Understanding the chemistry of customized carbonaceous nanomaterials

Anja Kristiina Aarva
Understanding the chemistry of customized carbonaceous nanomaterials

Anja Kristiina Aarva

A doctoral thesis completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall TU1 of the school on 15 December 2022 at 12 o'clock.

Aalto University
School of Electrical Engineering
The Department of Electrical Engineering and Automation
Microsystems Technology
Supervising professors
Prof. Tomi Laurila, Microsystems Technology

Department of Electrical Engineering and Automation, Aalto University, Finland

Department of Chemistry and Materials Science, Aalto University, Finland

Thesis advisors
Dr. Miguel A. Caro, Research Fellow

Department of Electrical Engineering and Automation, Aalto University, Finland

Preliminary examiners
Dr. Deyu Lu, Center for Functional Nanomaterials, Brookhaven National Laboratory, USA

Prof. Robert Vajtai, Research Professor, Rice University, USA

Opponent
Dr. Annika Bande, Group Leader, Helmholtz-Zentrum Berlin für Materialien und Energie, Germany
Abstract

Carbon-based nanomaterials have exceptional properties, that can lead to new advanced applications in many fields of technology and science, for instance: energy solutions, protective and biomedical coatings, and biosensing.

This thesis is focused on computational methods used to understand better the structure-property relationships of these materials. We look deeply into surface chemistry of carbon nanomaterials and develop a new method for characterization of carbonaceous materials by combining computational and experimental data.

In addition to elemental carbon, carbonaceous nanomaterials often contain other elements, like hydrogen and oxygen. As the properties of the materials depend on their chemical composition we carried out a detailed analysis about reactivity of amorphous carbon surfaces to obtain more information on this connection. In this work, a combination of novel computational methods were employed: density functional theory-level calculations, with machine learning-based structure construction and classification of the chemical environments present in the structures. This approach enables us to achieve statistically meaningful data-set, with high accuracy and low computational cost.

Once we have studied the reactivity of carbon, careful material characterization is in place, so that it is possible to compare computational results with experimental samples. This comparison is valuable in two ways: we want to understand the experimental data better but, on the other hand, there is a need to make sure that the computational samples reflect experimental results in a realistic way. In other words, our models enable computational experiments.

X-ray spectroscopy provides an extremely accurate means for analyzing the surface chemistry of substances in general. However, interpretation of the experimental data, especially when the sample is disordered, can prove to be challenging. The computational method that we introduce in this thesis can recreate the X-ray spectra from first principles, which provides a long searched connection between macroscopic and atomic level structures.

As a conclusion, the information gained by combining novel computational methods with experimental science can be used to develop customized materials, e.g., for electrochemical detection of dopamine, which is used as a model system in this study, but also for several other promising applications.

Keywords Nanomaterials, Carbon, Electrochemistry, Density functional theory, Machine learning

ISSN (printed) 1799-4934 ISSN (pdf) 1799-4942
Location of publisher Helsinki Location of printing Helsinki Year 2022
Tekijä
Anja Kristiina Aarva

Väitöskirjan nimi
Hiilipohjaisten nanomateriaalien pintakemian syvällinen analyysi

Julkaisija
Sähköteknian korkeakoulu

Yksikkö
Sähköteknian ja automaation laitos

Sarja
Aalto University publication series DOCTORAL THESESES 174/2022

Tutkimusala
Elektroniikan integrointi ja luotettavuus

Käsikirjoituksen pvm
12.10.2022
Väitöspäivä
15.12.2022

Väittelyluvan myöntämispäivä
28.10.2022
Kieli
Englanti

Monografia
Artikkeliväittöskirja
Esseeväittöskirja

Tiivistelmä
Hiileen pohjautuvilla nanomateriaaleilla on lukuisia ominaisuuksia, jotka tekevät niistä kiinnostavia tutkimuskohteita useilla tieteen aloilla. Mahdollisia sovelluskohteita löytyy esimerkiksi energiaratkaisuiden, pinnoitteiden ja sähkökemiallisten antureiden parista.

Tässä työssä tutkitaan näiden materiaalien rakenteen ja ominaisuuksien vähistä suhdetta laskennallisia menetelmiä hyödyntäen. Tarkastelun kohteena on erityisesti hiilen pintakemia. Työssä on kehitetty uusi menetelmä hiilipohjaisten materiaalien karakterisointia varten. Menetelmä perustuu kokeellisten tulosten systemaattiseen analyysiin laskennallisten tulosten avulla.


Röntgenspektroskopiaan perustuvilla mittauksilla on mahdollista analysoida materiaalien pintakemian erittäin tarkasti. Mutta, erityisesti, jos näytteen rakenne on järjestäytymätön, kuten amorfisen hiilen tapauksessa, tulosten tulkinta voi olla hyvin haastavaa. Tässä työssä esiteltä laskennallinen menetelmä tarjoaa uuden työkalun tulosten tulkintaan ja mahdollistaa tulosten tulkinnan atomitasolla.

Lopputapaukseilla voidaan todeta, että laskennallisten menetelmien yhdistäminen kokeellisten tulosten tulkitsemiseksi tukee uusien räätläöityjen nanomateriaalien kehitystyötä. Näille materiaaleille on kysyntää esimerkiksi sähkökemiallisten antureiden suunnittelussa, mutta myös lukuisilla muilla korkean teknologian aloilla.

Avainsanat
Nanomateriaalit, Hiili, Sähkökemia, Tiheysfunktionaliteoria, Koneopposiminen

ISSN (painettu) 1799-4934 ISSN (pdf) 1799-4942

Julkaisupaikka Helsinki Painopaikka Helsinki Vuosi 2022

Understanding the chemistry of customized carbonaceous nanomaterials

Anja Aarva\textsuperscript{1}

\textsuperscript{1}Department of Electrical Engineering and Automation,
School of Electrical Engineering, Aalto University, 02150 Espoo, Finland
(Dated: November 3, 2022)
ACKNOWLEDGEMENTS

In all cases, computational resources were provided by CSC – IT Center For Science LTD., which is gratefully acknowledged. I also want to thank Aalto Science-IT for all the support and teaching.

Prof. Tomi Laurila, thank you for giving me this opportunity. I don’t know what I would have done without it. This was my dream already long time ago; and now it is finally happening.

Dr. Miguel A. Caro, thank you for all the help and input – and friendship – you have given me during all these years. I know it’s a lot from you. It has been priceless.

Prof. Volker Deringer, I want to thank you for all the encouraging collaboration, feedback and presence, whenever I needed you. And, not to mention, thank you for the wonderful visits I got to experience; I’ll never forget.

Dr. Sami Sainio, year by year I understand you better. I admire your strong will and determination. Please, don’t ever change!

Prof. Emilia Peltola, you have provided a perfect example of how to make it in this crazy Peter Pan world. I’m not sure if I’m up to that, but it feels good to see that you can and you will. Thank you!

All the rest of the Microsystems Technology group – including the ones that have already moved on with their lives – thank you for being there providing company and support.

Grandparents of our children, thank you for taking care of the kid’s when I couldn’t. Also; thank you for parental advisory. I never thought I would say that.

Juuso, my nearest and the dearest, sorry for my absence during the past few years, but try to hang along. Thank you for all the support and kindness. I love you.
Ada and Noa, you will always be my true masterpieces.
CONTENTS

Acknowledgements 2
Abbreviations and Symbols 6
Publications included in this thesis 8
Author’s contribution 9

I. Introduction 11

II. Experimental framework: an overview and motivation behind the computational work 13
A. Electrochemical measurements 13
B. Experimental X-ray spectroscopy characterization 16

III. Basics of density functional theory 17
A. Theoretical background 17
B. Technical details about obtaining the structure of the systems 20
C. X-ray spectroscopy calculations 20

IV. Machine learning for material science of carbon-based materials 23
A. GAP: an interatomic potential for obtaining the structure 23
B. SOAP: a tool for classification of the single sites within the structures 26
C. MDS: two dimensional visualization of multidimensional system 28

V. Results and discussion 30
A. Effects of changing the potential on adsorption energy 30
B. Understanding the structure of the disordered carbon materials 33
C. Characterization by computational X-ray spectroscopy: General discussion 34
D. Part I: X-ray spectroscopy fingerprints of clustered a-C motifs 38
E. Part II: Fitting experimental X-ray spectra with re-clustered fingerprint spectra 40
F. Expanding the method to other disorganized carbon based materials: graphene, graphite, and their oxidized forms 44
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-C</td>
<td>Amorphous carbon</td>
</tr>
<tr>
<td>AA</td>
<td>Ascorbic acid</td>
</tr>
<tr>
<td>ASE</td>
<td>The atomic simulation environment</td>
</tr>
<tr>
<td>C 1s/C K-edge</td>
<td>Atomic- or X-ray notation, respectively, that refers to the energy that is needed to excite the electrons during the measurement when carbon in the sample is being studied</td>
</tr>
<tr>
<td>CA</td>
<td>Cathecol</td>
</tr>
<tr>
<td>CSM</td>
<td>Continuum solvent model</td>
</tr>
<tr>
<td>CV</td>
<td>Cyclic voltammetry</td>
</tr>
<tr>
<td>D</td>
<td>Diamond</td>
</tr>
<tr>
<td>DA</td>
<td>Dopamine</td>
</tr>
<tr>
<td>DAox</td>
<td>Dopaminequinone</td>
</tr>
<tr>
<td>DFT</td>
<td>Density functional theory</td>
</tr>
<tr>
<td>DLC</td>
<td>Diamond like carbon</td>
</tr>
<tr>
<td>DOS</td>
<td>Density of states</td>
</tr>
<tr>
<td>DV</td>
<td>Double-vacancy</td>
</tr>
<tr>
<td>$E_{\text{ad}}$</td>
<td>Adsorption energy</td>
</tr>
<tr>
<td>G</td>
<td>Graphene</td>
</tr>
<tr>
<td>GAP</td>
<td>Gaussian approximation potential</td>
</tr>
<tr>
<td>GO</td>
<td>Graphene oxide</td>
</tr>
<tr>
<td>GPAW</td>
<td>DFT code within the projector-augmented wave method framework</td>
</tr>
<tr>
<td>GPR</td>
<td>gaussian process regression</td>
</tr>
<tr>
<td>LDOS</td>
<td>Local density of states</td>
</tr>
<tr>
<td>MDS</td>
<td>Multidimensional scaling</td>
</tr>
<tr>
<td>ML</td>
<td>Machine learning</td>
</tr>
<tr>
<td>O 1s/O K-edge</td>
<td>Atomic- or X-ray notation, respectively, that refers to the energy that is needed to excite the electrons during the measurement when oxygen in the sample is being studied</td>
</tr>
<tr>
<td>SLAC</td>
<td>Stanford synchrotron radiation lightsource</td>
</tr>
</tbody>
</table>
SOAP Smooth overlap of atomic positions
SSRL synchrotron storage ring and beamline (SSRL) stations
SV Single-vacancy
ta-C Tetrahedral amorphous carbon
vdW Van der Waals forces
XAS X-ray absorption spectroscopy
XPS X-ray photoelectron spectroscopy
ΔKS Delta Kohn-Sham, which describes the onset of XAS spectra and can be related to XPS spectra

<table>
<thead>
<tr>
<th>Functional groups</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-COOH</td>
<td>Carboxylic acid bonded to a carbon atom</td>
</tr>
<tr>
<td>-H</td>
<td>Hydrogen bonded to a carbon atom</td>
</tr>
<tr>
<td>-O</td>
<td>Oxygen bonded to a carbon atom</td>
</tr>
<tr>
<td>-OH</td>
<td>Hydroxyl group bonded to a carbon atom</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbols</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>Elementary charge</td>
</tr>
<tr>
<td>$E_{II}$</td>
<td>Repulsive nucleus-nucleus interaction</td>
</tr>
<tr>
<td>h</td>
<td>Planck’s constant</td>
</tr>
<tr>
<td>$\hat{H}$</td>
<td>Hamiltonian operator</td>
</tr>
<tr>
<td>$m_e$</td>
<td>Electron mass</td>
</tr>
<tr>
<td>$M_I$</td>
<td>Nuclei mass</td>
</tr>
<tr>
<td>$r_i$</td>
<td>Coordinate of an electron</td>
</tr>
<tr>
<td>$R_I$</td>
<td>Coordinate of a nucleus</td>
</tr>
<tr>
<td>$\hat{T}_{\text{elec}}$</td>
<td>Kinetic operator for the electrons</td>
</tr>
<tr>
<td>$\hat{T}_{\text{nuc}}$</td>
<td>Kinetic operator for the nuclei</td>
</tr>
<tr>
<td>$\hat{V}_{\text{ext}}$</td>
<td>Attraction between the nuclei and the electrons</td>
</tr>
<tr>
<td>$\hat{V}_{\text{int}}$</td>
<td>Repulsion between the electrons</td>
</tr>
<tr>
<td>$\psi$</td>
<td>General function</td>
</tr>
<tr>
<td>$\omega$</td>
<td>General eigenvalue</td>
</tr>
<tr>
<td>$\hat{\Omega}$</td>
<td>General operator</td>
</tr>
<tr>
<td>$\nabla$</td>
<td>Nabla</td>
</tr>
</tbody>
</table>
PUBLICATIONS INCLUDED IN THIS THESIS

The list of publications that are included and summarised in this thesis:


AUTHOR’S CONTRIBUTION

Publication [1], a book chapter, represents the big picture of the work that has been carried out in our group. This publication contains results from many members of our group: Dr. M. A. Caro, Dr. S. Sainio and Dr. T. Palomäki, who have all provided inspiration and support during the defendant’s studies. The authors contribution to the results involve a-C surface reactivity, LDOS calculations and computational X-ray spectroscopy. All the work has been carried out under supervision of Prof. T. Laurila.

In publication [2] the author has contributed all of the computational results and visualization, except the electrostatic potential calculations, which are done by Dr. M. A. Caro, who also provided assistance whenever it was needed. Prof. T. Laurila controlled the work from electrochemical point of view.

Dr. M. A. Caro is the first author in publication [3] and the defendant the second. The first author introduced SOAP clustering, analysis of the clustered sites and H-probe analysis. The second author computed adsorption energies of the functional groups, LDOS of the sites in question, and bond lengths as well as the bond angles between the functional groups and the probed carbon sites. Prof. V. L. Deringer together with Prof. G. Csányi provided the vast data-set of a-C created via ML-based potential. Prof. T. Laurila was the supervisor of the work.

The defendant is the first author in this publication [4] as well as in the following. Here, in Part I the author, made the comparison between graphene, a-C and diamond, computed all the fingerprint spectra, both XAS and XPS, for C 1s and O 1s, as well as tabulated (ΔKS) values. Prof. V. L. Deringer provided the ML-based a-C data-set. Dr. S. Sainio and Prof. T. Laurila brought experimental point of view to the work. Dr. M. A. Caro provided a schematic image about the working principles of XAS, as well as assistance and supervision throughout the work.

Publication [5] is Part II of the preceding work, and the distribution of work is very much the same. The first author performed reclassification of the structures and all the fitting
calculations of the experimental samples, and visualization of the results. The structures were originally introduced by Prof. Deringer, who also gave brilliant support in visualization and provided a schematic image of the fitting scheme workflow. Experimental spectra were provided by Dr. S. Sainio and Prof. T. Laurila aided in interpretation of the experimental spectra. The fitting scheme as well as the code for fitting is created by Dr. M. A. Caro. The work was carried out under supervision of Dr. M. A. Caro.

In publication [6], as the first author, the defendant calculated all the computational spectra, XAS and XPS, analyzed the results, reclassified the spectra, and selected appropriate spectra for fitting and carried out the fitting. Dr. S. Sainio kindly provided the experimental spectra. Prof. Deringer and Prof. T. Laurila gave valuable input about the chemistry of the graphene and its many forms. Dr. M. A. Caro provided a new improved code for X-ray calculations and the fitting was done with his code. The work was supervised by Prof. T. Laurila.
I. INTRODUCTION

Carbonaceous nanomaterials have the potential to revolutionize several fields of technology, including energy solutions, protective and biomedical coatings, and biosensing. The feasibility of these materials has been demonstrated, for example, to be used as electrode materials for electrochemical detection of biomolecules, e.g., dopamine (DA), \textit{in vivo} [1, 2, 7–11], which has been one of the main interests of our research group within recent years. In this thesis, we utilize computational methods to understand better what happens during electrochemical measurements [2, 8]; study chemistry of carbon in general [3, 12]; and develop a new method for interpretation of experimental characterization of carbon-based materials via X-ray photoelectron spectroscopy (XPS) and X-ray absorption spectroscopy (XAS) [1, 4, 5, 11, 13, 14].

The motivation behind studying the chemistry of DA, in particular, is that changes in the DA level have been associated with several neurological disorders, including Parkinson’s disease and depression, and the real-time determination of these levels would greatly help in developing better diagnostic and treatment methods [7, 15, 16]. The electrochemical behavior of biomolecules often exhibits inner-sphere redox characteristics, which means that their oxidation/reduction behavior strongly depends on the chemistry between the reactant and the electrode surface [17, 18]. Changes in the potential across the electrochemical cell originate the forces which initiate and drive the reactions [2, 9], but not only; several complicated reaction paths can follow. Adsorption in particular plays an important role in the oxidation reaction path in the case of DA, but it is not necessarily controlled by the potential changes [9, 18, 19]. In the beginning of this work we have aimed to study DA adsorption and electron transfer computationally during potential changes in aqueous solution on different carbonaceous surfaces [2, 8]. All adsorption energy calculations have been carried out at the density functional theory (DFT) level.

Carbon-based materials often also contain elements other than carbon, especially hydrogen and oxygen, and the physicochemical properties of the materials depend on their chemical composition [7, 8, 20–23]. For this reason, we continued the work with a detailed analysis of reactivity of amorphous carbon (a-C) surfaces [3]. Amorphous carbon is an optimal candidate for studying the surface chemistry of carbonaceous materials since it has a variety of differently bonded carbon, and the reactivity of the surface sites depends on the
bonding coordination of the site in question. The bonding coordination of a particular site in a-C reflects what can be anticipated to be the behavior on graphite and diamond surfaces meaning, whether the site is $sp^2$ or $sp^3$ bonded, respectively. In this work we have combined DFT-level computation with machine learning (ML)-based structure construction [12] and classification of the chemical environments present in the structures [3, 24].

In order to obtain an even more in-depth understanding of the connection between the properties and the structure of these materials, careful material characterization is needed. X-ray absorption spectroscopy (XAS) and X-ray photoelectron spectroscopy (XPS) are extremely powerful methods for analyzing the electronic structure and, thus, the composition of the material in question [16, 25, 26]. However, since the materials discussed here are so complex, interpretation of the experimental data is challenging and can lead to severely biased results. A simulation tool that can recreate spectra from first principles can aid in understanding the experimental data remarkably. Computational references can be especially useful when the spectra of more complicated materials such as a-C and graphene oxide are interpreted. In order to do this, we have again combined DFT calculations with unsupervised ML-based data clustering [4, 5, 11, 13]. In other words, we employ DFT-based simulations to calculate the XAS spectra of different pure and functionalized carbon sites. We also calculate the $\Delta$ Kohn-Sham ($\Delta$KS) energy shifts which define the onset of the XAS spectra and can be used to interpret XPS spectra. Finally, we fit experimental spectra with our simulated XAS and XPS references to provide \textit{ab initio} atomic-level chemical information about carbon nanomaterials. Characterization by utilizing computational references lets us extend the scope of this work beyond electrochemistry, since the method is generally applicable [4–6, 27].

In conclusion, we combine high quality experimental data with novel simulation techniques to study and characterize carbon-based materials, including a-C, diamond and graphene – with functionalization. The information gained with these methods can be used to create tailor-made materials for a variety of applications.
II. EXPERIMENTAL FRAMEWORK: AN OVERVIEW AND MOTIVATION BEHIND THE COMPUTATIONAL WORK

A. Electrochemical measurements

The main focus of this thesis is in using computational methods in order to understand experimental reality at the atomic level. In other words, the goal is to shed light on to those areas that cannot been studied experimentally, and at the same time, to find connections between experimental and computational results.

Sources of error are often difficult to determine because there are so many variables: possible impurities, errors caused by the devices or code, human error, etc. However, if computational and experimental results support each other, it is safe to say that we are standing on fairly solid ground.
FIG. 2. The reactions of neurotransmitters, especially in vivo, can be very complicated and there are several active interfering agents and other variables. Dopamine (DA), which belongs to a group of cathecols (CA), is given here as an example of a multitude of pathways that are possible for these molecules [15]. (Copyright T. Palomäki, reproduced with permission). Electrochemical reactions (E), that are accompanied with electron transfer, are denoted with double arrows and chemical reactions (C) are denoted with one arrow, respectively.

Since the main focus of this thesis is about computational work that is connected with the experimental results, we will discuss the experimental background only briefly. However, experimental work stands behind the computational studies and provides the motivation for the calculations as well as works as a driving force and helps to find out where computational work should be targeted. Publication [1] provides an overview about the concept and the
Let us start with an introduction to electrochemistry. In the following sections we will focus on the electrode materials in question and on the characterization that is done via X-ray spectroscopy.

An ideal cyclic voltammetry (CV) response of a reversible reaction is given in Figure 1. In the figure a molecule ‘A’ is first oxidized to a new form ‘B’. The reaction is accompanied by removal of an electron from A. Then when the potential of the electrochemical cell is shifted backwards, the reaction is reversed and B reduces back to A. However, needless to say, the situation is often more complicated. There can be several reactants in the electrolyte solutions and peak potentials (see Fig. 1) that correspond to certain reaction paths may overlap. Also, it is possible that there are unknown reaction products and unknown reactions, which cannot be interpreted from CV curves without prior knowledge.

The dopamine (DA) oxidation path is given as an example in Figure 2. Ascorbic acid
AA) is usually the most interfering molecule in the solution, because of its concentration and because of its oxidation behavior. Dopaminequinone (DAox) is a reaction product of DA oxidation, which may be reduced back to DA completely, but not always. It may react chemically further to form new reaction products which can also be electrochemically active. The reaction products may subsequently polymerize and agglomerate on the electrode surface, which causes degradation of the electrode performance [9, 13]. DA and AA are both inner-sphere redox systems [18]. This means that their oxidation/reduction behavior depends strongly on the interaction between the molecule and the surface of the electrode [17]. This is where computational science can help: to see beyond what can be studied experimentally [2, 8]. More information about electrochemical measurements and of biomolecules on carbon-based electrodes can be found, for instance, from [7, 15, 16, 29, 30].

B. Experimental X-ray spectroscopy characterization

X-ray photoelectron spectroscopy (XPS) is used for measuring the composition and electronic structure of a sample in question. It is a widely used method for characterizing condensed matter phases. Solids, surfaces and interfaces are the most common type of samples to be studied with XPS [26].

The measurement is carried out by reflecting stable X-rays, whose energy is known, to the sample. A monochromator is used to control that the beam contains only single wavelength X-rays that hit the sample. When the beam reaches the sample core electrons from the atoms in the specimen are excited. Excited electrons are collected by a detector. Core electrons of different sites, or atoms, in the sample have different binding energies. When the energy of the X-ray beam is carefully targeted around selected energy range, the resulting binding energy spectrum reveals the composition of the sample [16, 26].

Different elements can by studied one at the time when, for instance, C 1s and O 1s can be plotted separately. Or one can scan the whole sample with a wider energy range (called wide scan) to obtain absolute intensities of the elements in the sample. This way the proportions of the elements can be compared. However this may lead to loss of accuracy, thus for studying individual elements and their bonding environment, focusing on their specific energy range is often preferred [16]. The main defining factor of the accuracy of the measurement is however the resolution of the instrument, which can vary a lot. The best
resolution can be achieved with synchrotron facilities [26]. Which takes us to discussion about X-ray absorption spectroscopy XAS.

XAS measurements are done with a synchrotron lightsource, for instance the Stanford Synchrotron Radiation Lightsource [28], which has been used to obtain the experimental XAS spectra used in this work. A schematic image of the system is depicted in Figure 3 from [16]. The main difference between XPS and XAS, even though they are closely related, is that in case of XPS the electrons are excited from the system and their binding energy is measured, but in the case of XAS electrons are excited to the conduction band (See Fig. 5). In other words, with XPS electrons are excited beyond the vacuum level, whereas XAS measures how much energy is needed to excite an electron from an occupied state to a conduction state. Since more sophisticated instruments are used to perform XAS measurement, it is a far more accurate method and XAS spectra contains a lot more information than XPS spectra from the same sample [4, 16]. However, XPS is more widely available and utilized more often in experimental science.

III. BASICS OF DENSITY FUNCTIONAL THEORY

A. Theoretical background

In principle density functional theory (DFT) lets us examine physical chemistry of the system in question ab initio, i.e., via quantum mechanical calculations instead of using experimental input. In other words, properties that cannot be measured can be studied, and then, can be compared with the experiment. Both methods, computational and experimental, complement each other. In this section we go through the very basics of DFT in order to explain the origin of the results. The formulae are based on [31–35].

DFT is originally based on the Hohenberg and Kohn theorems [36] and later on it has been developed to be used for several purposes. The main focus is on finding the total energy of the system, in principle the calculation – based on Schrödinger equation – is exact. However, in practice, approximations have to be made in forms of exchange-correlation functionals originally based on the Kohn-Sham (KS) ansatz [37]. In order to accommodate different kinds of systems to be studied with DFT, a huge variety of these functionals have been developed. The one used in this work is Perdew, Burke and Ernzerhof (PBE) [38] which is
a safe choice with carbon as well as with the excited state calculations [39]. There is also a wide range of programs, based on the atomic simulation environment (ASE) [40, 41], that could be used in order to employ DFT. Calculations in this thesis are done with GPAW calculator [42–44].

The total energy of the system is the expectation value of the Hamiltonian operator, which has to be minimized in order to find the electronic ground state of the system. In principle this means that we need to find an eigenvalue for an operator:

\[ \hat{\Omega} \psi = \omega \psi, \quad (1) \]

where \( \Omega \) is the operator, \( \omega \) is the eigenvalue and \( \psi \) is the function. In other words we have [31]:

\[ (\text{operator})(\text{function}) = (\text{constant factor}) \times (\text{the same function}). \quad (2) \]

We start looking at the quantum mechanical theory with the Hamiltonian operator \( \hat{H} \) [31, 32]:

\[ \hat{H} = \hat{T}_{\text{elec}} + \hat{V}_{\text{ext}} + \hat{V}_{\text{int}} + \hat{T}_{\text{nuc}} + E_{\text{II}}. \quad (3) \]

Each term of the equation represents different factors, that are included in the calculations in order to compute the geometry and the electronic structure of the system in question. The first one, \( \hat{T}_{\text{elec}} \), is the kinetic energy operator for the electrons. The second one, \( \hat{V}_{\text{ext}} \), describes how the nuclei are attracting the electrons, whereas the third, \( \hat{V}_{\text{int}} \), gives the interactions between the electrons. The fourth term, \( \hat{T}_{\text{nuc}} \), is the kinetic energy operator for the nuclei. And the last one, \( E_{\text{II}} \), describes repulsion between the nuclei. The formula can be written out as follows:

\[ \hat{H} = -\frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 - \frac{1}{4\pi \epsilon_0} \sum_{i,j} \frac{Z_i e^2}{|\mathbf{r}_i - \mathbf{R}_j|} + \frac{1}{8\pi \epsilon_0} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \frac{\hbar^2}{2M_I} \sum_I \nabla_I^2 + \frac{1}{8\pi \epsilon_0} \sum_{i \neq j} \frac{Z_I Z_J e^2}{|\mathbf{R}_i - \mathbf{R}_j|}, \quad (4) \]
where capital letters denote symbols related to nuclei and small letters are related to electrons.

According to the Born-Oppenheimer approximation, the mass of the nuclei goes to infinity compared to the mass of the electrons \([32]\). Thus, the nuclei move extremely slowly compared to the electrons. This allows us to neglect the movement of the nuclei, i.e., nuclear kinetic energy \(\hat{T}_{\text{nuc}}\), when electronic structure of the system is computed. In other words, the coordinates of the nuclei work only as parameters in the calculation. The Born-Oppenheimer approximation can be connected with the Franck-Condon principle which states that an electronic transition takes place much faster than the nuclei can respond \([31]\) to. The Franck-Condon principle is an essential concept both in electrochemistry as well as in spectroscopy: the nuclei remain still while the electrons move. Computationally, after the optimal electronic structure is achieved, the structure of a system can be relaxed, meaning that it is let to find its geometrical energy minimum, the nuclei are let to move according to the forces that has emerged between the atoms in the sample. The criterion of successful structure relaxation is achieving a structure with minimum tension – or force – between the atoms. In other words, when the movement of the nuclei was first discarded during the relaxation of the electronic structure, it is then brought back in form of forces acting between the atoms. Then the electrons follow nuclei and the electronic structure is relaxed again. Altogether, the system is relaxed to reach the energy minimum in both; geometrical and electronic structure.

It is a known fact that Kohn-Sham eigenvalues that are being computed in order to obtain electronic structure, have no real exact meaning \([32]\). The only exception is the approximated highest eigenvalue just before the ionization energy, which defines not only the beginning of the band gap of the material, but also the onset of the spectra – both are important features related to this thesis. The difficulty is to find the connections, between experimental and computational science. I cannot remind the reader enough that here we are not comparing functionals or codes, which has already been done in \([39]\), for instance. This is where comparing of the trends between computational and experimental results comes exciting and interesting. Unfortunately there are no exchange functionals that can be in perfect agreement with the experiment. We found and used the safest option: PBE and VdW, in fact, more expensive combinations can be less accurate \([32, 39]\). The problems can be similar as in the experimental circumstances, which have typical (but not the same)
issues, caused by different things, like contamination or oxidation. Shifting of the onset is one of them. Changing to a functional that is more expensive can only make the calculation more inaccurate, because the functional is not calibrated for the purpose [39].

B. Technical details about obtaining the structure of the systems

Every structure employed in this work has been checked to obtain both; converged geometric and electronic structure. Some examples of the structures are presented in Figure 4. It has been an enormous project to calculate, for instance, optimal lattice parameters and finding the lowest energy configuration for all pristine, doped, functionalized and defected carbon-based models containing approximately – more or less – 200 atoms each. The DFT code GPAW was used in all cases [42].

Also ML based structures [12, 45, 46] were further relaxed at the DFT level. The coordinates for the molecules, DA, AA and DAox, that were used in adsorption energy calculations were originally obtained from [47]. In all cases, generalized gradient approximation PBE [38] functional and Van der Waals (VdW) corrections were employed via the method developed by Tkatchenko and Scheffler [48, 49].

In the case of doping as well as adsorption energy study [2, 8], a continuum solvent model (CSM) [50, 51] was used in order to simulate aqueous environments. The doping study [2], which will be discussed later (Sec. VA), and is included in this thesis, is aimed to have a more detailed information of what happens to the molecules and electronic structures of the systems at carbonaceous surfaces when the potential of the surface, and thus its charge, is changed. In other words, to shed light on to what happens during electrochemical reactions. This kind of information cannot be obtained experimentally. Doping is a novel method to study electron transfer between the surfaces and the molecules in question.

C. X-ray spectroscopy calculations

After the ground state structure – both electronic and geometric – has been established, excitation calculations can be carried out. Figure 5 shows a pedagogical example of how the XAS spectra depend on the electronic structure of the sample. The XPS signal, known as Δ Kohn-Sham (ΔKS) in this thesis corresponds to the onset of the XAS spectra. The
method is based on the Khon-Sham approach introduced in Refs. [32, 55]. In this work we are utilizing the GPAW implementation by Ljungberg et al. [42, 44, 56, 57]. The theory as well as the method has been described in great detail in the aforementioned references as well as in publications [4–6].

It is important to note that when the convergence of calculated XAS spectrum is studied, in addition to energy values one must also take intensity values into account. In addition to convergence testing, several other tests were carried out in order to validate the method; only few examples are described in this thesis. We wanted to make sure that the code performs well for our purposes with carbon and oxygen containing substances. The study about the differences between X-ray spectra signals of graphene, amorphous carbon and diamond is explained in more detail in Section V D and in Refs. [1, 4]. The method was also tested with small molecules: benzene and formic acid (Fig. 6). The effect of the nearest neighbors of the site that is being probed was tested with a so-called “graphene test” (Fig. 7).
FIG. 5. Schematic image of how the calculated XAS spectra reflect the electronic structure, i.e., density of states (DOS) of the material in question [4] (Copyright: https://pubs.acs.org/doi/10.1021/acs.chemmater.9b02049).

The small molecules, benzene and formic acid, serve well for validation of the method. The first one has been used for testing and validation also previously [20] and the latter contains oxygen and has a sharp peak, that can be related to carboxylic acid, which is anticipated to be one of the most solid experimental references – also in literature. The experimental spectra are taken from [52–54], respectively. This is a different version from the figure depicted in the last publication of this thesis [6]. The one in the article is selected for clarity. This one here is selected because it contains more information in a form of having additional spectra. Here, in case of formic acid, gaseous and crystalline samples are being compared. The panel about formic acid demonstrates that both experimental and computational spectra can differ depending on how the samples are fabricated or simulated, respectively. However, the main peak positions remain the same. There is a lot of fluctuation between carbon-based experimental samples, depending on how the samples are prepared. The computational studies that are presented in this thesis, as well as this experimental work Ref. [14], aim to show trends to understand the main reasons behind this variety.

The graphene test shows that an addition of a functional group, -H and -O in this case, on a certain selected site, site (a) in Figure 7, marked with a red circle, changes the XAS fingerprint spectra only locally. The effect of the functionalization on the shape of the
FIG. 6. Validation test with small molecules: benzene and formic acid. The experimental spectra are from Refs. [52–54]. The calculated C 1s spectra and C 1s experimental spectra are plotted together to show well they agree. The lower panel, formic acid, shows that the spectra can differ due to the sample preparation process, however the agreement between computationally and experimentally obtained spectra is obvious.

spectra declines quickly: see sites from (a) to (f) in Figure 7. Site (f) is far away from the functionalized site and resembles pristine graphene. The effect of the defect is discussed more in Sec. V D. Results and conclusions of the X-ray spectroscopy studies are also summarized in Sec. V C.

IV. MACHINE LEARNING FOR MATERIAL SCIENCE OF CARBON-BASED MATERIALS

A. GAP: an interatomic potential for obtaining the structure

When it comes to future prospects of DFT, or some other computationally more demanding method, they allow us to provide data for ML-algorithms. One cannot predict anything with machine learning only; the model has to be taught. Computationally more demand-
FIG. 7. The graphene test shows how local the effect of functionalization is (Ref. [4]) (Copyright: https://pubs.acs.org/doi/10.1021/acs.chemmater.9b02049).

In this thesis, from ML point of view, we focus on utilizing ML-based structures, smooth overlap of atomic positions (SOAP) clustering and visualization via multidimensional scaling (MDS).

Since this thesis is intended to be presented to a very interdisciplinary audience, let us start from the very basic principles, based on Refs. [58–60]. In supervised machine learning there are three main concepts: 1) the data set, which should be as large as possible, 2) a computational model that connects the input with the output and, finally, 3) a loss function for estimating the error that the model gives compared to the input. The data, which is the input, is used to teach the model to produce the same outcome with a lot less computational effort, or not to mention human effort, than what was used to produce the input data. This is done by interpolating, or predicting, what happens between the original data points. And,
thus, learning the space between the data points. The size and quality of the data set limits how much the ML model can learn, because it cannot “see” beyond the training set.

The performance of the model can be validated by comparing the results with the original test set. In other words, if one tries to calculate or predict where some original data point, that was used as part of the input, lies, how close or far does one get with the ML algorithm? However, this kind of validation only tells if the model correlates with the input data. It does not guarantee that the model is perfect. If the input data has been wrong, i.e., low quality, or there’s not enough data to cover what one wants to learn, it may not correlate with reality. Thus comparison with experimental work – or common sense – is always in place. Data usually has some “features” – also called “descriptors” – that can be observed, computed or measured easily. Then there are questions that are a lot harder to answer, because the outcome is not that obvious. Such as: “What is the true structure of a-C?”.

A ML algorithm and a combination of proper features can lead to better understanding of the link between the features and the desired property, called “label” in ML jargon. This can apply generally on almost anything; from structure of a chemical compound to some random images. In addition, proper “hyperparameter” optimization has to be carried out. Hyperparameters are selected when the input is given to the algorithm. They differ from “free parameters” that are defined when the model is fitted to the data; and there might be a significant number of them. It is important to remember that the algorithm itself not understand anything about chemistry or physics – or whatever it is that is being computed. A human touch is required to make sure that the results are correct.

The clustering method that we are using here, presented in more detail in Sec. IV B is an “unsupervised method”. The difference between “supervised” and “unsupervised” is that the first one requires that the labels that are being studied are defined by someone or something, whereas in the case of the latter labels are automatically defined. Although, according to [58], the distinction between the two can be blurry.

Now that we have gone through the basics – and some vocabulary – we can focus on the applications that are used in this thesis. Most of the a-C structures used in publications [3–5] are created with Gaussian approximation potential (GAP) [12, 45, 59]. SOAP [3–6, 61, 62], behind this potential, will be introduced in Sec. IV B.

It needs to be mentioned that we are not creating these structures within the framework of this thesis, we are only adopting them to be used for our purposes, thus, they will be
discussed only briefly. The GAP potential that was used to create the structures that we employed for the calculations, has been trained with DFT-level data [45]. SOAP was used to find suitable descriptors, i.e., to classify atomic sites from DFT data in order to help the potential to learn the gaps between the original data points. In this case, interatomic potential, or force field, is interpolating with an assumption that similar sites possess similar atomic energies and interatomic forces; which then can lead to larger structures that are very near to DFT accuracy, but can be created with a lot less computational effort. This amount of large structures could not have been created at the DFT level.

Surface samples were realized by carving from a melt-quenched bulk sample, created with GAP, and subsequent annealing. Details of the computational protocol, as well as the original structures, are published in [12]. For the work presented in this thesis, all of the structures were further relaxed at the DFT level using the GPAW code with the PBE functional and VdW corrections.

A very profound overview of Gaussian process regression (GPR) and machine-learning methods that are used in computational chemistry and materials science is provided in a review by Deringer et al. [59]. GRP together with SOAP provide the platform for using GAP in order to simulate structures that can then be further utilized in computational work, which aims to understand the chemistry of the structures in question. In this thesis we focus on surface chemistry. In addition to what is presented here, further examples of how ML potentials can be used to increase our physicochemical understanding can be found from the recent work by Caro et al. [63, 64].

B. SOAP: a tool for classification of the single sites within the structures

Machine learning based structure construction allows us to have computational models, that are big enough to provide good statistics. However, it also brings an issue, especially in case of more disordered materials such as a-C and defective G and GO. How to classify logically thousands of geometrically inequivalent sites?

For this reason we employ a featurizing algorithm: SOAP (smooth overlap of atomic positions) [3–6, 61, 62]. SOAP is also the basis behind the GAP potential. Similar environments can be grouped together by utilizing an unsupervised machine learning technique, which in this thesis is based on k-medoids [24, 65]. The term “medoid” denotes the sites that are
the most equivalent to the center of each cluster. In other words, SOAP is a ML-based tool that lets us to have an automated approach for reaching a high dimensional representation of all the sites in the samples according to their chemistry. In our case, according to the geometry, i.e., bonding environment of the site with respect to its nearest neighbors. In general, when chemical compounds are studied by utilizing SOAP, the atomic sites in the sample that are invariant with respect to translation, rotation and permutation are expected to behave similarly.

Clustering algorithm uses kernels (or so-called ML jargon functions) to optimize the result. Similarity is computed with SOAP kernel, which is a dot product of two SOAP vectors:

\[
    k^{\text{SOAP}}(i, j) = (q_i \cdot q_j)^\zeta, \tag{5}
\]

where \( q_i \) and \( q_j \) are the SOAP vectors, that are computed in order to obtain the characteristics of the sites \( i, j \) in the sample, respectively. Typical values for \( \zeta \) are small positive numbers, e.g., \( \zeta = 4 \) [45, 66]. The kernel is defined between 0 (totally dissimilar sites) and 1 (equivalent sites).

From the kernel, a measure of similarity – or distance between the sites – can be defined as follows:

\[
    D^{\text{SOAP}}(i, j) = \sqrt{1 - k^{\text{SOAP}}(i, j)}, \tag{6}
\]

where \( D^{\text{SOAP}}(i, j) \) is the similarity matrix which includes the whole data set.

The accuracy of the method can be improved by adding more features, such as local density of states (LDOS) kernel or Delta Kohn-Sham (\( \Delta KS \)) kernel. This may be necessary, if several elements are present in the sample, which makes the task a lot more complicated. Here is an example of how to include an extra feature, \( \Delta KS \) kernel, in this case:

\[
    k^{\Delta KS}(i, j) = \exp\left(-\frac{(\Delta KS_i - \Delta KS_j)^2}{2\sigma_{\Delta KS}^2}\right), \tag{7}
\]

where \( \Delta KS_i \) and \( \Delta KS_j \) are the \( \Delta KS \) values of sites \( i \) and \( j \), respectively. LDOS kernel can be obtained in an analogous manner. These values – \( \Delta KS \) and LDOS – must be computed at the DFT-level.
Finally, the combined kernel, $k^{\text{tot}}(i, j)$, is a linear combination of SOAP and $\Delta$KS kernel (or LDOS kernel):

$$k^{\text{tot}}(i, j; x) = x k^{\text{SOAP}}(i, j) + (1 - x) k^{\Delta\text{KS}}(i, j).$$  (8)

This kernel, $k^{\text{tot}}(i, j)$, can then be used to create the similarity matrix, $D(i, j)$, which is named here at a more general level, but the equation doesn’t differ from $D^{\text{SOAP}}(i, j)$. In all cases, in addition to aforementioned $\zeta$, several other hyperparameters must be optimized. The SOAP cutoff, which defines the area in the sample that is being observed, i.e., the radius that defines spherical region around each atomic site that the SOAP algorithm takes into account. Then $\sigma_{\text{atom}}$, which tells the algorithm how rough or smooth the area within the cutoff should be. In case of a combined kernel there is also $x$ in the linear combination and $\sigma_{\Delta\text{KS}}$ in the $\Delta$KS kernel, to name two. Also the right number of clusters must be fixed (for more information look for SOAP [3–6, 61, 62] and $k$-medoids [24, 65]). In this thesis, stochastic methods were used for hyperparameter and cluster size optimization until the combinations of the clusters were sound and in agreement with chemical intuition.

The only difference between $D^{\text{SOAP}}(i, j)$ and $D(i, j)$ is that the latter contains additional feature. In principle, one could add as many features as needed, as long as computational time (and human time) allows. It must be kept in mind that DFT-level calculations – not to mention GW-level [67–69] – are much more expensive than ML calculations, and thus, the benefits of using very fast ML techniques could be lost.

C. MDS: two dimensional visualization of multidimensional system

In order to effectively visualize the clustering results – obtained by $k$-medoids algorithm – again a new ML-based tool, MDS (multidimensional scaling), was employed [3, 5]. Figure 8 illustrates the workflow. When SOAP compares the differences, or distances, between the atomic sites, the resulting matrix can be difficult to interpret with a naked eye. MDS provides means to do so by reducing dimensionality. It is, again, an iterative method, that aims to preserve the original distances, but projects the data points into a 2D visualization plot. The python-based code that we are using is introduced in more detail in Ref. [70].

Visualization is important in order to detect whether or not the clustering has been
FIG. 8. Schematic image of the workflow and connection between data management via clustering and visualization via MDS. First raw data is classified according to its properties into clusters that can be connected with desirable features. Then MDS visualization provides means to study connection, or similarity, between the clusters.

In other words, \( k \)-medoids and MDS go hand in hand. If the 2D plot shows that the clusters are “all over the place”, so to speak, there must be something wrong with the clustering. In other words, in Figures 8 and 9 data points that represent a cluster are neatly grouped together, which means that the clustering has been successful. If that was not the case the data points would be scattered. There are also other ways to validate and control the clustering scheme, such as making sure that the bond lengths within the clusters in question remain approximately the same, or inspecting that the X-ray spectroscopy signals of the clustered atomic sites do not differ too much, as has been done in [6]. Then again, when X-ray spectroscopy is being studied, these outliers may have an interesting explanation, which could help us to understand experimental results. This, of course, demands a lot of intense manual labor. A MDS plot reveals quickly if there are outliers in the data set. Having said that, both SOAP-based representation as well as the MDS algorithm require hyperparameter and parameter optimization, optimization of the number of the clusters, and optimization of the number of random initializations with respect to possible medoids, to see how many times the calculation will have to run in order to produce the best result. Consequently, it
is quite tricky when two methods are combined, especially, if several elements are involved. Even if it is computationally cheap and fast, the human time that is consumed, quickly comes expensive. However, the MDS algorithm provides an error estimation, and the result with least “stress” between the data points, should be the best, which is convenient. Also it makes it easier to locate the medoids of the clusters on a 2D map. An example of projecting clustering into 2D MDS map is given in Figure 9 from [5].

V. RESULTS AND DISCUSSION

A. Effects of changing the potential on adsorption energy

Study about doping of the carbonaceous materials, that are discussed throughout this work, is a case study to probe the potential dependency of electron transfer between carbon-based electrode and molecules that are being examined in experimental work [2]. This study as well as Ref. [8] provides the first links between computational and experimental work,
from the point of view of this thesis. Simulating potential differences in aqueous solution is far from straightforward with DFT, and with our doping approach [2] – second publication in this thesis – we are trying out a new method to do so.

We have chosen to approach the problem indirectly via shift of the Fermi level, and at the same time, changing the work function. This is in fact the same basic phenomenon that takes place when potential of an electrode is changed by using a potentiostat. In that particular study, this has been done by introducing dopant atoms. This means, that by replacing 0–3 carbon atoms by either with boron or with nitrogen, where boron reduces the number of electrons in the system and nitrogen increases it – thus we can move the Fermi-level down and up, respectively. It was tested that adding more dopant atoms to the systems of this size (125–144 atoms) did not give more added value with respect to what was our intention, which will be explained next in more detail.

Boron works as an electron acceptor, whereas nitrogen works as an electron donor. The introduction of boron causes that the Fermi level of the surface decreases, and thus, the work function increases. Analogously, with nitrogen, it is the other way around. In practice, this is equivalent to tuning the surface potential externally. The goal was not to study the effects of doping atoms on adsorption itself, and therefore, dopant atoms are positioned as far away from the adsorption site as possible, so that they wouldn’t interfere with the adsorption process.

The molecules in question are dopamine (DA), ascorbic acid (AA), and dopaminequinone (DAox). AA is included, because it is the most common interfering molecule in DA measurements. The reaction paths of neurotransmitters can be extremely complicated (see Sec. II A). An example of an anticipated oxidation/reduction path is given in Figure 2. As can be seen from the first step DAox is an oxidation product of DA. Both DA and AA are inner-sphere redox systems [18]. The term inner-sphere suggests that the oxidation/reduction behavior of these molecules strongly depends on the interaction between the analyte and the surface of the electrode [17]. The adsorption energies of these molecules on the surfaces in question were calculated both in vacuum as well as in dielectric medium, i.e., simulated aqueous environment via CSM [50, 51].

We are not aiming to present exact, or absolute, numbers. Instead, we seek trends within the accuracy of the method. This approach applies to all aspects in Ref. [2]: adsorption energies, potential variation and differences in behavior of graphene, amorphous carbon
FIG. 10. The changes in the interaction between the electrode surface (in this case doped graphene) and dopaminequinone, when dopant atoms are added. The figure shift from boron-doped to undoped, and finally to nitrogen-doped graphene; in vacuum (a)-(c) and in dielectric medium (d)-(f). Charge density is moving from the violet to the green regions. This indicates adsorption especially in case of (f). DAox on graphene has been chosen as an example system for visualization because it provides the clearest image of the phenomena, but the rest of the systems behave in a very similar manner [2] (Copyright: https://doi.org/10.1063/1.4986521).

The results suggest that DAox forms the strongest interaction with the surfaces of the electrodes. Figure 10 shows how the electron density is transferring between the surface and the molecule. Aqueous environment is enhancing and stabilizing the changes in electron density. The electron density was determined by using Bader analysis [71, 72].
This reaction product – DAox – may, thus, be the cause of agglomeration of the organic material on the electrode surface, and so, be the cause of degradation of the electrode performance [9, 13]. Interestingly, in vacuum diamond surface seems to be more resistant to this behavior, implying that \( sp^3 \)-bonded carbon would not be deactivated so easily. However, when dielectric medium is added the situation is reversed. Adsorption energies of DA and AA differ mainly only on graphene, which suggests that \( sp^2 \)-bonded carbon on the electrode surface could aid in increasing the selectivity of the measurement. The reaction mechanisms are complicated and not completely known. Some experimental studies have established hypotheses about the reaction paths [9, 73, 74]. Computational studies, like the one presented here, are the way to find out more detailed information about the reactions that cannot be studied experimentally at the elemental level.

B. Understanding the structure of the disordered carbon materials

In this study [3], we combine ML based interatomic potentials, ML clustering techniques, and density functional theory calculations in order to understand reactivity of a-C surfaces. With these tools a comprehensive, qualitative and quantitative, analysis is performed. We focus on reactivity towards hydrogen (-H), oxygen (-O), hydroxyl group (-OH), and carboxylic acid (-COOH), which are the functional groups that are anticipated to be present on carbonaceous surfaces [20]. In this article we focus on a-C, but the presented new systematic and general methodology can be applied for several disordered materials, such as graphene oxide, as we show in the last article of this thesis [6].

As stated before, understanding the surface chemistry of the electrode material is the key for understanding electrochemical reactions. This opens a possibility for these materials to be utilized for real-time electrochemical detection of biomolecules \textit{in vivo} [7] that could be used in a several medical applications. In addition to that, carbon can be used in a variety of other applications, as is discussed in [4]. Carbon-based materials are widely applied as protective coatings and, because they are bio-compatible, they are also applicable for biomedical coatings [75, 76].

Other elements, such as hydrogen and oxygen, which are mainly studied in this thesis, can be present in elemental form or as part of different functional groups on the surface. Naturally, the presence of these elements has an effect on the properties of the material in
question. Not only the electrochemical properties [8, 20, 21] are affected but also, for instance, the band gap [22], and even mechanical properties and the friction performance of the materials [23] can be varied. This is why it is important to have a closer look at the reactivity of carbon-based materials as has been done in [3]. Later on we have extended the DFT-and ML-based computational collaboration for characterization by combining computations with experimental work. This will be discussed in more detail in the section VC.

This work [3], however, is focused on the reactivity of differently bonded carbon. First, carbon sites were clustered by employing $k$-medoids. In other words, they were separated according to their chemical environment based on the connection with their nearest neighbors. Bond angles and bond lengths were the main cluster defining factors. In this case only carbon atoms were considered when clustering was carried out. Six clusters were found to be the optimal solution (Fig. 11). After the clustering process the sites were functionalized with the previously mentioned groups: -H, -O, -OH and -COOH (Fig. 12). The reactivity of the sites was probed in different ways. Calculated adsorption energies ($E_{ad}$) are reported in the article. The main finding is, that the reactivities of the carbons sites depend strongly on their geometry. Different clusters are clearly separated when adsorption energies are compared. Moreover the strength of the bond between functional group and the surface correlates strongly with the local density of states (LDOS) of the carbon site in question (Fig. 13).

### C. Characterization by computational X-ray spectroscopy: General discussion

This section is probably the most important part of this work. We use computational tools directly to interpret experimental data. And, vice versa, we verify that our computational models are realistic compared to experimental samples. Traditional deconvolution of experimental X-ray spectroscopy, both XAS and XPS, is challenging, to say the least, because the reference data varies and can be even controversial in the literature [14, 20, 77–92]. Ab initio calculations can overcome these obstacles, since there are no prior expectations – or human introduced bias – that could affect the result. In this work, DFT (Sec. III), ML-based structures (Refs. [12, 45, 46, 59]) as well as unsupervised ML techniques (Sec. IV) are employed in order to get a deeper understanding about the carbonaceous materials in question.
FIG. 11. The six typical motifs that were found in a-C. Each figure represent the medoid of the cluster in question. Red atoms are the atomic site that is being clustered and yellow atoms are the nearest neighbors [3] (Copyright: https://pubs.acs.org/doi/10.1021/acs.chemmater.8b03353).

FIG. 12. Functional groups that are being studied in this work. Carbon, oxygen and hydrogen atoms are represented in yellow, dark red and white, respectively. [3] (Copyright: https://pubs.acs.org/doi/10.1021/acs.chemmater.8b03353).

The method relies on obtaining an optimal set of computational samples by utilizing DFT- [4, 93] or ML-based [12, 45, 46] model structure construction followed by structure relaxation. Relaxation is done with fixed parameters in order to maintain a systematic and uniform approach (Sec. III and Refs. [3, 4]). Then, after the optimized structure is achieved, excited state calculations are carried out with that model. Then, finally, the computational spectra can be compared with experimental spectra. With this fitting method (Secs. V E- V F and Refs. [5, 6, 11, 13]) we can provide semi-quantitative estimations of the concentrations
FIG. 13. (a) Adsorption energies ($E_{ad}$) of the functional groups vs. integrated local density of states (LDOS) for each site in each cluster. (b) Average local density of states (LDOS) of each cluster. Note that the cluster 6 has so low reactivity that adsorption energies were not reported. Vertical dashed lines describe the integration interval. Cluster 1 is excluded from the plot because of too small sampling. [3] (Copyright for (b): https://pubs.acs.org/doi/10.1021/acs.chemmater.8b03353 SI).

of elements bonded in different ways.

The method is limited to DFT accuracy. However, that is a tiny problem compared to a situation where the database of chemically sound bonding environments is limited. In other words, there needs to be enough computational samples to represent the complexity of the
materials in question. For this reason we have used large databases of both ML- and DFT-based models in order to get wider understanding of the X-ray spectroscopy fingerprints of pristine and functionalized amorphous carbon, graphene and diamond. Having said that, there would be a lot of room for research to widen the approach for different elements, such as nitrogen.

It needs to be said that in experimental reality the experimental sample is never absolutely pure; whatever the chemical or physical fabrication, transportation or preservation process might be. That is, simply because the yield of the sample fabrication process cannot be 100% and transportation as well as preservation method can damage the experimental sample. Thus, careful characterization is always in place in order to draw systematic connections between the behavior of the sample and its structure, and in order to validate the results – whatever they might be. For instance, in our case, the results of electrochemical measurements are compared with X-ray spectroscopy characterization to make sure that the electrode that is being used during electrochemical detection actually is what it is anticipated to be.

As mentioned before, we cannot compare the characterization results directly or completely with literature references. That is difficult because there is so much variation within the literature references with XPS and XAS [14, 16, 25, 77–92]. Computational interpretation cannot be bribed. The model is what it is. When ab initio calculations are compared with an experiment, we loose the element of human error – or expectation. But when computational references agree with literature, it is truly safe to say that we are on solid ground. This method will be even more accurate in the future when GW technique [67, 68] is being developed further. In addition, GW corrections can be made computationally less expensive by utilizing machine learning. In any case, the computational references have smaller weight of human handling and no environmental effects.

On the other hand, computational references have indeed attracted a lot of attention in interpretation of X-ray spectroscopy. In case of XPS, earlier research has shown that computational references can be used in deconvolution of experimental work [94–103]. Calculated references have also been employed in the field of XAS characterization [104–110]. For instance, a large computational database has already been created for crystalline/more organized materials: The Materials Project website [107, 111, 112]. There have also been similar attempts, that we are using here, to quantify X-ray spectroscopy results by utiliz-
ing different levels of theory [105, 109, 110] or by employing ML-based methods [107, 108]. Some attempts to combine different analytical methods in parallel have also been carried out [109, 110, 113].

However, in case the of carbon and organic molecules the situation is more complicated. For instance The Materials Project concept [107, 111, 112], does not perform well with carbon based structures [112]. The GPAW implementation by Ljungeberg et. al. [42, 44], which we are using, is designed and validated with carbon and oxygen. In addition to structure construction of a-C and other disoriented materials [59, 63, 64], great ML related tools have been created, for instance, in order to understand organic chemistry and to change the understanding of medicinal chemistry [114] – these approaches might even be called game changers. But they do not yet cover characterization, at least not at X-ray spectroscopy-level. That said, these tools provide a lot of potential for utilizing this method, that we are presenting here, in several applications. Also with constantly increasing computational resources, it looks promising – assuming that there would also be more human resources – that this technique could be used in a lot wider context, than what is presented in this thesis. As a personal note, I would like to add that, this approach has been referred to as an avantgarde in the interpretation of experimental results in many forms.

D. Part I: X-ray spectroscopy fingerprints of clustered a-C motifs

Previously in Ref. [2] we compared calculations between graphene, a-C and diamond. They behave differently because of their inherently different electronic structures. Naturally, this applies also to X-ray spectroscopy; the materials in question have different peak positions (see Fig. 5). This is also true when functional groups are being measured [1, 4], although it is now often clear when literature references are being studied [14, 16, 25, 77–92]. Because of this incoherence we started looking into creating computational tools that could not only explain the differences between graphene, a-C and diamond, but that could also rationalize characterization of complex materials, such as a-C.

In addition to functional groups, we can study defects, that are often present in experimental samples. This is particularly interesting in the case of graphene, which is in its pristine form extremely well organized, but when defects and/or functionalization is introduced, the X-ray spectroscopy signal is dramatically changed (Sec. VF and Ref. [6]).
FIG. 14. Calculated C 1s spectra of pristine a-C surface sites compared with functionalized sites of the same kind. Columns labeled 2 to 6 correspond to different motifs. The first row shows the spectra of the sites belonging to the clusters which the motifs 2-6 represent without any functionalization. The subsequent rows show how the spectra of the sites belonging to each cluster change as functional groups are introduced. These groups, -H, -O, -OH and -COOH, are depicted on right side of the figure. The atom that is being probed is marked with a circle. Motif 1 is discarded from the figure, since there was not enough sites for statistically meaningful comparison in the data-set. Also functional groups were not added to sites in cluster 6, since these sites are fourfold coordinated (sp$^3$), and they are not reactive [3]. Thick colored lines show the average XAS spectra of each cluster, individual spectra of the sites within the clusters are plotted with light-grey lines, the standard deviation around the average is shown with dark-grey shaded areas [4] (Copyright: https://pubs.acs.org/doi/10.1021/acs.chemmater.9b02049).
However, defects were first taken to be part of this study [4] for a different reason: pristine graphene does not offer a favorable bonding environment for functional groups, whereas SV defect does. SV defect was also used in testing of how addition of a functional group affects the spectra of other sites in the sample, especially the nearest neighbors. The graphene test is explained in Section III C. Figure 7 provides also information about how the sites around the defect differ from pristine sites. There are clear geometry induced differences between the sites (b) - (e) (Fig. 7), that are explained by the fact that they are located around the defect, which affects the bond lengths and the bond angles. Site (f) is located far away from the functionalized site and X-ray fingerprint corresponds to pristine graphene. In other words, Figure 7 also shows how an ideal graphene site differs from the sites around the defect.

In amorphous carbon there are no geometrically equivalent sites. The material is “defective” by default; bond lengths and bond angles differ in every case. Luckily, clustering provides the means to study also a-C systematically (Sec. IV B). After the new classification method for a-C [3] was established, we were able to use the clustering also for other purposes. In Ref. [4], we employ the classified structures in order to understand experimental X-ray spectroscopy – XAS and XPS to be exact – better. Thanks to ML-based large data-set of computational samples, we were able to calculate statistically sufficient amount of spectra from clustered sites (published also via Zenodo [115]). These spectra represent the variety of differently bonded carbon and oxygen sites present in a-C samples – and they can be directly compared with experimental X-ray spectroscopy results. The method is valid for both C 1s and O 1s spectra. This allows us to examine qualitatively which motifs or functional groups are present in the experimental sample.

E. Part II: Fitting experimental X-ray spectra with re-clustered fingerprint spectra

In the first part of this X-ray spectroscopy study [4] we computed spectra of carefully classified a-C samples, which enabled us to give qualitative estimations of how each type of carbon cluster or carbon site that is bonded to a functional group contributes in experimental spectrum. In the second part of the study [5] we are taking the idea even further: the ambitious goal is to interpret experimental spectra quantitatively. In order to do that we
need a code for fitting and sufficient database of spectra for both carbon and oxygen. In addition, we explore whether or not XAS and XPS spectra of the same sample can be fitted in parallel; this will be discussed a bit later on.

Let us start with the fitting code. This in-house made code allows us to fit an experimental spectrum with a weighted linear combination of calculated spectra. During the fitting, the complete shapes of the fingerprint spectra are utilized, in contrast to traditional fitting, where only peak positions – based on literature references – are used [14, 16, 25]. However, also this novel way of thinking has its “flaws”. Or more precisely, as accurate characterization methods XAS and XPS can be, there lies an inherent problem: overlapping features. When the whole shape of the spectrum is investigated, some features may cover each other and then it becomes impossible to tell which type of bonding is dominating around the particular region, or in the whole spectrum.

To overcome this issue, the fitting code lets us to focus on certain energy range, for instance, the tail of a spectrum can be ignored, since it contains very little information. This way, the idea of fitting the whole shape of the spectrum, instead of single peak positions, is not lost, only the energy range is slightly limited. Another – inevitable – way to improve the quantitative fitting method is re-clustering in order to reduce complexity, as has been done in Refs. [5, 6]. Without this procedure, the fitting code becomes unstable, because it cannot handle too many spectra with overlapping features. The ideal is to use approximately 5-10 fingerprint spectra, that represent certain chemical environment, to fit an experimental spectrum in order to keep the fitting algorithm stable. As has been mentioned before, computational results do not lie and they cannot be bribed, but these aforementioned adjustments do bring some human touch to the interpretation.

That said, in Ref. [5] we have employed clustering again. However, this time the situation was more complicated than before for the following reasons: 1) we needed to reduce the number of reference spectra, 2) we needed to make a difference, not only between differently bonded carbon sites, but also between the sites that are bonded to other elements; hydrogen and oxygen in different ways, and in addition, 3) we needed to rule out overlapping features.

The workflow of our approach is described in Figure 15. Clustering of oxygen sites was a more simple task and the outcome is in good agreement with chemical intuition. Clustering of carbon sites, however, was a lot more complicated since there is so much more data, or sites to probe, and huge variety of bonding environments, and in addition, we
FIG. 15. The workflow of our approach. First we collect data: the structure, ΔKS values and individual XAS spectra from DFT calculations. Then the data is clustered according to a multi-species SOAP kernel or with additional ΔKS kernel, depending on whether C 1s or O 1s is being handled. After this we have the average spectra – both XPS and XAS – of each cluster to be used as the fingerprints of each cluster. Finally, the experimental spectrum can be fitted with these computational spectra [5] (Copyright: https://pubs.acs.org/doi/10.1021/acs.chemmater.9b02050).

needed to somehow address the issue with overlapping features. Oxygen sites were clustered with SOAP kernel only; whereas with carbon we included ΔKS-kernel as well in order to have satisfying classification of all the carbon sites for fitting. More detailed description of clustering kernels is given in Section IV B. Figure 9 illustrates, by utilizing MDS, how the data – both oxygen and carbon – was redistributed by the new clustering scheme.

After we had created suitable sets of fingerprint spectra (published via Zenodo [115]), which are the average spectra of each carbon and oxygen clusters, we were able to start fitting the experimental spectra. Basically, the fitting code minimizes the space between calculated, i.e., weighted linear combination of fingerprint spectra, and experimental spectrum within a given energy range. The calculated weights of the linear combination reveal how much each fingerprint spectrum contributes to the fit.

We were also exploring a possibility of parallel fitting of XAS and XPS spectra from the same sample simultaneously. In principle, these methods should give the same results of the consistency of the sample. The results that were obtained by parallel fitting are,
FIG. 16. (a) Computational a-C O 1s XAS fingerprint spectra of each oxygen cluster and (b) experimental a-C O 1s XAS spectra fitted with this data. (c) Computational a-C C 1s XAS spectra [for the legend, see panel (e) below], are used to fit experimental spectra which are depicted in (d). The fitting ranges, \((E_{\text{min}}, E_{\text{max}})\), are indicated with dashed lines. Panel (e) Is showing the calculated normalized a-C C 1s XPS fingerprint spectra for every carbon cluster. And finally, (f) illustrates the calculated a-C C 1s XPS spectra fitted to experimental spectra [5] (Copyright: https://pubs.acs.org/doi/10.1021/acs.chemmater.9b02050).

however, somewhat controversial. One reason for that is that, even tough XPS is more widely used as a characterization method, XAS is far more sensitive and XAS spectra contain a lot of information that XPS cannot provide. Then again, XAS implicitly contains all the information that the XPS does, which in this work is incorporated via \(\Delta KS\) data. My personal opinion is that so far we do not have one final answer to this dilemma, but the
issue would be worth looking into in more detail in the future.

All in all, the O 1s and C 1s fingerprint spectra and the fitting results obtained with this novel quantitative fitting method are depicted in Figure 16. The fit is not perfect, it cannot be perfect, since we cannot cover the plethora of chemical environments that can be, and are, present in the experimental sample. However, it should be noted that the trends between the experimental sample and computational fit remain the same. Instead of aiming for a perfect fit, we have used the computational data, from which we know exactly where the reference spectra are arising from. This is a clear advantage compared to traditional fitting with various literature references. In other words, we can compare how much each fingerprint spectrum weights compared to the others in the given data-set. One important, and intuitive, conclusion is that C 1s spectra holds more information about how the carbon is bonded in the sample, whereas O 1s spectrum gives more detailed information about functional groups. This conclusion is confirmed when we move from a-C to graphene based samples in the next, and final, Section V F.

F. Expanding the method to other disorganized carbon based materials: graphene, graphite, and their oxidized forms

In the last publication [6] we turned our focus from a-C to graphene-based $sp^2$-rich carbonaceous materials. Given the huge scientific and industrial attention that these 2D or 3D materials possess [116–119], it is quite limited what we actually know about these substances and their properties. Already in 2010, Dreyer et al. published a Critical Review [120] about synthesis and structure of GO. Their main conclusion was that there is no single GO but that the structure, properties and nature of GO depend strongly on the quality of the precursor material, i.e., the graphene or graphite source, as well as the oxidation (or reduction) protocol [120]. This particular article, among several other studies, provided the motivation and demand for us to develop a new method for characterization of this type of materials.

First we got started with systematic analysis of computational graphene models [6, 121], moving from pristine perfectly organized graphene sheet, to something that is almost amorphous, with all its functional groups and defects. This information provides the means to inspect the spectra obtained from experimental samples qualitatively. The assembly of XAS spectra calculated from this sliding scale of carefully selected structures is presented in Fig-
FIG. 17. Calculated C 1s XAS spectra of graphene samples with systematic increase of defect and/or oxygen concentration: (a) pristine graphene sheet, (b) graphene with single vacancy (SV) defect, (c) defective sample with high vacancy concentration (both SV and DV), (d) graphene with some oxygen, (e) graphene with even more oxygen, and finally (f) graphene with high oxygen and defect concentration. The upper row depicts samples without oxygen, whereas the oxygen-containing samples are depicted in the lower one. Pristine graphene features are depicted with dashed lines as references. Calculated individual spectra of the sites are plotted with grey lines, which shows how the number of inequivalent local chemical environments increases when the sample develops from pristine to something that is nearly amorphous. Green solid lines represent the average C 1s spectrum of the sample in question. Schematic images of the corresponding structures are shown next to the spectra [6] (Copyright: https://pubs.acs.org/doi/10.1021/acs.jpcc.1c03238).

As materials, graphene, graphite or their oxidized forms can be as complicated as amorphous carbon structures, and the interpretation of their experimental characterization is just as difficult. Fortunately, we had already developed tools to do so; all we needed was a new data-set based on \( sp^2 \)-rich carbon. In this work, we employed structural models
FIG. 18. (a) Computational XAS spectra of defective (SV and DV) graphene samples without any functionalization, i.e., these samples do not carry oxygen or hydrogen. The depicted spectra are averages calculated from the whole sample, except in the case of pristine graphene, since in that sample all the sites are symmetry equivalent. Corresponding structures are depicted in Figure 17. As in case of Figure 17, the pristine spectrum has been published before in [5] and SV spectrum in [14]. Note that the calculated spectra are depicted as they appear, they are not shifted according to any literature reference. (b) Fingerprint spectra of the defects: SV and DV. Only the sites around the defect were taken into account when the average spectra were calculated. Schematic images of SV and DV are depicted next to the plot. [6] (Copyright: https://pubs.acs.org/doi/10.1021/acs.jpcc.1c03238).

created by Kumar et al. [121] and our previously computed results from graphene based models [4, 5]. Then, this was again followed by SOAP-based clustering, with new hyper-parameter configuration. After that, based on this classification, we obtained a new set of fingerprint spectra, that can be used in fitting of experimental spectra of $sp^2$-based carbon samples. Thanks to the clustering method, the same fingerprint spectra can be used for fitting 2D and 3D materials, since the original sites are disconnected from the original struc-
FIG. 19. (a) Examples of computationally fitted experimental C 1\textsubscript{s} XAS spectra of \textit{sp}\textsuperscript{2}-based carbon samples. The sharp peak around 291.7 eV in experimental data reflects the excitonic effects which refers to long range order. This peak cannot be reproduced computationally with the method we are using, which is discussed in Refs. [4–6]. (b) In the case of the GO sample, there is plenty of oxygen in the sample. Both C 1\textsubscript{s} [in panel (a)] as well as O 1\textsubscript{s} [in panel (b)] can be fitted by using the carbon reference spectra (C 1\textsubscript{s}) and the oxygen reference spectra (O 1\textsubscript{s}), respectively. (c) Shows the absolute O 1\textsubscript{s} XAS spectra intensities of the experimental samples. In graphene and HOPG samples the oxygen content is so low that only the experimental C 1\textsubscript{s} spectra are fitted computationally. Experimental spectra are taken from Ref. [14] and the fitting data can be found from [6].
known that only fractions of oxygen is present in the experimental sample, it makes sense to discard C 1s fingerprint spectra, that correspond to carbon bonded to oxygen during fitting. Instead, we included fingerprint spectra of certain defects – SV and DV, depicted in Figure 18 – in order to get a better idea of how defective or not the experimental sample is. Figure 19 (c) shows in more detail how the amount of oxygen can be evaluated from absolute intensities of experimental spectra [6, 14]. Fitted C 1s spectra are depicted in Figure 19 (a) and Figure 19 (b) describes the O 1s fit of the only experimental sample that has significant amount of oxygen to be fitted. Again, as is mentioned in previous section, this work confirms that C 1s fingerprint spectra are more important for identifying different forms of carbon in the sample, whereas O 1s fingerprint spectra offer more information when different functional groups are being studied.

VI. SUMMARY

In this thesis we present several examples that are powered by DFT- or/and ML-based solutions, and introduce also other computational methods, that are employed in order to understand and to complement experimental science [1–6].

The big picture of how experimental and computational work can be united is given in the first publication [1]. This book chapter describes different areas that are being studied in our multidisciplinary group and provides motivation for the subsequent computational work.

In the second publication [2] we are exploring what happens during electrochemical measurements. The purpose of this work is to build a bridge between computational and experimental work. It is a case study of dopamine, ascorbic acid and dopaminequinone adsorption on carbonaceous electrode surface, while the potential of the surface is being changed. We present a novel indirect method for controlling the potential. It is done by adding dopant atoms to the computational model. These dopants either increase or decrease the electron density of the surface, and thus, the Fermi level as well as the work function is shifted. Then we observe how the electron density between the surface and the molecules changes and how this affects the adsorption energies of the molecules. Since adsorption is highly surface chemistry dependent and adsorption is considered to be an important state during the oxidation/reduction process of these molecules, adsorption energy calculations have also
been employed in Refs. [3, 8].

Carbonaceous materials are traditionally categorized according to how \( sp, \ sp^2 \) or \( sp^3 \) rich they are [75]. However, in the case of amorphous carbon the complexity of the structure creates a demand for a closer investigation in order to understand the behavior of this material better. In the second publication [2], we overcome this issue by comparing \( \alpha-C \) with graphene (\( sp^2 \) rich) and diamond (\( sp^3 \) rich). In the third publication [3] the classification is taken further, but utilizing machine learning featurizing technique SOAP [61, 62]. Furthermore, this is done with a large database of structures, that are created with ML-based potential GAP [12, 45, 59]. Having enough data – which would be impossible to create with DFT only – allows us to study reactivity of \( \alpha-C \) in more detail at statistically meaningful level.

After having this new approach for classifying amorphous and other disordered structures, we decided to apply it in characterization. In the fourth publication, Part I [4], we are computing X-ray spectroscopy (XAS and XPS) fingerprints of clustered \( \alpha-C \) structures – both pristine and functionalized. This study helps us to see where the profile of experimental \( \alpha-C \) spectrum is arising from. In other words, we can estimate which motifs are the most predominant ones in the experimental sample. The X-ray spectroscopy interpretation method, including DFT- and ML-based structures, SOAP clustering, MDS and finally fitting the experimental spectra, are explained in more detail, in addition to this publication, in Refs. [5, 6], which are also included in this thesis and will be discussed next.

In the previous publication, we investigate the spectra qualitatively. In the fifth publication, Part II [5], we move on to making quantitative analysis and observations. This is done by introducing a fitting method: linear combination of selected reference spectra are matched with the experimental spectra. Weights in the linear combination show how much each reference spectra is contributing to the results that corresponds the best with the experimental spectrum. However, because of the overlapping features in the spectra, that are classified in the previous publication, we cannot use them in fitting. After all, while fitting, we are using the whole shape of the spectrum, not just peak positions. In addition, the algorithm does not give stable results with too many spectra. In case of carbon, there is so much data that we need to reduce complexity in order to have satisfying results. This is done by re-clustering the carbon sites by utilizing not only geometrical and chemical input, but we also add Delta Kohn-Sham values to find connections between the features in
the spectra. With a reasonable – not too large – set of reference spectra we can then fit experimental C 1s spectrum to see how carbon is bonded in the sample. With oxygen containing functional groups, clustering is more simple, and it agrees with chemical intuition, i.e., even more information about functionalization can be obtained by fitting experimental O 1s spectrum of the same sample.

Finally, in the sixth publication of this work [6], we combine the best parts of aforementioned publications [4, 5]. Once again computational work is being used in order to understand experimental results. Relevant applications of the method can also be found from Refs. [11, 13, 14]. In this publication [6], we turn our focus from a-C to graphene, graphene oxide and $sp^2$ rich carbon in general. These forms of carbon can be just as disoriented and complex as a-C, so it makes sense to utilize similar methods in order to interpret experimental results. However, especially the structure of graphene oxide has been under debate for a long time; it depends a lot of the precursor material as well as the oxidation or reduction protocol that have been chosen during sample manufacturing. Computationally, we have a perfect control over what our samples contain, and this way we can offer a systematic approach for examining experimental X-ray spectra qualitatively (see Fig. 17). For quantitative analysis, we employ clustering again. All the carbon sites as well as the oxygen sites were classified in order to provide fingerprints for computational fitting of graphene, graphene oxide and also multilayered functionalized graphite-based materials (see Fig. 19). That is, because clustering removes the sites that are being probed from the original structure, and thus, they can be treated individually. In other words, these fingerprints can represent any similar site in any kind of graphene- or graphite-based structure. In general, computational fitting of spectroscopy has an enormous potential in the future.

As a conclusion, we can state that we have successfully built new connections between computational and experimental science, from which the multidisciplinary community can benefit. When calculation and experiments support and complement each other, we can truly say that we have gained a better understanding of the phenomena that surround us.


[41] The Atomic Simulation Environment (ASE) is a set of tools and Python modules for setting up, manipulating, running, visualizing and analyzing atomistic simulations, https://wiki.fysik.dtu.dk/ase/, (accessed 06.05.2021).


[48] A. Tkatchenko and M. Scheffler, “Accurate molecular van der waals interactions from ground-


[60] Lecture material, course created by Caro M. A., (accessed 08.10.2021).


[73] A. N. Patel, S. Tan, T. S. Miller, J. V. Macpherson, and P. R. Unwin, “Comparison and


58


