
Anja Aarva,*† Volker L. Deringer,‡ Sami Sainio,∥ Tomi Laurila,⊥ and Miguel A. Caro †⊥§

ABSTRACT: Carbonaceous materials, especially tetrahedral amorphous carbon (ta-C), can form complex functionalized surface structures and are thus promising candidates for applications in biomedical devices and electrochemistry. Functional groups at ta-C surfaces have been widely studied by spectroscopic techniques; however, interpretation of the experimental data is extremely difficult, especially in the case of X-ray photoelectron spectroscopy (XPS) and X-ray absorption spectroscopy (XAS). The assignments of experimental XPS and XAS signals are normally based on references obtained from molecular or crystalline samples, which are simplified approximations for the far more complex amorphous structures. Here, we use extensive density functional theory (DFT) simulations to predict XAS and XPS signatures for carbon-based materials in more realistic environments, building on large data sets of structural models generated by a machine-learning (ML) interatomic potential. The results indicate clear signatures: individual fingerprint XAS spectra and distinctive XPS binding energy distributions, both in terms of center and broadness of the signal, for chemically different groups. The results point out what kind of structural information can and cannot be extracted with X-ray spectroscopy. This study will enable a deeper physicochemical understanding of experimental data and ultimately theory-based identification and quantification of functional groups in carbonaceous materials.

I. INTRODUCTION

Carbonaceous materials, amorphous carbon (a-C) in particular, where carbon can form several different kinds of structures and can be bonded in various ways (sp, sp², and sp³),¹–³ possess many desirable properties. Tetrahedral a-C (ta-C), a form of a-C containing a high degree of sp³-bonded carbon atoms and resembling diamond, is extremely hard, wear resistant, and optically transparent.³ In addition, these materials are biocompatible and resistant to bacterial adhesion. These inherent properties open up countless possibilities for industrial and technological applications. Carbon-based materials are widely used as protective and biomedical coatings,⁴,⁵ but they are also promising materials for real-time electrochemical detection of biomolecules in vivo⁶ that could be used in a variety of medical applications.

Amorphous carbon materials usually also contain elements other than carbon, such as hydrogen and oxygen. These elements can have a marked effect on the electrochemical properties of the materials,⁶–⁸ their electronic properties, such as band gap,⁹ and their mechanical properties which determine their friction performance.¹⁰ Oxygen and hydrogen can be present in the elemental form or as a part of different functional groups that bind to the surface. Often, chemical reactions take place at material surfaces. In particular, this is the case in the context of interfacial electrochemistry, where reactions on the electrode surface are driven by the changing potential of the electrochemical cell as this potential is externally shifted. In practice, most electrochemical reactions of relevance involve at least one step where an electron is transferred between a chemical species adsorbed on the surface and the electrode surface. In these cases, both the thermodynamics and the kinetics of the corresponding redox reactions are heavily influenced by the surface chemistry of the electrode. Thus, understanding the chemistry of the interface plays an important role in comprehending the reactions in question. Improved knowledge about the structure and, especially, the composition of the surface and present functionalities would aid in the development of many of the technologies mentioned above.

Previously, these carbon-based materials have been characterized in detail, for instance, with Raman spectroscopy, transmission electron microscopy (TEM), and time-of-flight elastic recoil detection analysis (ToF-ERDA) and by utilizing X-ray spectroscopy.⁶,⁷,¹¹–¹⁶ Their chemistry has also been widely studied by using electrochemical methods, such as cyclic voltammetry (CV) utilizing various outer and inner...
The main objective of this study is to pave the way toward an improved understanding of experimental X-ray spectroscopy of complex materials by incorporating atomic-scale knowledge from computational simulation. To better understand the roles of differently bonded carbon and to support the results gained with amorphous carbon structures, graphene and diamond samples were added as well-established reference structures. At the same time, graphene and diamond samples, under realistic application conditions, rarely contain only $sp^2$- or $sp^3$-bonded carbon and contain more or less structural defects depending on the origin of the sample. Thus, understanding $a$-C X-ray spectroscopy in more detail can also aid in deconvolution of spectra of nonideal graphene, graphene oxide, and diamond samples. Therefore, amorphous carbon serves as an ideal case study because of its inherent complex structure and its scientific and technological importance.

We provide a systematic methodology, based on a combination of DFT and machine learning (ML), to establish the links between characteristic, representative, atomic motifs and their corresponding XAS and XPS fingerprints. In the present paper (Part I), we generate a database of such fingerprint spectra, to understand certain qualitative spectral characteristics. In Part II, we will show a first attempt at quantitative fitting of experimental spectra of carbonaceous materials using the computational fingerprint spectra as building blocks.

II. X-RAY SPECTROSCOPY: THEORY AND INTERPRETATION

The motivation behind this work is to tackle the challenges in the interpretation of experimental XAS and XPS spectra of solids. Before looking at the issue in detail, we provide a small account of the theory behind XAS and XPS techniques. In Figure 1 we show schematically the working principle of X-ray spectroscopy. The close resemblance between XAS spectra and conduction band DOS is also indicated.

![Figure 1](https://example.com/fig1.png)

**Figure 1.** Schematic image of the working principle of XAS spectroscopy. The close resemblance between XAS spectra and conduction band DOS is also indicated.

spectroscopy with the aid of diamond and graphite as examples. In an initial-state picture, core electrons, namely 1s states in carbon, lie at energetic levels that depend on the atomic environment. For instance, 1s core levels in diamond lie approximately 2 eV deeper than in graphite, when referenced to their respective Fermi levels (since diamond has a band gap, this difference increases to ~5 eV if the reference is taken at the conduction-band minimum). When these core states are excited by X-ray light, the electronic structure of the sample rearranges in the presence of the core hole, leading to the breakdown of the initial-state picture. The energy difference...
between the initial and final state gives the energy of the transition. In other words, the true “initial state” is not the core state, but the ground state of the whole system. The excitation energy is the energy difference between ground state and excited state, where the excited state incorporates all effects of electronic structure reorganization. The actual transition energies measured experimentally, and the difference in these energies between various atomic environments (i.e., the chemical shift for XPS), are therefore more appropriately understood within this final-state picture and also in the context of quasi-particle excitations. With this limitation in mind, we find the initial-state picture useful in illustrating how XAS works.

In an XAS measurement, after a sample is irradiated with X-ray light of energy \( \hbar \omega \), the core electron is promoted to an empty conduction band state, \( \psi_c \), which lies \( \hbar \omega \) above the core level. This electron is detected, and the conduction band is probed. The distribution of conduction band states is characteristic of each material, meaning that XAS allows us to retrieve spectra which look different depending on the sample. The strength or probability of the transition, also known as the cross section, does not only depend on the availability of empty conduction band states but also on their symmetry. This cross section (c.s.) is proportional to the dipole matrix element between the initial (core) and final (conduction) state:

\[
c.s. (1s \rightarrow \psi_c^*) \propto \langle 1s \mid \mathbf{R} \mid \psi_c^* \rangle^2
\]

Therefore, the XAS spectrum resembles, but is not equivalent to, the conduction band density of states (DOS), as is evident from the figure.

The working principle behind XPS is similar to that behind XAS, with some differences. In XPS, after a core electron has been photoionized, its kinetic energy, i.e., the remaining energy after the core electron has been removed from the sample, is measured. Since the energy of the incident light beam is fixed, this kinetic energy can be subtracted from the energy of the incident X-ray photon to give the core electron’s binding energy. In XAS the X-ray beam energy is scanned over a certain range, corresponding to the spectrum of a particular element, and the fraction of the incident photons that are absorbed at each energy value determines the shape of the spectrum. Thus, each excited core electron in XPS yields only one characteristic energy, rather than a spectrum of energies, as is the case in XAS. This makes XAS more sensitive than XPS and better at identifying the chemical nature of local atomic environments. Therefore, interpretation of XPS is even more intricate than interpretation of XAS.

In a molecular orbital theory picture, in diamond bonding is due to \( \sigma \) orbitals exclusively, whereas in graphite/graphene bonding is due to both \( \sigma \) (in-plane) and \( \pi \) (out-of-plane) orbitals. \( \sigma \) bonds are more stable than \( \pi \) bonds and thus lie at lower energy in the valence band. These \( \sigma \) and \( \pi \) states give rise to corresponding antibonding states, \( \sigma^* \) and \( \pi^* \), in the conduction band, which lie at higher and lower energy, respectively. The characteristic first peak in experimental XAS spectra of \( sp^2 \)-containing carbon materials is due to a \( 1s \rightarrow \sigma^* \) transition. These transitions are schematically depicted in Figure 1. The “shoulder” in XAS spectra of high \( sp^2 \)-containing carbon materials is due to a \( 1s \rightarrow \pi^* \) transition.

A critical point to note is that each core state (i.e., each atomic environment) is irradiated with X-ray light over a given range of energies, which gives access to all excitations within that range. Therefore, each core has a continuous XAS spectrum associated with it which contributes to the overall XAS spectrum of the whole sample. Since in ideal diamond and graphite all the sites are equivalent by symmetry, every core level has the same XAS signature (Figure 1). However, things get significantly more complicated when moving to disordered materials, such as a-C, because each atomic site is unique and therefore (1) the chemical shift for that core state (which determines the position of the transition on the energy scale) and (2) the local character of the conduction band (whose symmetry determines the amplitude, or cross section, for the transition) will also be unique. Therefore, the corresponding XAS spectrum for that site will generally differ from the spectra of other sites. As will be discussed in detail in the next section, atomic sites which resemble each other will also present similar XAS spectra. When performing an experimental measurement, all of these individual spectra will be overlaid together to give the overall XAS spectrum of the sample under study. This spectrum in turn will bear the imprint of the sample’s specific composition and nanostructure: for bulk a-C, the most characteristic feature would be the \( sp^2/\pi^* \) ratio.

Because of the existence of characteristic spectral signatures for specific atomic motifs that we have discussed, interpretation of experimental XAS spectra usually relies on fitting to reference data that correspond to sharp features in the spectra. However, a single peak position is not sufficient to capture the whole effect that a fingerprint spectrum of a certain site has on the total spectrum. For instance, the characteristic \( 1s \rightarrow \pi^* \) transition in graphite can also be observed in a-C and attributed to the presence of \( sp^2 \)-bonded carbon. Often, when the experimental spectra are fitted, a molecular reference that corresponds to a certain feature is used. However, this fitting strategy overlooks the fact that, as we showed in Figure 1, each single atomic motif present in the sample contributes characteristically throughout the full continuum of the core-to-conduction band transitions. Therefore, decoupling the contribution of, say, an \( sp^2 \) site at the low-energy end of the spectrum (the \( \pi^* \) transition) from the rest of the spectrum results in a gross simplification of the actual physics taking place during measurement. To work around this issue, in Part II of this study we use the idea of computational “fingerprint spectra” of representative atomic motifs, introduced here, as reference data for fitting, and therefore understanding, experimental XAS spectra of carbon-based materials.

III. METHODS

III.A. Structural Models. Because amorphous materials lack long-range order, the first ingredient for any computational study on the properties of a-C is a suitable atomic model of the material. In this work we use three different kinds of a-C models: (1) a bulk sample created by geometry optimization followed by pressure correction; (2) a reconstructed surface generated by cleaving from the first bulk sample; and (3) a large set of a-C surfaces created by melt-quench simulations with a ML based interatomic potential, to ensure large enough sampling. All these model systems were built using the supercell approach, where “infinite” surfaces are constructed from finite-size slabs by employing periodic boundary conditions. Regardless of the generation method, all structures were ultimately fully relaxed with the density functional theory (DFT) code GPAW, using the PBE functional with respect to geometry and electronic structure prior to computing the XAS spectra and core binding energies, to ensure methodological consistency. Examples of
the supercells employed in this study are depicted in Figure 2. In the following we provide a brief account of each generation method.

The geometry-optimized a-C bulk sample, with a total of 512 atoms, was created by Laurila et al.\(^2\) at the DFT level with a method explained in detail in ref 2, where pressure relaxation allows us to generate a-C samples with mass densities in good agreement with experiment. This structure was used to study the properties of bulk a-C. From this bulk supercell, also an a-C surface was created by “anchoring” the bulk to a diamond substrate and adding a certain amount of vacuum on the top and then allowing for surface reconstruction. However, surfaces generated in this way would be too large for carrying out numerous DFT calculations, as has been done in this work, and thus a reduced surface sample was subsequently created by retaining only the top few angstroms and passivating dangling bonds in the bottom with hydrogen.\(^1\) The surface sample used in this work is 7 Å thick, and it consists of 326 atoms, which is manageably small compared to state-of-the-art DFT codes. In these a-C calculations a \(k\)-point sampling of 2 \(\times 1 \times 1\) was used.\(^4\)

Unlike in the case of graphene and diamond, there are no symmetry-equivalent sites, or equivalent carbon atoms, in the a-C samples. Thus, much bigger sampling is needed to describe the complexity of the material at a desirable level. For this purpose, we use the large set of melt-quench surface reconstructions that were produced and analyzed in the preceding recent studies.\(^5,6\) These samples were created by using the ML-based Gaussian approximation potential (GAP) for a-C developed by Deringer and Csányi.\(^6\) The GAP potential was trained from DFT data and optimized for liquid and amorphous carbon. In preceding work,\(^7\) bulk a-C samples were generated by melt-quenching, the vacuum regions were inserted, and the so-obtained slab structures were further heated, annealed, and finally cooled down to stabilize the surface structure. This melt-quench surface data set consists of a total of 50 slabs, each containing 216 atoms. The method is described in detail in ref 57. For the present work, the structures were further relaxed with DFT using \(2 \times 2 \times 1\) \(k\)-point sampling. These samples were added to the bulk and surface samples, which were generated at the DFT level. Together, they make up the database of pure (nonfunctionalized) a-C structures used in this study.

The reference diamond surface was simulated by introducing a single vacancy defect\(^8\) into a graphene sheet,\(^6\) with a thickness of 10 Å. Two types of termination were used at the bottom of the slab: hydrogen and carbon termination, corresponding to 330- and 360-atom systems, respectively. The termination method had a negligible effect on the characteristic spectra and the adsorption energies of the functional groups, but in some cases carbon termination aided the convergence of XAS calculations. The \(k\)-point sampling in the case of diamond surfaces was \(2 \times 2 \times 1\). For the reference graphene structures two types of surfaces were used: pristine and defective. The size of the graphene sheets was \(9 \times 5 \times 1\) primitive unit cells, and thus there were 180 carbon atoms in a pristine graphene sheet. Defective graphene was simulated by introducing a single vacancy defect\(^9\) into a graphene sheet of the same size as the pristine sample. In both cases, \(k\)-space integration was performed using a Monkhorst–Pack (MP) grid\(^10\) with \(2 \times 2 \times 1\) \(k\)-point sampling.

To dismantle the complexity of the amorphous material and compare to crystalline carbon, the local atomic structures were classified, in terms of the occurrence of characteristic atomic motifs, by employing ML-based clustering techniques.\(^11\) The local atomic environments were characterized with a state-of-the-art many-body descriptor known as “smooth overlap of atomic positions” (SOAP).\(^12-14\) Similar environments were grouped together using an unsupervised ML technique, \(k\)-medoids.\(^12,15\) SOAP-based clustering relies on geometrical features within a certain cutoff radius centered at a given atom. In a recent study,\(^16\) we showed that this clustering technique also allows us to identify motif reactivity toward the functional groups present in this study in a robust way. More details about the clustering approach and the reactivity of the surface sites is presented in ref 63. In Part II of this work,\(^17\) SOAP-based clustering is employed again in order to aid in reconstruction and fitting of the experimental spectra.

III.B. Surface Functionalization. The a-C surfaces introduced in the previous section were studied on their own (as pure carbon) and also systematically functionalized with hydrogen (C–H), oxygen as a ketone, an epoxide or an ether depending on the adsorption site (C–O/C–O–C), hydroxyl group (C–OH), and carboxylic acid (C–COOH) in order to obtain the fingerprint spectrum of each group on a certain site type. In order to functionalize the a-C surface, the most favorable sites needed to be identified. This was done by comparing the local density of states (LDOS) of the surface atoms. The sites that have the highest density of states around the Fermi level are taken to be the most reactive. Explicit adsorption energy calculations for those sites show the expected strong correlation with the LDOS.\(^15\) The functional groups were placed on these sites one at a time, to avoid creating excessively complicated systems that could lead to ambiguous results open to various interpretations.

Adsorption energies are calculated as the difference between the total DFT energy of the surface with the adsorbed group (\(E_{ads}\)) and
the sum of the total energies of the isolated surface (\(E_{\text{surface}}\)) and the group in vacuum (e.g., for -H, \(E_H\)). For instance, the adsorption energy for -H, \(E_{\text{ads}}\), is

\[
E_{\text{ads}} = E_{\text{tot}} - (E_{\text{surface}} + E_H)
\]

where \(E_{\text{tot}} = E_{\text{surface}} + E_H\)

(3)

In all cases, an adequate amount of vacuum was set above the system to ensure that the energy is converged to the correct value. In these adsorption energy calculations, van der Waals corrections were included via the method developed by Tkatchenko and Scheffler.73

To have graphene as a reference, the defective sample was included in this study to ensure that the adsorption energies (\(E_{\text{ads}}\)) of the functional groups considered here were in the range of covalent bonding: pristine graphene is found to be quite inert toward functionalization. This observation is in good agreement with the formation energy calculations of functionalized pristine and defective structures presented in ref 63. It is crucial to point out that computational fingerprint X-ray spectra can be obtained for any functional group in any relaxed structure, but these groups would bear no resemblance to the experimental reality unless the group is really bonded to the surface. On pristine graphene only oxygen has a binding energy that is strong enough to correspond to covalent bonding (Table 1). Introduction of the single vacancy defect creates a favorable site for attachment of functional groups, and the adsorption energies on that site are correspondingly much larger. Adsorption energies of the functional groups on reconstructed diamond, pristine and defective graphene, and geometry-optimized surface-reconstructed a-C are presented in Table 1. More comprehensive data for \(E_{\text{ads}}\) values obtained on melt-quenched a-C surfaces have been published in ref 63.

### III.C. X-ray Spectra Calculations

Using the discussed structural models as starting points, we aim to utilize DFT-based simulations of XAS spectra in order to rationalize the experimental results. We extend this idea to quantitative fitting in Part II of this work.3 For the XPS and XAS calculations we use the method implemented in the GPAW code59,60 by Ljungberg et al.,7 which is well suited for periodic systems containing carbon and oxygen. Different approaches to calculate XPS or XAS spectra exist, as is discussed in refs 41, 45, 74, and 75; and even dynamical core hole screening effects can be included, for instance, by using time-dependent density functional theory (TDDFT) or the computationally less demanding Mahan–DeDominicis (MND) theory.76,77 Susi et al.45 (XPS) have carried out systematic studies about different core-level constraining methods and the effect of level of theory on the error when computational results are compared to experiment. Both studies showed the PBE0 level of theory to be sufficient to reach computed results that are remarkably close to measured values. In addition, Susi et al.44 compared the method implemented in GPAW with results obtained with the all-electron FHI-aims code78 and found them to be in very good agreement. Thus, we choose to rely on the GPAW implementation, that has been shown to yield spectra in good agreement with experiment in our previous work,7 that dealt with sp²- and sp³-rich, as well as amorphous, carbonaceous materials.

XAS simulations consist of two steps. In the first step, we calculate the DFT energy difference between the system’s ground state and an excited state where a core-hole is created in the site of interest and the core electron is promoted to the system’s conduction band. This energy difference is known as the \(\Delta\)Kohn–Sham (\(\Delta\)KS) energy shift. In other words, within this method, in the excited state calculations, the atom that is being excited is represented with a core-hole PAW setup (similar to a pseudopotential) where one electron is removed from the core (1s) state, but the system is kept neutral. These values can be used to interpret XPS spectra, since \(\Delta\)KS values correspond to core-level excitation energies, where one of the core electrons is excited above the Fermi level, and this method has been shown to compare well with experiment.44–46,79,80 In particular, for metals and semiconductors the Fermi level is routinely used as the reference level in XPS measurements (instead of the vacuum level, used for molecules), making comparison between \(\Delta\)KS shifts and experimental XPS more or less straightforward.81 Thus, \(\Delta\)KS values allow us to estimate XPS as well as align XAS spectra on the energy scale (since the \(\Delta\)KS value corresponds to the lowest-energy transition). The \(\Delta\)KS values obtained for defective graphene in this work, both carbon and oxygen K-edge, are in very good agreement with the \(\Delta\)KS values of equivalent defective graphene sites presented by Susi et al.44 The average difference is approximately 0.12 eV. In fact, in most cases the absolute difference is much smaller (<0.03 eV) and the average is affected by few outliers, which most likely stand out because the relaxed geometries of the samples differ slightly.

We also mention a partial limitation of the \(\Delta\)KS approach, related to strong excitonic effects in disordered carbon. In particular, for amorphous carbon the core exciton binding energy for sp³ sites is approximately 0.5 eV larger than in diamond (where it is ca. 0.2 eV).82,83 The shift toward lower energies of the 1s → π⁺ transition energy by about 0.7 eV in a-C compared to graphite can also be explained in terms of the inability of carriers in a-C to efficiently screen the core hole due to the strong disorder potential and consequent high degree of localization. The same effect has been observed in graphene oxide, where the core exciton binding energy is about 0.54 eV.85 The \(\Delta\)KS approach introduces these excitonic effects which are not present in XPS measurements, where the electron is removed from the sample and a bound electron—core hole pair is not formed. However, we note, given the values for exciton binding energies in disordered carbons listed above, and as has been already highlighted in the context of the present methodology,14 that this underestimation is systematic. Therefore, a shift of all \(\Delta\)KS by the same amount, as we do in this work, allows us to compare our computational results to experiment directly.

In the second step, the so-called Haydock recursion method is used to compute the cross section for the transitions between a core level and the different conduction band states. This allows us to obtain the XAS spectra. The Haydock recursion method saves computational effort since it is an iterative method, based on Green’s functions, that is designed to avoid computation of unoccupied states. This method does not allow excitonic features present in XAS spectra to be captured, for instance, the strong excitonic feature corresponding to the 1s → π⁺ transition in diamond.84 On the other hand, the energy shift and the alignment of the second band gap of diamond is captured well. A more detailed description of the method can be found in refs 74 and 85.

In all cases, system sizes and k-point sampling were tested to ensure sufficient convergence with respect to binding energies and XAS spectra. Because of the existence of local (atomic) magnetic moments in a-C, because a defect can introduce magnetization to the system,86 and due to the excited state calculation protocol used, all calculations were carried out with spin polarization. The results are discussed in Section IV.

### Table 1. Adsorption Energies (in eV) of Functional Groups on Reconstructed Diamond, Pristine Graphene, Defective Graphene, and a-C

<table>
<thead>
<tr>
<th></th>
<th>C–H</th>
<th>C–O</th>
<th>C–OH</th>
<th>C–COOH</th>
</tr>
</thead>
<tbody>
<tr>
<td>diamond</td>
<td>−2.5</td>
<td>−4.1</td>
<td>−2.4</td>
<td>−1.9</td>
</tr>
<tr>
<td>prist.</td>
<td>−0.9</td>
<td>−2.2</td>
<td>−0.8</td>
<td>−0.2</td>
</tr>
<tr>
<td>def. graphene</td>
<td>−4.3</td>
<td>−6.3</td>
<td>−3.9</td>
<td>−3.5</td>
</tr>
<tr>
<td>a-C (average)</td>
<td>−4.1 ± 0.6</td>
<td>−5.3 ± 0.9</td>
<td>−4.4 ± 0.6</td>
<td>−4.0 ± 0.6</td>
</tr>
</tbody>
</table>

*Values on a-C are averages of energies calculated on different sites of surfaces reconstructed using geometry optimization.*
IV. RESULTS AND DISCUSSION

IV.A. Qualitative Analysis of Spectra: From Diamond and Graphene to a-C. As we have discussed in Section I and discussed in more detail in Part II, well-defined crystalline materials have easily interpretable XAS spectra associated with them, since it is straightforward to attribute the presence of specific features in the spectra to the atomic motif from which they originate. Therefore, we will use functionalized diamond and graphene curves to obtain relatively simple XAS spectra associated with them, to put the significantly more complex XAS spectroscopy of a-C into context.

In order to construct a simulated carbon K-edge X-ray spectrum which is directly comparable to experiment, we need to combine pure carbon sampling with the fingerprint spectra of carbon bound to functional groups. These fingerprint spectra are calculated by placing the core hole on the carbon atom to which the functional group in question is bound (i.e., the C atom that shares a bond with it) yielding a carbon K-edge spectrum for the site. Alternatively, the core hole can be placed on an oxygen atom, which can be present on the surface as part of these functional groups, thus providing us with the oxygen K-edge spectrum for that particular site. Depending on the group in question and how the carbon (or the oxygen) is bonded to the carbon matrix that forms the surface, an individual spectrum with characteristic features arises.

A comparison of calculated C K-edge fingerprint XAS spectra for pure carbon samples (published previously in ref 7), together with functionalized graphene, a-C, and diamond, is presented in Figure 3. For a-C we sample the different atomic motifs discussed in Section III.A. The trends are clear: moving from the sp\(^2\)-rich graphite/graphene sample to the sp\(^3\)-rich diamond bulk/surface, the characteristic \(\pi^*\) peak (ca. 285 eV on the figure) disappears and the \(\sigma^*\) characteristics (ca. 291 eV on the figure) are enhanced. Amorphous carbon, that has carbon in all forms (sp, sp\(^2\), and sp\(^3\)), lies somewhere in between the two. Note that computational energy values do not exactly match experimental values because a small systematic underestimation of the \(\Delta KS\) energies (the onset of the spectra) is present in DFT calculations.45 When the shift is calculated, the error is different for different elements, such as carbon and oxygen44 or carbon and nitrogen.76 For instance, in the case of pure sp\(^3\) carbon an approximately 1 eV shift toward higher energies would be needed to match the experimental spectrum.77 Nevertheless, Figure 3 clearly shows how the spectra of differently bonded carbons differ from each other, and how the features change when a functional group is introduced to the site. At the same time, this figure highlights the relevance of using the features present throughout each whole individual spectrum when experimental data are interpreted, rather than only certain peak positions.

The first step toward unfolding the complexity of a-C X-ray spectra and making the connection between spectroscopy and the material’s nanostructure is to unfold the structural complexity of a-C itself. That is, how do the spectroscopic features of a-C in the scale from ideal graphene to ideal diamond correlate with the transition between sp\(^2\)-rich and sp\(^3\)-rich chemical bonding? A systematic way to deal with this problem is to use machine-learning-based clustering techniques to classify and identify atomic motifs by grouping together all the atomic sites which resemble each other. We used this approach in a recent study to classify carbon atoms into six distinct groups: one sp\(^3\) motif (cluster 6), two different sp\(^2\) motifs (clusters 4 and 5), two different sp motifs (clusters 2 and 3), and a highly unstable undercoordinated motif where the carbon atom has only one neighbor (cluster 1).63 The rationale behind clustering is that of achieving a compact representation of typical structural motifs and computing corresponding properties that can be assigned to said motifs. This is presently the case regarding characteristic X-ray spectra. In the remainder of this section we discuss how this classification scheme can help in the qualitative understanding of X-ray spectroscopy of a-C.

IV.B. Qualitative Understanding of a-C XAS. XAS spectra of individual atoms are full of sharp features. With a-C the situation is more complicated since there are no equivalent sites. With large-enough structural sampling, individual peaks are merged together, becoming wider and more representative. For this purpose, we employ the melt-quench surface-reconstructed samples discussed in Section III.A that consist of 50 slabs, each containing 216 atoms. For a more detailed analysis than just comparing between sp, sp\(^2\), and sp\(^3\), all sites are clustered using the method introduced in ref 63. Cluster 1 (not presented here) consists of extremely reactive undercoordinated sites located at the end of a short carbon chain. Clusters 2 and 3 consist of different kinds of sp sites along a carbon ring. Clusters 4 and 5 consist of sp\(^2\) sites that differ slightly with respect to bond lengths and angles. Cluster 6 consists of sp\(^3\) sites. Based on their structural similarity, one would expect that sites that belong to a certain cluster give similar spectra and that the fingerprint spectra of different motifs differ from each other. This is expected because the clusters have different geometries and local electronic proper-
ties due to their specific atomic environment, as well as exhibit different kinds of affinity toward functional groups. Motif 1 appears in our large data set only three times, and it is thus not possible to obtain adequate statistics. It is also expected that, even though this motif appears in simulated surfaces, these sites are so reactive that they would be terminated immediately during the manufacturing process. For these reasons motif 1 is left out of the discussion in this study.

Motif-specific spectra of pure carbon and functionalized carbon sites in a-C surfaces are depicted in Figure 4, with the pure carbon spectra given on the top row. As can be seen, clusters 2 and 3, which contain $sp$ sites in a chain-like ring structure, show strong $\pi^*$ characteristics around $\sim 285$ eV. Cluster 3 has also a clear peak at $\sim 282$ eV, which can be associated with high reactivity. It has been shown that motif 3 indeed presents the strongest reactivity toward functional groups and is, as a matter of fact, significantly more reactive than the other, at first glance similar, $sp$ motif. The main structural difference between these two $sp$ motifs, which drives the difference in reactivity, is a flatter bond angle for the less reactive one (motif 2). Also clusters 4 and 5, which contain $sp^2$ sites, show $\pi^*$ features. On the other hand, the spectra from cluster 6, which contains only $sp^3$ sites, have only $\sigma^*$ features.

These surface spectra, together with spectra from a bulk sample, will form the basis of our pure carbon reference database when we fit the experimental spectra in Part II of this work.

We now move on to functionalized a-C surface sites. For clarity, we use the convention, throughout the paper, that the atom where the core hole is set is marked in bold face. For instance, when discussing a carbon double-bonded to an oxygen, we write $C\equiv O$ to denote a carbon core hole and $C\equiv O$ to denote an oxygen core hole. The fingerprint spectra of carbon sites that are bonded to functional groups are depicted

---

**Figure 4.** Predicted carbon K-edge spectra of pure carbon surface sampling compared with functionalized sites. Columns labeled 2 to 6 correspond to different types of motifs, and the first row gives the spectra for the bare motifs, that is, in the absence of adsorbed functional groups. The subsequent rows show how the spectra of these sites change as the functional groups are introduced. No functional groups are added to type-6 atoms, which are already 4-fold coordinated ($sp^3$). Individual spectra are plotted with thin light-gray lines, averages are given with thick colored lines, and uncertainty intervals (the standard deviation around the average) are plotted as dark-gray shaded areas. Schematic images of the corresponding motifs and the functional groups are depicted next to the spectra. The atom from which the core electron is excited and which, therefore, holds the core hole, is marked with a circle. Dashed lines at 285 and 287 eV are added to guide the eye.
in Figure 6 below the pure carbon surface spectra. The oxygen K-edge spectra are depicted in Figure 5. Here we focus on groups that are bonded to sites belonging to clusters 2−5. Cluster 6, which consists of sp\(^3\) sites, was found to be unfavorable for functionalization. Every time a functional group was brought to the site and the structure was relaxed, bond breaking occurred in the surrounding carbon matrix; thus, the site no longer represented the original motif. Hence, no spectra could be obtained for functional groups on sites that belong to cluster 6. Again, in Figure 6, it is evident that each cluster gives individually characteristic spectra. When oxygen was brought to an sp\(^3\) site (clusters 2 and 3), and the structure was subsequently relaxed, bond angles at the site approached 120°. XAS spectra of the carbons bound to this oxygen show a clear π* peak, which is however shifted by ∼+2 eV compared to pristine carbon without an adsorbant. When the oxygen was brought to an sp\(^2\) site (clusters 4 and 5), and the system was relaxed, bond angles approached 109°. In this case, the XAS spectra of the carbons that are bonded to the oxygen lose their π* features and then show very distinctive σ* characteristics.

In Figure 4 we only show the spectra of carbon bonded to a ketone type of oxygen (C═O). However, oxygen can be bonded in different ways, all of which give different kinds of characteristic spectra. Figure 6 presents the spectra of ether and epoxide groups [(a and c) C−O−C and (b and d) C−O−C]. In these groups, the oxygen atom is simultaneously bonded to two different carbons and thus the clustering scheme based on a single C adsorption site is no longer valid. It is interesting to see how epoxides, which have a very constrained geometry, have also quite well-defined characteristic spectra, whereas in ethers, where there is more variation with respect to the bond angle and the surrounding ring structure, the spectra are more scattered. In the case of ethers and epoxides, however, based on the shape of the spectra it would seem impossible, in an experimental setting, to separate between the two.

In the case of C−H and C−OH the characteristic π* peaks are located at approximately ∼285 eV and ∼287 eV, respectively, and thus they overlap with pure carbon and C═O π* peaks, respectively. However, if we look at the shape of the whole spectrum, we notice that these spectra are unique for every combination of adsorbant plus adsorption site.

In the case of C−COOH, two different carbons (and two different oxygens) can be probed. When attached to motifs 2 and 3 (sp sites), C−COOH shows a double peak, ∼285 eV and ∼288 eV, located at around the onset of the spectra. When attached to motifs 4 and 5 (sp\(^2\) sites), C−COOH shows quite sharp σ* characteristics. The C−COOH XAS spectrum is more, but not completely, independent from the adsorption site and, interestingly, all C−COOH spectra have very intense π*
Lorentzian pro-Voigtian lineshapes, i.e., a combination of Gaussian and instrumental resolution whereas Lorentzian broadening corresponds to the lifetime of the excitation. In this work we are choosing to use a 0.3 eV width for both.

We must note that ΔKS values obtained from DFT calculations, i.e., the calculated differences between (core-hole) excited-state and ground-state total energies, lack the accuracy of higher-level approximations, such as the GW method. However, even though absolute ΔKS values may not mirror experimental XPS exactly, they can provide a meaningful description of relative energies. Also, the core level energies can vary depending on the exchange-correlation density functional used in the DFT calculations. In the case of XAS, the choice of functional affects the overall shift of the spectrum, but the relative positions of certain features of the spectra are significantly less affected. Unfortunately, the GW method is still computationally very expensive and thus out of reach when large data sets of periodic systems are studied, as done in this work. In this study we choose to use the PBE level of theory, which has been shown to be a good option with carbon-based materials. In fact, in some cases, increasing the level of theory can actually increase the error when results are compared to experiment, since even though higher level functionals may

**IV.C. Qualitative Understanding of a-C XPS.** We now turn our attention from XAS spectroscopy to XPS spectroscopy. ΔKS values have been shown to be good guidelines for the assessment of experimental XPS spectra. In Figures 8–10 we present carbon K-edge ΔKS values of pure carbon sites at the surface (Figure 8) and functionalized carbon sites (Figure 9) and O K-edge ΔKS values of the functional groups (Figure 10). The predicted spectra are obtained by applying Voigtian lineshapes, i.e., a combination of Gaussian and Lorentzian profiles. Gaussian broadening describes the instrumental resolution whereas Lorentzian broadening corresponds to the lifetime of the excitation. In this work we are choosing to use a 0.3 eV width for both.

We must note that ΔKS values obtained from DFT calculations, i.e., the calculated differences between (core-hole) excited-state and ground-state total energies, lack the accuracy of higher-level approximations, such as the GW method. However, even though absolute ΔKS values may not mirror experimental XPS exactly, they can provide a meaningful description of relative energies. Also, the core level energies can vary depending on the exchange-correlation density functional used in the DFT calculations. In the case of XAS, the choice of functional affects the overall shift of the spectrum, but the relative positions of certain features of the spectra are significantly less affected. Unfortunately, the GW method is still computationally very expensive and thus out of reach when large data sets of periodic systems are studied, as done in this work. In this study we choose to use the PBE level of theory, which has been shown to be a good option with carbon-based materials. In fact, in some cases, increasing the level of theory can actually increase the error when results are compared to experiment, since even though higher level functionals may

**Figure 7.** Graphene test that shows that addition of a functional group on a certain site [site (a), marked with a red circle] changes the XAS fingerprint spectra only locally. The effect of functionalization declines quickly: sites (a)–(f). Site (f) is far away from the functionalized site and corresponds to pristine graphene. The image also shows how an ideal graphene site differs from the sites around the defect.

**Figure 8.** C K-edge ΔKS values of pure carbon sites at the surface. The spectra are weighted according to how often the sites appear in the data set. Filled curves show the shape of the spectrum when contributions from all motifs are summed together, and the normalized individual spectra are plotted below. Schematic images of motifs 2–6 are depicted in Figure 4.
result in more accurate ground state energies, they may not be suitable to be used with core hole setups. Since in a-C every site is unique with respect to its local atomic environment and has an individual ΔKS value associated, just like it has an individual XAS spectrum, we can provide ΔKS distributions for similar and dissimilar sites. In any case, only with a high-enough sampling can we estimate how widely spread these ΔKS values are, i.e., how wide the corresponding XPS peaks will be. From Figures 9 and 10 it is clear that each group has its own fingerprint ΔKS range.

In the case of carbon XPS spectra, C−H, C−OH, and C−COOH values can still be separated according to the original adsorption site. However, especially C−H and C−COOH ΔKS values overlap substantially with the values of pure carbon, i.e., within the energy range of ~282−285 eV. Pure carbon ΔKS values below 282 eV correspond to highly reactive sites that would most likely be terminated with a functional group in real experimental conditions. Telling these motifs apart from plain carbon sites seems to be an impossible task for XPS. The C−OH XPS signal lies a bit higher on the energy scale. C=O and C−C−OOH ΔKS values are more narrowly distributed and therefore should, in principle, stand out more clearly from the plain carbon signature. Interestingly, in the case of oxygen K-edge spectra, the situation is the opposite, and C−OH, C−COOH, and C−COOH ΔKS values are narrowly distributed whereas C=O values are more widely spread. In addition, ether and epoxide ΔKS values, the C K-edge peak position ~285.1 eV and the O K-edge peak position ~530.4 eV, differ substantially from ketone values (C K-edge ~284.4−284.7 eV and O K-edge ~526.7−528.3 eV). XPS signals coming from ethers and epoxides would overlap with the ones coming from the hydroxyl group. Once again, oxygen as a functional group shows more complex behavior compared to other groups included in this study. All the XPS peak positions are summarized in Table 2 (for carbon) and Table 3 (for oxygen).

V. CONCLUSIONS

In this study we have introduced the concept of fingerprint spectra and discussed its usefulness for qualitative interpretation of experimental XAS and XPS data. We have illustrated the difficulty in understanding experimental results with the important case of disordered carbonaceous materials, in particular a-C. We have performed a detailed comparison between X-ray spectroscopic characteristics of graphene, a-C, and diamond, and we carried out a qualitative analysis of their XAS and XPS spectra, including the intricate contribution arising from hydrogen and oxygen-containing functional
groups commonly found in these materials. Furthermore, we provide an extensive database of XAS fingerprints as well as ΔKS (XPS) values to be used in interpretation of experimental data. This database consists of ab initio results that can be directly compared with experiment. The data sets will be made publicly available in the near future. In the meantime, the authors will provide the data on request.

We have shown how certain features of the individual fingerprint spectra overlap and, thus, have illustrated the difficulty in identifying these characteristics from experimental spectra. On the other hand, some motifs have very clear characteristics, and, indeed, these features can be used to detect the functional group or a plain carbon motif in question, directly from the experimental data. A new methodology to provide an extensive database of XAS fingerprints as well as ΔKS (XPS) values to be used in interpretation of experimental data. This database consists of ab initio results that can be directly compared with experiment. The data sets will be made publicly available in the near future. In the meantime, the authors will provide the data on request.

We have shown how certain features of the individual fingerprint spectra overlap and, thus, have illustrated the difficulty in identifying these characteristics from experimental spectra. On the other hand, some motifs have very clear characteristics, and, indeed, these features can be used to detect the functional group or a plain carbon motif in question, directly from the experimental data. A new methodology to perform quantitative fitting of experimental X-ray spectra of carbonaceous materials, based on the concepts introduced and discussed in this paper, will be presented in Part II of this study.53

**REFERENCES**


(62) The k mesh used has twice as many k points along the k axis than along k, because in real space the cell dimensions along y are twice as large as those along x.


