ KS augmentation. Oxygen clustering with respect to K-edge spectra closely follows the chemistry of the sites. Our clusters are labeled according to the functional groups to which they pertain: O<sub>(C-O-C)</sub> epoxides/ethers, O<sub>(C-COOH)</sub> hydroxyl group in carboxylic acid,  $O_{(spC=O)}$  ketone bonded to a sp site,  $O_{(C-OH)}$  hydroxyl group,  $O_{(C-COOH)}$  ketone in carboxylic acid, and  $O_{(sp^2C=O)}$  ketone bonded to a  $sp^2$  site. We follow the convention that the O marked in bold face is the one where the core hole is placed. To graphically represent the data, we use multidimensional scaling (MDS), a dimensionality reduction method to visualize the similarity between individual data points in a certain data set.<sup>29,36</sup> The MDS representation of oxygen clustering, Figure 2a, shows how functional groups are clustered together and how similar/dissimilar the sites are compared to each other.

In the case of carbon the situation is more complicated, because our carbon samples not only consist of functionalized sites but also contain pure carbon sites at the surface and in the bulk. In order to address the structural complexity of a-C we



Figure 2. (a) Clustering of oxygen sites. (b) Clustering of carbon sites. The medoids of the clusters are marked with gray symbols. Functional groups and pure carbon sites are depicted next to their corresponding areas in the MDS scheme.

have utilized a large sampling of different carbon sites, bulk and surface. A 512-atom bulk sample was added to the database of surface samples to reproduce experimental conditions as realistically as possible, since even in surface-sensitive methods at least  $\sim 10$  Å depth from the sample's surface is scanned.<sup>26</sup> Even at these depths, a-C surface samples already resemble the bulk.<sup>37,38</sup> This means that, in experimental C K-edge data from a carbon sample, signals coming from the bulk are always present. By contrast, O K-edge data consist only of signals coming from functional groups, which are primarily at the surface, although the presence of some dissolved oxygen has also been suggested.<sup>3</sup> Thus, to obtain meaningful results, the carbon sampling needs to be significantly larger than for oxygen, which leads to a larger variety of spectra and  $\Delta KS$ values. In addition,  $\Delta KS$  values are more widely spread and the values for sites with different chemistries can sometimes overlap. Also some characteristics of the fingerprint XAS spectra are similar across atomic motifs, as has been discussed in the previous section. Because of this, the SOAP kernel alone was not sufficient to satisfy all three conditions that we required from a kernel to suit our purposes. Therefore, carbon sites were clustered with the augmented SOAP/ $\Delta$ KS kernel with the following parameters: SOAP cutoff of 2.5 Å,  $\sigma_{\text{atom}}$  = 0.2 Å, x = 0.55,  $\zeta = 4.5$ , and  $\sigma_{\Delta KS} = 1$  eV.

To understand experimental carbon K-edge spectra we found eight clusters, from which we obtain eight fingerprint spectra characteristic of the corresponding atomic motifs, to be sufficient (Table 2). Again, we follow the convention that the C atom that contains the core hole is marked in boldface. Carbon is clustered as follows:  $C_{(C-COOH)}$  for the second carbon in carboxylic acid;  $C_{(sp^2C_A)}$  for pure  $sp^2$  carbon, the carbon bonded to hydrogen, and carbon that binds to the carboxylic acid group;  $C_{(spC)}$  for pure sp carbon;  $C_{(C-COH)}$  for pure  $sp^3$  carbon, the carbon bonded to hydrogen, and carbon that binds to the carboxylic acid group;  $C_{(spC)}$  for pure sp carbon;  $C_{(cp-COH)}$  for pure  $sp^3$  carbon, the carboxylic acid group;  $C_{(C=O/C-OH)}$  for ketone and the hydroxyl group;  $C_{(sp^2C_B)}$  for pure  $sp^2$ , the

Table 2. As Table 1 but Now for Carbon Atoms in Different Environments, i.e., C K-Edge Clustering<sup>a</sup>

cluster	group	core
C <sub>(C-COOH)</sub>	carboxylic acid	С-СООН
$C_{(sp^2C_A)}$	plain <i>sp</i> <sup>2</sup> , hydrogen, carboxylic acid	С ( <i>sp</i> <sup>2</sup> ), С–Н, С–СООН
C <sub>(spC)</sub>	plain <i>sp</i> (very reactive)	<b>C</b> ( <i>sp</i> )
$C_{(sp^3C_A)}$	plain <i>sp</i> <sup>3</sup> , hydrogen, carboxylic acid	С ( <i>sp</i> <sup>3</sup> ), С–Н, С–СООН
C <sub>(C=O/C-OH)</sub>	ketone, hydroxyl	С=О, С-ОН
$C_{(sp^2C_B)}$	plain <i>sp</i> <sup>2</sup> , hydrogen, carboxylic acid	С ( <i>sp</i> <sup>2</sup> ), С–Н, С–СООН
$C_{(sp^{3}C_{B})}$	plain <i>sp</i> <sup>3</sup> , hydrogen, carboxylic acid	С ( <i>sp</i> <sup>3</sup> ), С–Н, С–СООН
C <sub>(C-O-C/C-OH)</sub>	ether/epoxide, hydroxyl group	С-О-С, С-ОН
Soo Figuro 2h	for visualization Note how	r plain or differently

"See Figure 2b for visualization. Note how plain or differently functionalized carbon sites are clustered together because of overlapping  $\Delta KS$  values.

carbon bonded to hydrogen, and carbon that binds to the carboxylic acid group;  $C_{(sp^{3}C_{B})}$  for pure  $sp^{3}$  carbon, the carbon bonded to hydrogen, and the carbon that binds to the carboxylic acid group; and  $C_{(C-O-C/C-OH)}$  for epoxide, ether, and the hydroxyl group. This carbon clustering scheme is depicted in Figure 2b, according to MDS representation. The MDS scheme reveals that even though it would at first seem more practical to cluster all  $sp^{2}$  sites or  $sp^{3}$  sites together forming two clusters, the sites in  $C_{(sp^{2}C_{A})}$  and  $C_{(sp^{2}C_{B})}$ , as well as sites in  $C_{(sp^{3}C_{A})}$  and  $C_{(sp^{3}C_{C}B)}$ , do differ with respect to their geometry and  $\Delta KS$  values and thus four clusters in total are needed to describe the variety of  $sp^{2}$  and  $sp^{3}$  sites. Unlike oxygen, where clustering leads to island-like graphical representation on the MDS plot (Figure 2a), for carbon there is a continuous distribution of sites, corresponding to the smooth transition between structural motifs observed in a- $C.^{23,25}$  The majority of the sites in a certain cluster are as listed

above, but there are some exceptions, for instance, some sp sites can be clustered together with  $sp^2$  sites because of overlapping  $\Delta KS$  values. We must point out that this is only one among many possible ways of how to cluster the carbon sites. Among several clustering schemes that we tested, this was found to be the best solution in terms of satisfying the three conditions mentioned before.

In the following we discuss how these classification schemes for C and O sites allow us to construct motif-resolved fingerprint spectra and use these spectra to fit experimental spectra.

**IV.B. Fingerprint Spectra of the Atomic Motifs.** Each of the motifs identified in the previous section has its own fingerprint spectrum associated with it. As discussed before, representative  $\Delta KS$  values could not be obtained if the clustering relied only on the geometrical features of the C sites. On the other hand, if clustering relied only on a  $\Delta KS$  kernel, because of overlapping values, we would lose all the connection to the chemistry of the site. To further complicate things, the contents of clusters  $C_{(sp^{3}C A)}$  and  $C_{(sp^{3}C B)}$  seem to be the same, but the sites are separated into two different clusters because they have different  $\Delta KS$  distributions. If we had clustered the spectra only according to chemical intuition (structure), this would have resulted in too many spectra for fitting to be practical, because some of them would overlap. These considerations highlight again the difficulty found, also experimentally, in the attribution of spectroscopic features of XAS spectra.

Based on the  $\Delta KS$  values presented in Part I,<sup>4</sup> it is evident that the  $\Delta$ KS values of the C–H and C–COOH cores overlap with the values of pure carbon cores. Also the characteristic features of the XAS spectra overlap with what can be seen from pure carbon spectra. Thus, clustering C-H and C-COOH together with pure carbon from bulk and surface samples is justified. It is unlikely that these two could be experimentally detected separately, since signals coming from pure carbon, which is present in much larger amounts, overshadow both XAS and XPS signatures coming from them. Recognizing overlapping characteristics seems indeed to be a problem when experimental XAS and XPS spectra are interpreted. Here, relying on our atomistic results, we are able to provide information about not only what can be seen in the spectra but also what cannot be detected, since we can trace back spectroscopic features directly to their atomic origin. The presence of other functionalities is much more easily resolved. For instance, C-COOH, which forms cluster  $C_{(C-COOH)}$ , has very clear and distinctive fingerprint characteristics. Hence, from those characteristics, C-COOH as a group present on the surface can be easily detected. Clusters  $C_{(C-COOH)}$ ,  $C_{(C=O/C-OH)}$ , and  $C_{(C-O-C/C-OH)}$  are the most interesting ones when functional groups are being studied and identified. Most importantly, we conclude that, in order to help in unfolding the complexity of overlapping signals in carbon Kedge spectra, one should study them together with oxygen Kedge spectra from the same sample.

**IV.C. Reconstruction of Experimental Spectra.** To reconstruct experimental spectra, we fit them to a weighted linear combination of our computed fingerprint spectra. For the fit, we use an in-house code that minimizes the following residual with respect to the fitting parameters:

$$R(E_0, \{a_i\}) = \int_{E_{\min}}^{E_{\max}} dE |S^{\exp}(E) - S^{\sin}(E; E_0, \{a_i\})|$$
(6)

where the simulated spectrum  $S^{sim}$  is given by

$$S^{\rm sim}(E; E_0, \{a_i\}) = a_0 + \sum_i a_i S_i^{\rm sim}(E - E_0)$$
(7)

The *i* index runs over all the fingerprint spectra used in the fit,  $a_0$  is a vertical shift to correct for a possible nonzero baseline in the experimental spectrum, and  $E_0$  is a horizontal shift to account for the systematic DFT underestimation of XPS and XAS energies.<sup>7</sup> The relative contribution of motif i to the fit, and therefore its abundance in the sample, can be computed from  $a_i / \sum_i a_i$ . The integration domain  $[E_{\min}, E_{\max}]$  determines the region where the fit is made. By determining the fitting range one can set the focus on certain desired features of the spectrum. Obviously, the choice of the range can have a large effect on the result and, thus, when the fitting is carried out several ranges should be tested to make sure that the chosen region is reasonable, i.e., that characteristic features associated with differently bonded carbon are reproduced well and not too much emphasis is put to the part of the spectrum corresponding to energy values beyond the ionization potential, i.e., when the electron is removed from the system in the experiment. The overlapping characteristic features of the fingerprint spectra that we have discussed above can lead to the existence of multiple optima in the fit. Overlapping characteristics cause inaccuracy in the predicted relative proportions of certain fingerprints, but we must keep in mind that when the fit is carried out in the traditional way (fitting with Gaussians peaks centered at certain energies instead of considering the shape of the whole spectrum), these overlaps are not taken into account at all. The only way to minimize this inherent arbitrariness in the fit is by reducing the number of degrees of freedom by carefully reducing the number of fingerprint spectra to the minimum, fitting XPS and XAS in parallel (as will be discussed below), or both.

Our implementation can minimize one residual, XAS or XPS independently, or a linear combination of XAS and XPS residuals, where the  $a_i$  are optimized so that both fits are satisfactory simultaneously for the same values of  $a_i$ . To obtain optimal fitting parameters, we randomly sample within predetermined intervals using a Monte Carlo approach. The best fit among all tested combinations (that is, the combination of fitting parameters that leads to the lowest residual) is chosen. The fitting procedure is first demonstrated by fitting experimental graphite oxide and diamond samples. To fit the spectrum of graphite oxide, a highly functionalized form of graphite, we use our graphene and functionalized graphene computational fingerprint spectra as references.<sup>4</sup> Then we employ the fitting scheme to interpret experimental XAS and XPS spectra, both carbon and oxygen K-edges, of an a-C sample. Our in-house fitting code and data sets required to run it will be made publicly available in the near future, after we have improved the user interface; in the meantime, reasonable requests for code and/or data will be accommodated by the authors whenever possible.

In the case of amorphous carbon, we can fit the experimental XAS and XPS spectra, either carbon or oxygen K-edge, with a weighted combination of the average fingerprint spectra representing each cluster discussed before. Experimentally, the fitting of the spectra is done one spectrum at a time, that is, XAS and XPS separately. This can give rise to XAS and XPS fits pointing to a differing chemistry for the same sample. To avoid this contradictory result, here we aim at using all the information available, and thus we fit XAS and XPS spectra in



**Figure 3.** (a) Fingerprint C K-edge spectra that are used in fitting the XAS spectrum of graphite oxide. (b) Calculated graphene C K-edge spectrum compared with the experimental graphite oxide spectrum (from ref 39). The calculated ideal  $sp^2$  spectrum (from ref 40) is also plotted to show the difference between pristine and oxidized samples. The fitting range,  $[E_{min}, E_{max}]$ , is indicated with vertical dashed lines. (c) Fingerprint C K-edge spectra that are used for diamond fitting. (d) Calculated diamond C K-edge spectrum compared with experimental diamond spectrum (from ref 41). The calculated ideal  $sp^3$  spectrum (from ref 40) is plotted to show the difference between the pristine and the oxidized samples. Again, the fitting range,  $[E_{min}, E_{max}]$ , is indicated with dashed lines. Since the experimental graphite oxide and diamond spectra are taken from different studies, the energy resolution of the measurement equipment differs slightly (0.1 and 0.05 eV, respectively). This difference does not have a large impact on the fitting procedure, since the overall experimental signal broadening (which also includes thermal and disorder effects) and the corresponding smearing parameter used in the fit are significantly larger.

parallel, whenever possible. More importantly, we also compare the results for the carbon and oxygen K-edge spectra to each other. Unsurprisingly, if the XAS and XPS spectra are fitted separately, a better fit can be obtained since more degrees of freedom (fitting parameters) are available. However, the results obtained in this fashion may lead to contradictory interpretations of the material's chemistry, as already discussed. In any case, knowing that XAS is more sensitive than XPS and that in XAS fitting we can use the cross sections for a continuum of possible transitions, more emphasis is put on fitting of the XAS spectra. Thus, even though XAS and XPS spectra can be fitted in parallel, we choose the result of the XAS fit to weight more than the result of the XPS fit.

Even though our fitting code gives percentages for the contribution of each spectrum that has been included in the fitting, these numbers should not be directly used as exact quantitative measures. They can, however, give approximate estimates of the proportions between different motifs present in the experimental samples. It is important to note that our current simulations consist only of the four most important functional groups anticipated to exist on carbonaceous surfaces. The chemistry between oxygen and carbon is far more complicated than that and, in addition, a-C samples are known to also contain small amounts of other elements, such as nitrogen ( $\sim$ 1 atom %),<sup>42</sup> which is not included in our

simulations. Here, our aim is to recognize and quantify the groups that have the largest effect on the chemical characteristics of the surface.

Computationally, we can resolve the precise transition energy between a core state and a conduction band state at 0 K. Considerations when attempting numerical fitting of experimental data include (1) the resolution of the instrument used in the experimental measurements, (2) the variety of atomic environments of the element in question (which in this work has been included by sampling over a variety of structures), and (3) the temperature of the sample, which causes thermal motion of the atoms and, in turn, broadening of the spectra. These points are taken into account by introducing a Gaussian smearing of 0.5 eV of the calculated spectra. This value roughly corresponds to the broadening of a typical experimental XPS spectrum for a pure gas-phase sample (which we have estimated from the full width at halfmaximum).<sup>43</sup> Pure molecular gas-phase samples are free from contaminants that would be present in bulk and surface experiments and, thus, can offer unbiased insight into the resolution of the instrument in question.

**IV.D. Fitting Diamond and Graphene Spectra: Benchmarking the Fitting Procedure.** In this section, as an introduction to the fitting scheme, we show how the experimental spectra are reconstructed and fitted by using the



**Figure 4.** (a) Calculated a-C O K-edge XAS fingerprint spectra for every oxygen cluster. (b) Calculated a-C O K-edge XAS spectra compared with two different experimental spectra. The fitting range,  $[E_{min}, E_{max}]$ , is indicated with dashed lines. (c) Calculated a-C C K-edge XAS fingerprint spectra for every carbon cluster (the clusters are named in panel (e)). (d) Calculated a-C C K-edge XAS spectra compared to two different experimental spectra (for the legend, see panel (f)). Again, the fitting range,  $[E_{min}, E_{max}]$ , is indicated with dashed lines. The fitting ranges differ between samples in order to minimize the fitting error. (e) Calculated normalized a-C C K-edge XPS fingerprint spectra for every carbon cluster. (f) Calculated a-C C K-edge XPS spectra compared to two different experimental spectra.

fingerprint spectra as building blocks (Figure 3). In this study, instead of only looking at peak positions that are traditionally used in fitting, we examine the whole spectrum. Furthermore, some groups show distinctive characteristics only in the C K-edge spectrum or in the O K-edge spectrum, not necessarily in both, and, thus, these two should always be investigated together. To benchmark this procedure and to compare different types of carbon surfaces, we have calculated and fitted spectra for experimental diamond and graphite oxide samples (from refs 41 and 39, respectively). As discussed, the calculated spectrum is a weighted sum of fingerprint spectra for a set of atomic motifs.

In the case of pristine diamond surfaces and graphene, all atomic sites are equivalent to one another. Real samples contain a number of defects, impurities, and functionalizations. For instance, the structure and the degree of oxidation of graphene, or graphene oxide, varies substantially depending on the origin of the sample.44-46 Nevertheless, graphite and diamond samples still display the strong  $sp^2$  or  $sp^3$  characteristics, respectively, from the pristine samples. Thus, the pure carbon reference spectra are obtained from the ideal  $sp^2$ (graphene) spectrum and a combination of  $sp^2$  site spectra around the single vacancy defect of graphene. Introducing a single vacancy into the structure creates a favorable site for functional groups to bind. For diamond, we sample  $sp^3$  and  $sp^2$ sites present in reconstructed diamond surfaces. The fingerprint spectra that are used in the fit are depicted in Figure 3a,c. Note that here, in both cases, we include also the spectra of carbon bonded to hydrogen, which in the case of amorphous carbon has overlapping characteristics with pure carbon sites.<sup>4</sup> It has been previously shown that hydrogen bonded to carbon in graphene shifts  $\Delta KS$  values compared to pristine carbon, but the values can overlap with defective carbon sites.<sup>6</sup> Moreover, the calculated  $\Delta KS$  values used in this work to determine the onset of the XAS spectra are in very good agreement with the values presented by Susi et al. in ref 6.

The simulated XAS spectra are compared to the experimental ones in Figure 3b,d. It is evident that the experimental spectra are very well reproduced. The fit is not perfect, since (1) our samples are simplified representations of a far more complex reality, where real samples actually contain a number of different kinds of defects, surface steps/terraces, etc., and (2) our DFT-based methodology to compute the reference spectra has inherent inaccuracies associated with it. The  $\Delta KS$  values, which are used to interpret XPS spectra and provide the energy-scale alignment of calculated XAS spectra, are systematically underestimated (approximately by 1 eV in the case of carbon). Nevertheless, relative energies are well reproduced.<sup>6,7</sup> In addition,  $\Delta KS$  energies (and, to a lesser degree, intensities in XAS calculations) depend on the exchange-correlation density functional used in the DFT calculations.<sup>7,8,47-49</sup> In this study we use the PBE level of theory, which is a reasonable choice with large data sets of periodic systems and has been shown to perform well for carbon-based materials.<sup>6-8,49</sup> In other words, the overall shapes of the spectra are correctly reproduced, and we can confidently

identify where certain characteristic peaks originate from. It must be emphasized that the shapes of the experimental spectra would not be correctly reproduced if we were using pure carbon reference spectra only (also given in Figure 3b,d). The characteristics observed in experimental spectra can only be reproduced by introducing and utilizing the fingerprint spectra of carbons that are bonded to specific functional groups.

We will not attempt to give quantitative estimates about the contents of the samples, but we can give a semiguantitative evaluation based on the fingerprint spectra. In this graphite oxide sample<sup>39</sup> the amount of pure  $sp^2$  sites exceeds the amount of defected sites. According to our simulations, there is hydrogen, oxygen as ketones, hydroxyl and carboxylic acid groups, and, finally, oxygen as epoxides present in the sample, in that order. Unsurprisingly, the  $sp^3$  content of the diamond sample<sup>41</sup> is high compared to  $sp^2$  bonded carbon, which is also present. The rest of the carbon sites in the sample are likely to be bonded to a functional group. The functional groups involved are oxygen, hydrogen, hydroxyl, and carboxylic acid groups, in that order. Oxygen-induced carbon features, in the case of diamond, seem to overlap with well-known diamond exciton characteristics (which cannot be predicted within our methodological framework<sup>50</sup>), and thus in our simulations the oxygen content is exaggerated. On the other hand, because of this overlap, the presence of oxygen can stay hidden when diamond spectra are interpreted traditionally.

**IV.E. Fitting a-C Spectra: A Detailed Analysis.** Having established that our methodology can extract semiquantitative chemical information from the benchmark simulations of diamond surfaces and graphene, we turn our attention to the far more complicated situation of interpreting experimental a-C spectra. Furthermore, in order to better establish the performance of the method, we fit experimental data from two different a-C samples separately. The fingerprint spectra of the motifs discussed in Section IV.B are depicted in Figure 4a,c (XAS) and Figures 4e and 5 (XPS). The fitted XAS oxygen and carbon K-edge XAS spectra for a-C are presented in Figure 4f.



Figure 5. Normalized fingerprint XPS spectra of the oxygen clusters compared with the experimental spectrum. The image clearly shows where certain clusters contribute on the binding energy scale.

We will first discuss the fitting of the oxygen K-edge spectra. Since both oxygen atoms in the C–COOH group, i.e., clusters  $O_{(C-COOH)}$  and  $O_{(C-COOH)}$ , should give signals in the same proportion, the calculated spectra of the two are combined for the fitting of the oxygen spectra. Clusters  $O_{(spC=O)}$  and  $O_{(C-COOH)}$  are the major contributors forming a peak at ~532

eV in the XAS spectrum. Both clusters  $O_{\rm (C-COOH)}$  and  $O_{\rm (C-OH)}$  give a distinctive peak nearly in the same region of the XAS spectrum (~535–536 eV), which in the experimental spectrum can be seen only as a change of slope or a wide shoulder. However, these two groups give a large contribution to the XPS spectrum in the stronger binding energy regions, and thus combined fitting of XAS and XPS overemphasizes this feature.

In order to explain this observation, a literature search was carried out. Even though peak positions used in deconvolution of the experimental spectra of carbonaceous substances that are bonded to oxygen fluctuate depending on the reference, O K-edge spectra are often fitted with three peaks. These peaks are usually attributed to (1) ~531 eV double-bonded oxygen, (2) ~533 eV hydroxyl or ether groups, and (3) ~534 eV single-bonded oxygen in carboxylic acid or ester groups (or chemisorbed water which is not included in our simulations).<sup>51–55</sup> The ordering of these peaks is in good agreement with our calculations. The calculated O K-edge XPS spectrum is made out of three components as well: (1) C–COOH, (2)hydroxyl group with overlapping ethers and epoxides, and (3) C-COOH, in that order. Peak 1, arising from the doublebonded carbon, comes from our C-COOH, i.e., from the double-bonded oxygen in carboxylic acid group. It is known based on the literature that in molecular or polymer references it is practically impossible to tell apart a ketone from another double-bonded oxygen experimentally.53 In fact, our calculations show that signals coming from ketones that are bonded to an a-C surface are widely distributed in energy because of the varying local atomic environments. These signals then result in a very wide peak near the onset of the spectra. Carboxylic acid groups, that stand higher from the surface, are less affected by the surface itself, and, thus, the calculated signals are in better agreement with molecular or polymer references from experiment. In other words, XPS does not capture well signals coming from ketones that are bonded to the a-C surface and the presence of hydroxyl and carboxylic acid groups is significantly overemphasized. In Figure 5 the fingerprint spectra of oxygen motifs are depicted together with the experimental spectra from both samples. However, in XAS fitting we can use the full spectrum, which allows us to detect also ketones that are practically hidden around the onset of the XPS spectra.

Hence, in the case of O K-edge we settle for fitting XAS spectra and XPS spectra separately, and we focus on XAS fitting in order to find better agreement with carbon K-edge spectra fitting, which will be discussed later. Most of the ketones present in the fit belong to the cluster  $O_{(spC=O)}$ , which is expected since the oxygen atom in a ketone bonded to an  $sp^2$ site  $(O_{(sp^2C=0)})$  is over coordinated. Even though the existence of these groups is observed in our simulations, these sites are quite reactive. There are also some epoxides or ethers  $(O_{(C-O-C)})$  present on the surface, which are not observed in carbon K-edge fitting due to the fact that the signals coming from these sites, both XAS and XPS, overlap with signals coming from other groups. The results from the fitting suggest that ethers and epoxides contribute in sample  $1 \sim 4-5\%$  and in sample 2  $\sim$ 11–12% to the overall spectrum. Ketones (sample 1 ~36% and sample 2 ~39%), the hydroxyl group (sample 1 ~26% and sample 2 ~14%) and carboxylic acid (sample 1 ~34% and sample 2 ~36%) are the main contributors to the oxygen K-edge XAS spectra of the a-C sample studied. However, one should be careful about how to interpret these



Figure 6. (a) Independently fitted a-C C K-edge XAS spectra. (b) Experimental a-C C K-edge XPS spectra fitted with fingerprint spectra weighted by independent XAS fitting.

percentages, in the context of the current overall accuracy of the method. Therefore, using these percentages as guidelines, we are confident to claim that these groups are present on the surfaces of both samples and that they each account for, roughly, one-third of the total oxygen-containing groups, but that said, the two samples have different compositions.

Inspection of the carbon K-edge spectra confirms the presence of these groups. To test our method further we will fit carbon K-edge spectra according to two different schemes. First, during the fitting of the carbon spectra, XAS and XPS spectra are fitted in parallel. Unlike with oxygen, all the carbon cluster XPS fingerprints show a non-negligible overlay with the experimental peak (Figure 4e). We have two more C core reference fingerprint spectra to fit than with oxygen. However, when we carry out parallel XAS and XPS fitting, the complexity of the problem is reduced. It can be seen from Figure 4d,f that, even though the XAS signals of the two samples differ significantly, the XPS signals are actually very similar. If the XAS and XPS spectra were fitted separately, obviously, better fits can be produced, but we choose to fit them together first in order to obtain consistent results from both measurements. This means compromising the accuracy of the XAS fit, but we settle for that because although again XAS captures signals more accurately than XPS, in the case of carbon, these signals may very well be coming from something that our model does not cover (such as carbon bonded to nitrogen). In this study we focus mainly on oxygen containing functional groups. We consider parallel fitting of XAS (weight 70%) and XPS (weight 30%) to be a safe choice in this case. However, it should again be kept in mind that the accuracy of the method is limited. This is specially true when one considers the much larger contribution from pure carbon motifs to the carbon K-edge spectrum. Based on parallel fitting it seems that approximately 6% of carbon in sample 1 and 4% of carbon in sample 2 is bonded to oxygen and that carboxylic acid, ketones, and the hydroxyl group are present roughly in the same proportions, making 1-2% each. In carbon K-edge spectra, signals coming from epoxides and ethers overlap with those coming from other motifs. According to the interpretation of oxygen K-edge spectra, epoxides and, especially, more stable ethers are also likely to be present. However, the characteristic signals arising from them are easily hidden by other oxygen-containing groups or even by pure carbon. Overlapping characteristics are precisely what limits quantitative analysis of XAS and especially XPS spectra.

As discussed in the previous section and in more detail in Part I of this study,<sup>4</sup> carbon bonded to hydrogen cannot be

separated from pure carbon by means of analyzing X-ray spectroscopy data. Motifs containing pure carbon, carbon bonded to hydrogen, and C-COOH form the majority of the sites, according to the results of the fits. When XAS and XPS spectra are fitted in parallel the estimated  $sp^2/sp^3$  ratio of both experimental samples studied here is quite high. According to the fitting results from parallel fitting, we estimate that approximately 40% of the C sites are  $sp^2$  hybridized and 50% are  $sp^3$  hybridized (including C and H neighbors). A small amount,  $\sim$  2%, of reactive and surface-specific sp-bonded carbon is also present. Amorphous carbon surfaces are known to be  $sp^2$ -rich and the bulk is more  $sp^3$ -rich. These results are in good agreement with recent ta-C growth simulations.<sup>37</sup> Sample 1 has a slightly higher  $\pi^*$  peak around 285 eV, but it must be kept in mind that also C-H and C-COOH contribute in that region. We therefore avoid drawing direct conclusions from the height of the peak only. However, when fitting is carried out in parallel this feature is not well reproduced since the XPS spectra of the samples are so similar (Figure 4).

In order to study the  $sp^2/sp^3$  ratio of the samples more carefully, we also fit the XAS spectra independently from XPS. The results of independent XAS fitting are shown in Figure 6. The calculated XPS spectra, interpolated with the weights obtained for the XAS fit, are depicted in Figure 6b. Interestingly, especially around the tails of the spectra, the XPS fit is qualitatively improved with respect to the parallel XAS/XPS fit result. Even though there are less fitting degrees of freedom in the parallel fit compared to the independent fit, it should be borne in mind that, as a structural analysis method, XAS is significantly more accurate than XPS, also when the measurements are carried out at the same time with the same high resolution equipment. This advantage of XAS versus XPS is also reflected on the computational results: when XAS spectra are fitted independently the differences between the samples stand out. The  $sp^2/sp^3$  ratio of sample 1 (2.3) is higher than the  $sp^2/sp^3$  ratio of sample 2 (0.8), according to the results of the fit. Small amounts, ~1%, of sp-bonded carbon remain present in both samples. According to the independent XAS-based fits, there is almost twice as much oxygen containing groups in sample 1 (~18%) as in sample 2 (~10%). In sample 1 ketones are present in large numbers (C-COOH ~ 3%, C=O ~ 12%, and C-OH ~ 4%), whereas in sample 2 there are more hydroxyl groups present (C-COOH ~ 2%, C=O ~ 2%, and C-OH ~ 6%). Nevertheless, our results strongly suggest that, when trying to establish the amount of carbon bonded to oxygen, one should look at the carbon K-edge spectra. By contrast, when the interest is in finding out relative proportions between oxygen-containing functional groups, oxygen K-edge spectra become more useful. All in all, we conclude on the basis of our analysis and fitting of the spectra, that oxygen is present on a-C surfaces predominantly as a ketone, as a part of carboxylic acid group or as part of a hydroxyl group.

# V. CONCLUSIONS

In this study we have (1) introduced a new way to classify functionalized carbonaceous surface atomic motifs by utilizing a combination of structural descriptors (multispecies SOAP kernel) and electronic descriptor (Gaussian  $\Delta KS$  kernel); (2) proposed a systematic fitting scheme for deconvolution of experimental XAS and XPS spectra, used here to understand the X-ray spectroscopic features of carbon-based materials, but generally applicable to other material systems; (3) benchmarked this fitting procedure by fitting experimental graphite oxide and diamond spectra; (4) fitted experimental a-C XAS and XPS spectra, either in parallel or separately, depending on the situation; and finally (5) combined O K-edge data with C K-edge data in order to provide semiguantitative estimates of the composition of experimental samples based on fitting to computational (ab initio) references. Thanks to the detailed knowledge about the microscopic origin of these spectroscopic signatures, i.e., the "fingerprint spectra", the present methodology allows us to not only identify what can be directly detected from the experimental spectra but also propose and depict an atomic-level picture of these materials.

Given the current resolution of XPS, we conclude that for highly complex and/or disordered materials XAS provides considerably more insight into the atomic-level structure of the sample under study, both experimentally and in simulation. As a matter of fact, the amount of quantitative, and even qualitative, information that can be extracted from XPS for these materials is very limited, since the signals from local atomic environments overlap. However, being a more widely available experimental technique, XPS analysis is more easily accessible than XAS and often XAS characterization is not carried out. In such cases, care must be exercised when attempting to extract quantitative structural information from XPS data. Despite that, the present fitting scheme based on computational references allows us to perform a more informed XPS fit compared to those relying on molecular references. Nonetheless, we emphasize that XAS should be chosen over XPS whenever both techniques are available.

#### AUTHOR INFORMATION

#### **Corresponding Author**

\*(A.A.) E-mail: anja.aarva@aalto.fi.

#### ORCID 💿

Anja Aarva: 0000-0002-0207-3145 Volker L. Deringer: 0000-0001-6873-0278 Sami Sainio: 0000-0002-9268-0124 Tomi Laurila: 0000-0002-1252-8764 Miguel A. Caro: 0000-0001-9304-4261

# **Present Address**

<sup>§</sup>(V.L.D.) Department of Chemistry, University of Oxford, Oxford OX1 3QR, United Kingdom.

#### Notes

The authors declare no competing financial interest.

# ACKNOWLEDGMENTS

Funding from the Academy of Finland (Project No. 285526) and the computational resources provided for this project by CSC - IT Center for Science are gratefully acknowledged. M.A.C. acknowledges personal funding from the Academy of Finland under Project No. 310574. V.L.D. acknowledges a Leverhulme Early Career Fellowship and support from the Isaac Newton Trust. M.A.C. and V.L.D. are grateful for traveling support from the HPC-Europa3 program under the auspices of the European Union's Horizon 2020 framework (Grant Agreement No. 730897). Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-76SF00515. S.S. acknowledges personal funding from Instrumentarium Science Foundation and the Walter Ahlström Foundation.

# ■ REFERENCES

(1) Robertson, J. Diamond-like amorphous carbon. *Mater. Sci. Eng.*, R 2002, 37, 129–281.

(2) Tiainen, V.-M. Amorphous carbon as a bio-mechanical coating—mechanical properties and biological applications. *Diamond Relat. Mater.* **2001**, *10*, 153–160.

(3) Laurila, T.; Sainio, S.; Caro, M. A. Hybrid carbon based nanomaterials for electrochemical detection of biomolecules. *Prog. Mater. Sci.* **2017**, *88*, 499–594.

(4) Aarva, A.; Deringer, V. L.; Sainio, S.; Laurila, T. Understanding X-ray Spectroscopy of Carbonaceous Materials by Combining Experiments, Density Functional Theory, and Machine Learning. Part I: Fingerprint Spectra. *Chem. Mater.* **2019**, DOI: 10.1021/acs.chemmater.9b02049.

(5) Barinov, A.; Malcioglu, O. B.; Fabris, S.; Sun, T.; Gregoratti, L.; Dalmiglio, M.; Kiskinova, M. Initial stages of oxidation on graphitic surfaces: photoemission study and density functional theory calculations. *J. Phys. Chem. C* **2009**, *113*, 9009–9013.

(6) Susi, T.; Kaukonen, M.; Havu, P.; Ljungberg, M. P.; Ayala, P.; Kauppinen, E. I. Core level binding energies of functionalized and defective graphene. *Beilstein J. Nanotechnol.* **2014**, *5*, 121–132.

(7) Susi, T.; Mowbray, D. J.; Ljungberg, M. P.; Ayala, P. Calculation of the graphene c 1 s core level binding energy. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *91*, No. 081401.

(8) Susi, T.; Scardamaglia, M.; Mustonen, K.; Tripathi, M.; Mittelberger, A.; Al-Hada, M.; Amati, M.; Sezen, H.; Zeller, P.; Larsen, A. H.; Mangler, C.; Meyer, J. C.; Gregoratti, L.; Bittencourt, C.; Kotakoski, J. Intrinsic core level photoemission of suspended monolayer graphene. *Physical Review Materials* **2018**, *2*, No. 074005.

(9) Scardamaglia, M.; Susi, T.; Struzzi, C.; Snyders, R.; Di Santo, G.; Petaccia, L.; Bittencourt, C. Spectroscopic observation of oxygen dissociation on nitrogendoped graphene. *Sci. Rep.* **2017**, *7*, 7960.

(10) Walter, M.; Moseler, M.; Pastewka, L. Offsetcorrected  $\delta$ -kohnsham scheme for semiempirical prediction of absolute X-ray photoelectron energies in molecules and solids. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *94*, No. 041112.

(11) Ozaki, T.; Lee, C.-C. Absolute binding energies of core levels in solids from first principles. *Phys. Rev. Lett.* **2017**, *118*, No. 026401.

(12) Walter, M.; Mangolini, F.; McClimon, J. B.; Carpick, R. W.; Moseler, M. Origin of carbon 1s binding energy variance in diamond. arXiv preprint arXiv:1902.02958, 2019 (accessed Oct. 25, 2019).

(13) Kahk, J. M.; Lischner, J. Accurate absolute core-electron binding energies of molecules, solids and surfaces from first-principles calculations. arXiv preprint arXiv:1904.04823, 2019 (accessed Oct. 25, 2019).

(14) Soldemo, M.; Vandichel, M.; Grönbeck, H.; Weissenrieder, J. Initial  $Fe_3O_4(100)$  formation on Fe(100). J. Phys. Chem. C **2019**, 123, 16317–16325.

(15) Öström, H.; Ogasawara, H.; Näslund, L.-Å.; Pettersson, L. G. M.; Nilsson, A. Physisorption-induced CH bond elongation in methane. *Phys. Rev. Lett.* **2006**, *96*, 146104.

(16) Schiros, T.; Haq, S.; Ogasawara, H.; Takahashi, O.; Öström, H.; Andersson, K.; Pettersson, L. G. M.; Hodg-son, A.; Nilsson, A. Structure of water adsorbed on the open Cu (1 1 0) surface: H-up, Hdown, or both? *Chem. Phys. Lett.* **2006**, *429*, 415–419.

(17) Schiros, T.; Nordlund, D.; Pálová, L.; Prezzi, D.; Zhao, L.; Kim, K. S.; Wurstbauer, U.; Gutiérrez, C.; Delongchamp, D.; Jaye, C.; Fischer, D.; Ogasawara, H.; Pettersson, L. G. M.; Reichman, D. R.; Kim, P.; Hybertsen, M. S.; Pasupathy, A. N. Connecting dopant bond type with electronic structure in n-doped graphene. *Nano Lett.* **2012**, *12*, 4025–4031.

(18) Mathew, K.; Zheng, C.; Winston, D.; Chen, C.; Dozier, A.; Rehr, J. J.; Ong, S. P.; Persson, K. A. High-throughput computational X-ray absorption spectroscopy. *Sci. Data* **2018**, *5*, 180151.

(19) Zheng, C.; Mathew, K.; Chen, C.; Chen, Y.; Tang, H.; Dozier, A.; Kas, J. J.; Vila, F. D.; Rehr, J. J.; Piper, L. F. J.; Persson, K. A.; Ong, S. P. Automated generation and ensemble-learned matching of X-ray absorption spectra. *npj Computational Materials* **2018**, *4*, 12.

(20) Zhovtobriukh, I.; Norman, P.; Pettersson, L. G. M. X-ray absorption spectrum simulations of hexagonal ice. *J. Chem. Phys.* **2019**, *150*, No. 034501.

(21) Zhovtobriukh, I.; Cabral, B. J. C; Corsaro, C.; Mallamace, D.; Pettersson, L. G. M. Liquid water structure from X-ray absorption and emission, NMR shielding and X-ray diffraction. *Sci. China: Phys., Mech. Astron.* **2019**, *62*, 107010.

(22) Wikfeldt, K. T.; Leetmaa, M.; Mace, A.; Nilsson, A.; Pettersson, L. G. M. Oxygen-oxygen correlations in liquid water: Addressing the discrepancy between diffraction and extended X-ray absorption fine-structure using a novel multiple-data set fitting technique. *J. Chem. Phys.* **2010**, *132*, 104513.

(23) Caro, M. A.; Aarva, A.; Deringer, V. L.; Csányi, G.; Laurila, T. Reactivity of amorphous carbon surfaces: rationalizing the role of structural motifs in functionalization using machine learning. *Chem. Mater.* **2018**, *30*, 7446–7455.

(24) Deringer, V. L.; Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *95*, No. 094203.

(25) Deringer, V. L.; Caro, M. A.; Jana, R.; Aarva, A.; Elliott, S. R.; Laurila, T.; Csányi, G.; Pastewka, L. Computational surface chemistry of tetrahedral amorphous carbon by combining machine learning and density functional theory. *Chem. Mater.* **2018**, *30*, 7438–7445.

(26) Stöhr, J. NEXAFS spectroscopy; Springer Science & Business Media: 2013; Vol. 25.

(27) Caro, M. A.; Zoubkoff, R.; Lopez-Acevedo, O.; Laurila, T. Atomic and electronic structure of tetrahedral amorphous carbon surfaces from density functional theory: Properties and simulation strategies. *Carbon* **2014**, *77*, 1168.

(28) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 184115.

(29) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.

(30) Nilsson, A.; Pettersson, L. G. M. Chemical bonding on surfaces probed by X-ray emission spectroscopy and density functional theory. *Surf. Sci. Rep.* **2004**, *55*, 49–167.

(31) Caro, M. A. Optimizing many-body atomic descriptors for enhanced computational performance of machine learning based interatomic potentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2019**, *100*, No. 024112.

(32) Caro, M. A. Fork of C. Bauckhage's k-Medoids Python implementation for enhanced medoid initialization. https://github. com/mcaroba/kmedoids (accessed Oct. 25, 2019).

(33) Bauckhage, C. Numpy/scipy Recipes for Data Science: k-Medoids Clustering; University of Bonn: 2015.

(34) Palomäki, T.; Caro, M. A.; Wester, N.; Sainio, S.; Etula, J.; Johansson, L.-S.; Han, J. G.; Koskinen, J.; Laurila, T. Effect of power density on the electrochemical properties of undoped amorphous carbon (a-c) thin films. *Electroanalysis* **2019**, *31*, 746–755.

(35) Palomäki, T.; Wester, N.; Caro, M. A.; Sainio, S.; Protopopova, V.; Koskinen, J.; Laurila, T. Electron transport determines the electrochemical properties of tetrahedral amorphous carbon (ta-c) thin films. *Electrochim. Acta* **2017**, *225*, 1–10.

(36) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(37) Caro, M. A.; Deringer, V. L.; Koskinen, J.; Laurila, T.; Csányi, G. Growth mechanism and origin of high sp3 content in tetrahedral amorphous carbon. *Phys. Rev. Lett.* **2018**, *120*, 166101.

(38) Davis, C. A.; Amaratunga, G. A. J.; Knowles, K. M. Growth mechanism and cross-sectional structure of tetrahedral amorphous carbon thin films. *Phys. Rev. Lett.* **1998**, *80*, 3280.

(39) Zhan, D.; Ni, Z.; Chen, W.; Sun, L.; Luo, Z.; Lai, L.; Yu, T.; Wee, A. T. S.; Shen, Z. Electronic structure of graphite oxide and thermally reduced graphite oxide. *Carbon* **2011**, *49*, 1362–1366.

(40) Sainio, S.; Nordlund, D.; Caro, M. A.; Gandhiraman, R.; Koehne, J.; Wester, N.; Koskinen, J.; Meyyappan, M.; Laurila, T. Correlation between sp3 -to-sp2 ratio and surface oxygen functionalities in tetrahedral amorphous carbon (ta-C) thin film electrodes and implications of their electrochemical properties. *J. Phys. Chem. C* 2016, *120*, 8298–8304.

(41) Shpilman, Z.; Gouzman, I.; Minton, T. K.; Shen, L.; Stacey, A.; Orwa, J.; Prawer, S.; Cowie, B. C. C.; Hoffman, A. A near edge X-ray absorption fine structure study of oxidized single crystal and polycrystalline diamond surfaces. *Diamond Relat. Mater.* **2014**, *45*, 20–27.

(42) Palomäki, T.; Chumillas, S.; Sainio, S.; Protopopova, V.; Kauppila, M.; Koskinen, J.; Climent, V.; Feliu, J. M.; Laurila, T. Electrochemical reactions of catechol, methylcatechol and dopamine at tetrahedral amorphous carbon (ta-C) thin film electrodes. *Diamond Relat. Mater.* **2015**, *59*, 30–39.

(43) Prosvirin, I. P.; Bukhtiyarov, A. V.; Bluhm, H.; Bukhtiyarov, V. I. Application of near ambient pressure gasphase X-ray photoelectron spectroscopy to the investigation of catalytic properties of copper in methanol oxidation. *Appl. Surf. Sci.* **2016**, *363*, 303–309.

(44) Loh, K. P.; Bao, Q.; Ang, P. K.; Yang, J. The chemistry of graphene. J. Mater. Chem. 2010, 20, 2277-2289.

(45) Loh, K. P.; Bao, Q.; Eda, G.; Chhowalla, M. Graphene oxide as a chemically tunable platform for optical applications. *Nat. Chem.* **2010**, *2*, 1015.

(46) Dreyer, D. R.; Park, S.; Bielawski, C. W.; Ruoff, R. S. The chemistry of graphene oxide. *Chem. Soc. Rev.* **2010**, *39*, 228–240.

(47) Triguero, L.; Pettersson, L. G. M.; Ågren, H. Calculations of near-edge X-ray-absorption spectra of gasphase and chemisorbed molecules by means of densityfunctional and transition-potential theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 8097.

(48) Takahashi, O.; Pettersson, L. G. M. Functional dependence of core-excitation energies. J. Chem. Phys. 2004, 121, 10339–10345.

(49) Michelitsch, G. S.; Reuter, K. Efficient simulation of near-edge X-ray absorption fine structure (NEXAFS) in density-functional theory: Comparison of core-level constraining approaches. *J. Chem. Phys.* **2019**, *150*, No. 074104.

(50) Ljungberg, M. P.; Mortensen, J. J.; Pettersson, L. G. M. An implementation of core level spectroscopies in a real space projector augmented wave density functional theory code. *J. Electron Spectrosc. Relat. Phenom.* **2011**, *184*, 427–439.

(51) Plomp, A. J.; Su, D. S.; de Jong, K. P.; Bitter, J. H. On the nature of oxygen-containing surface groups on carbon nanofibers and their role for platinum deposition – an XPS and titration study. *J. Phys. Chem. C* **2009**, *113*, 9865–9869.

(52) Biniak, S.; Szymánski, G.; Siedlewski, J.; Swiatkowski, A. The characterization of activated carbons with oxygen and nitrogen surface groups. *Carbon* **1997**, *35*, 1799–1810.

(53) López, G. P.; Castner, D. G.; Ratner, B. D. XPS O 1s binding energies for polymers containing hydroxyl, ether, ketone and ester groups. *Surf. Interface Anal.* **1991**, *17*, 267–272.

(54) Desimoni, E.; Casella, G. I.; Morone, A.; Salvi, A. M. XPS determination of oxygen-containing functional groups on carbon-fibre surfaces and the cleaning of these surfaces. *Surf. Interface Anal.* **1990**, 15, 627–634.

(55) Kundu, S.; Wang, Y.; Xia, W.; Muhler, M. Thermal stability and reducibility of oxygen-containing functional groups on multiwalled carbon nanotube surfaces: a quantitative high-resolution XPS and TPD/TPR study. *J. Phys. Chem. C* **2008**, *112*, 16869–16878.