

## OPEN ACCESS

IOP Publishing

Journal of Physics: Condensed Matter

J. Phys.: Condens. Matter **33** (2021) 154005 (20pp)<https://doi.org/10.1088/1361-648X/abd00>

# The nuts and bolts of core-hole constrained *ab initio* simulation for *K*-shell x-ray photoemission and absorption spectra

Benedikt P Klein<sup>1,2</sup> , Samuel J Hall<sup>1,3</sup>  and Reinhard J Maurer<sup>1,\*</sup> <sup>1</sup> Department of Chemistry, University of Warwick, Gibbet Hill Rd, Coventry, CV4 7AL, United Kingdom<sup>2</sup> Diamond Light Source, Harwell Science and Innovation Campus, Didcot, OX11 0DE, United Kingdom<sup>3</sup> MAS CDT, Senate House, University of Warwick, Gibbet Hill Rd, Coventry, CV4 7AL, United KingdomE-mail: [r.maurer@warwick.ac.uk](mailto:r.maurer@warwick.ac.uk)

Received 20 October 2020, revised 21 December 2020

Accepted for publication 22 January 2021

Published 26 February 2021



CrossMark

## Abstract

X-ray photoemission (XPS) and near edge x-ray absorption fine structure (NEXAFS) spectroscopy play an important role in investigating the structure and electronic structure of materials and surfaces. *Ab initio* simulations provide crucial support for the interpretation of complex spectra containing overlapping signatures. Approximate core-hole simulation methods based on density functional theory (DFT) such as the delta-self-consistent-field ( $\Delta$ SCF) method or the transition potential (TP) method are widely used to predict *K*-shell XPS and NEXAFS signatures of organic molecules, inorganic materials and metal–organic interfaces at reliable accuracy and affordable computational cost. We present the numerical and technical details of our variants of the  $\Delta$ SCF and TP method (coined  $\Delta$ IP-TP) to simulate XPS and NEXAFS transitions. Using exemplary molecules in gas-phase, in bulk crystals, and at metal–organic interfaces, we systematically assess how practical simulation choices affect the stability and accuracy of simulations. These include the choice of exchange–correlation functional, basis set, the method of core-hole localization, and the use of periodic boundary conditions (PBC). We particularly focus on the choice of aperiodic or periodic description of systems and how spurious charge effects in periodic calculations affect the simulation outcomes. For the benefit of practitioners in the field, we discuss sensible default choices, limitations of the methods, and future prospects.

Keywords: x-ray photoemission, density functional theory, NEXAFS, delta self consistent field, x-ray absorption, transition potential, metal–organic interfaces

 Supplementary material for this article is available [online](#)


(Some figures may appear in colour only in the online journal)

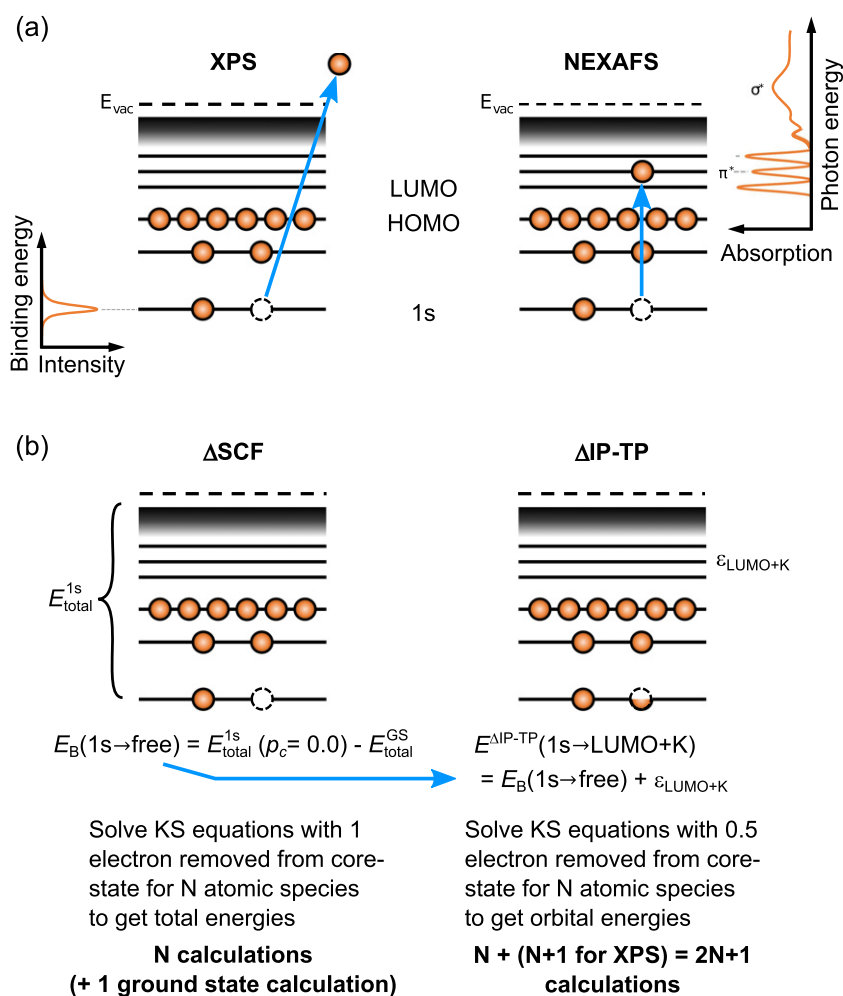
## 1. Introduction

X-ray photoelectron spectroscopy (XPS) and x-ray absorption spectroscopy (XAS, often called near edge x-ray

absorption fine structure, NEXAFS, or x-ray absorption near edge structure, XANES) play an important role in the characterization of materials and surfaces [1–3]. XPS is routinely used to provide information on the chemical composition and the oxidation state of elements in various systems, including oxides [4], two-dimensional materials such as graphene [5, 6], and metal–organic interfaces [7–9]. NEXAFS and, particularly, angular- or polarization-dependent NEXAFS can provide insight into the electronic structure and orientation of

\* Author to whom any correspondence should be addressed.

 Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



**Figure 1.** (a) Principles of x-ray photoelectron spectroscopy (XPS, left) and near-edge x-ray absorption fine-structure spectroscopy (NEXAFS, right). XPS detects photoelectrons emitted from core-levels after absorption of x-ray radiation. This provides information about the chemical environment of the core-states for each atom. NEXAFS probes secondary processes that are associated with the excitation of core-electrons into unoccupied valence states. This provides information about the unoccupied valence states of the system. (b) In this paper, we calculate XPS spectra with the delta-self-consistent-field ( $\Delta$ SCF) method (left) and NEXAFS spectra with the ionization-potential-corrected transition potential ( $\Delta$ IP-TP) method.

thin films [10], surfaces and adsorbates [11], and even liquids [12]. Both methods are particularly suitable for surface characterization: XPS is inherently surface sensitive due to the small escape depth of the photoelectrons from condensed matter and NEXAFS can make use of the same effect by choosing a suitable electron yield detection [2, 13].

In core-level spectroscopy (see figure 1(a)), the sample is exposed to an x-ray beam, leading either to the emission of a photoelectron from a core-level (XPS) or the excitation of the core-electron to unoccupied electronic states (NEXAFS). For XPS, a fixed photon energy is used and the spectrum is acquired by scanning over the kinetic energy of the photoelectrons. For NEXAFS, the photon energy is varied and the energy-dependent absorption is measured. Ideally, the measured spectrum exhibits sharp signatures which can be directly assigned to originate from individual core-levels (XPS, figure 1(a) left) or from transitions between core-levels and unoccupied states (NEXAFS, figure 1(a) right). Changes in the chemical composition, molecular conformation

or even orientation, or electronic structure of a sample will be reflected in the position, shape, and intensity of spectral signatures. Therefore both methods provide direct evidence of such changes.

However, the rich structure and overlapping features present in XPS and NEXAFS spectra often pose a challenge to the unambiguous assignment of all peaks. Here, the theoretical analysis and computational simulation of the electronic structure and transitions can contribute and produce additional insights. The first-principles-based simulation of core-level spectra is widely employed for this purpose. Established methods include the use of time-dependent density functional theory (DFT) [14, 15], many body perturbation theory methods such as GW (for XPS) and Bethe–Salpeter equation calculations (for neutral excitations) [16, 17], as well as multiconfigurational wave function methods and coupled cluster (CC) calculations [18–20]. In addition, more cost-effective DFT methods that approximate the effects of an excited core-hole within a ground-state DFT formalism have been proposed.

The latter methods include  $\Delta$ SCF DFT [21–23] to simulate XPS spectra and the transition potential (TP) method [24] to simulate NEXAFS spectra. Many variations of these methods have been proposed in the past, [25–27] with varying uptake in the community. These approximate methods are particularly suitable to study *K*-shell spectroscopy, i.e. the emission or excitation of 1s electrons, because higher lying states feature additional challenges such as spin–orbit coupling and delocalization over several atoms [28, 29].

Approximate core-hole excitation methods to simulate core-level spectroscopy differ in how many electrons are excited, how the core-hole constraint is imposed and localized [30], which underlying basis set is used, e.g. if all-electron local atomic basis functions or pseudopotential-plane-wave approaches are used [27, 31, 32], and how relativistic effects are treated. The empirical wisdom behind simulating reliable and accurate core-hole spectra is often very software- and application-specific and is rarely the focus of studies, but rather relegated to supplemental information, even as it is crucial for the achieved performance.

In this paper, we present our approach to approximate core-hole simulations and discuss methodological and technical intricacies. We analyze the dependence of simulated spectra with respect to all relevant simulation parameters to establish what considerations are required to produce reliable and converged XPS and NEXAFS spectra that are independent of numerical choices. We draw from our experience in simulating XPS and NEXAFS signatures of organic molecules adsorbed at metal interfaces [33–37] and highlight the challenges associated with simulating 1s XPS and *K*-edge NEXAFS spectra for organic molecules in gas-phase, in bulk crystals, and adsorbed at surfaces. After a brief introduction into the methodology and practical steps of XPS and NEXAFS calculations in section 2, we analyse how spectra depend on the choice of core-hole localization method (section 3.1), exchange–correlation functional (section 3.2), basis set (section 3.3), and model representation (cluster or periodic) before highlighting the challenges specific to bulk and surface calculations in section 3.5. On the latter topic, we specifically discuss the role of charging artefacts in periodic calculations. We further showcase the strength of analysing simulated NEXAFS spectra in terms of their angular and polarization dependence and by decomposing them into atomic and molecular orbital contributions to disentangle chemical shifts from final state screening effects (section 3.4).

## 2. Methods

### 2.1. *Ab initio* simulation of x-ray photoemission spectroscopy (XPS)

*Principle of XPS:* since the first observation of the photoelectric effect by Hertz and Hallwachs, photoemission spectroscopy has become one of the most important techniques for the chemical analysis of molecules and materials [38]. Upon exposure to x-ray radiation with frequency  $\nu$ , photoelectrons with kinetic energy  $E_{\text{kin}}$  and momentum  $\mathbf{k}$  can be detected

following the relationship:

$$E_{\text{kin},\mathbf{k}} = h\nu - E_{\text{B}} - \Phi. \quad (2.1)$$

In (2.1),  $E_{\text{B}}$  is the binding energy of the electron and  $\Phi$  is the work function of the material. A typical photoelectron spectrum plots the intensity of an electron detector signal against the binding energy (from highest to lowest binding energy). Depending on the used photon energy, the intensity  $I$  of the photoelectrons with kinetic energy  $E_{\text{kin},\mathbf{k}}$  is proportional to the photoelectron cross section  $\sigma_{\mathbf{k}}$  as follows [39]:

$$I_{\mathbf{k}}(h\nu) \propto n \lambda_{\text{in}} A J_{h\nu} \sigma_{\mathbf{k}}(h\nu), \quad (2.2)$$

where  $\lambda_{\text{in}}$  is the inelastic mean free path (IMFP),  $n$  is the number of atoms of this species per volume,  $A$  is the surface area covered by the x-ray spot, and  $J_{h\nu}$  is the flux density of incoming photons of frequency  $\nu$ . With few exceptions [40, 41], the IMFP is mostly neglected in simulations as, within the typical kinetic energy range (a few 100 electronvolts up to 1 keV), the IMFP of photoelectrons is of the order of 1 nm, which amounts to few atomic layers from the surface [13]. The cross section  $\sigma_{\mathbf{k}}$  itself can be related to the spectral function  $A_{ii}(E_{\text{kin},\mathbf{k}} - h\nu)$  of photoemission of an electron with momentum  $\mathbf{k}$  from core-state  $i$  via Fermi’s golden rule and the sudden approximation [17, 42]:

$$\sigma_{\mathbf{k}}(h\nu) = \sum_i |\langle \psi_{\mathbf{k}} | \mathbf{p} \cdot \mathbf{A}(\mathbf{r}) | \psi_i \rangle|^2 A_{ii}(E_{\text{kin},\mathbf{k}} - h\nu), \quad (2.3)$$

where  $\mathbf{A}(\mathbf{r})$  is the vector potential of the incident electromagnetic field and  $\mathbf{p}$  is the momentum operator.

Within many-body theory, the spectral function arises from the single particle Green’s function that describes the removal or addition of particles to the system [17] and describes the frequency-dependent probability of photoemission and the associated lineshape. The positions of the poles in the Green’s function correspond to the electron addition/removal energies with respect to the Fermi energy, i.e. the binding energies of the associated quasi-particles. This takes account of all relevant interaction processes, including collective screening, electron–electron scattering and relaxation effects due to the introduction of the core-hole, all of which arise from many-body correlation. For example on metals, XPS spectra typically show a characteristic asymmetry, which stems from the interaction of the photoelectron with electron–hole-pair excitations close to the Fermi level [43]. Full many-body calculations can provide very accurate XPS spectra that fully account for many spectral features [16, 17]. Such calculations are, however, associated with high computational cost and are all but unfeasible for systems larger than several tens of atoms, and therefore remain unpractical for organic crystals or metal–organic interfaces.

In practice, the main features of XPS spectra are often simulated by effective single particle approaches. Instead of evaluating the cross section from the full spectral function, it is instead approximated by a set of single particle transitions between effective Kohn–Sham states within DFT. By assuming independent electrons [ $A_{ij}(h\nu) = \delta_{ij} \delta(h\nu - \varepsilon_i)$ ], we

can simplify (2.3) to the well known Fermi's golden rule expression [44]:

$$\sigma_{\mathbf{k}}(h\nu) = \sum_i |\langle \psi_{\mathbf{k}} | \mathbf{p} \cdot \mathbf{A}(\mathbf{r}) | \psi_i \rangle|^2 \delta(E_{\text{kin},\mathbf{k}} - h\nu - \varepsilon_i). \quad (2.4)$$

If no selection rules are to be considered, for example if no angular dependence is required (as would be in angular resolved photoemission spectroscopy [45, 46]), the matrix elements in (2.4) are often neglected as all final states are assumed to be extended plane wave states. This amounts to associating each excitation that originates from an atom-centred core-state  $i$  with the same intensity, i.e. eliminating all but  $n$  and  $\sigma_{\mathbf{k}}(h\nu)$  from (2.2). The result of such a calculation is a spectrum with  $N$  infinitely narrow  $\delta$ -peaks where  $N$  is the number of atoms, each with the same intensity and located at the energetic position of the binding energy  $E_{\text{B}}$  of the respective atom's electrons.

This represents an effective single particle approach that neglects lifetime broadening due to many-body effects such as electron–electron correlation. These effects can be approximately accounted for by applying an empirical line broadening shape to the infinitely sharp binding energies. The line shape is in this case often approximated by assuming a Gaussian, Lorentzian, or convoluted Gaussian–Lorentzian (i.e. Voigt) function for each atomic peak. The summation of all peak functions leads to a spectral shape comparable to what is commonly observed in experiments. The Lorentzian broadening is justified with life-time effects of the excited states, while Gaussian broadening simulates experimental effects (e.g. analyser resolution). A simplified way to incorporate both kinds of broadening into the simulated spectra, which was chosen for this publication, is the use of a pseudo-Voigt function, which reduces the convolution needed for a proper Voigt function to a summation operation [47, 48]. When using such functions, the parameters for line-width (full width at half maximum, FWHM), and Gaussian–Lorentzian ratio of the broadening are chosen according to the expected or observed experimental resolution. Using these approximations to obtain intensities and line shapes, the problem of simulating XPS spectra is reduced to finding the binding energies  $E_{\text{B}}$  of the relevant electronic core-states.

*Ab initio calculation of core-level binding energies:* many different approaches exist to calculate core-level binding energies including CC [49], equation-of-motion CC (EOM-CC) [20], the GW method [16, 17], and real-time TD-DFT [50]. The simplest approach, however, is the use of Koopmans' theorem (2.5), which relates the ionization potential of an electronic state in Hartree–Fock theory with the negative of the relevant Hartree–Fock eigenstate [51]. Despite the fact that in DFT this relationship only holds for the highest occupied Kohn–Sham state [52, 53], Kohn–Sham (KS) energies of core-levels are often used to estimate the chemical shift contribution to the binding energy, i.e. the displacement of the core-level before removal of the electron (also termed initial state effect) [54–56].

$$E_{\text{B}}(i) = -\varepsilon_{i,\text{HF/KS}}. \quad (2.5)$$

In most cases, however, core-hole relaxation effects (or final state effects) cannot be neglected in the prediction of binding energies. The  $\Delta\text{SCF}$  DFT (or  $\Delta\text{SCF}$ ) method [21, 22, 57], which includes relaxation effects, is therefore the most common approach to simulate core-level binding energies [55, 58–61]. As shown in figure 1(b), left, the  $\Delta\text{SCF}$  method calculates binding energies as energy differences between two self-consistent KS DFT calculations, namely the ground-state calculation and the core-hole excited calculation, where the electron population  $p_c$  of the  $1s$  core-state has been constrained to remove one electron:

$$E_{\text{B}}(1s \rightarrow \text{free}) = E_{\text{total}}^{1s}(p_c = 0.0) - E_{\text{total}}^{\text{GS}}. \quad (2.6)$$

Other methods have been proposed that introduce core-holes equivalent to half an electron (so-called Slater transition state approach) [57], or two-thirds of an electron [62], although full removal of an electron has been shown to provide the best results [24]. Unfortunately, often a simple change of orbital occupation does not provide a sufficient constraint to localize the core-hole onto the atom-centred core-level [63, 64]. This is particularly true if the molecule contains symmetry-equivalent atoms with degenerate core-states. Many different approaches have been proposed in the past to overcome this, such as the maximum overlap method (MOM) [60], using a localized hole in a frozen core-state [25], shifting results between symmetric MOs [65], and localization via initialization with a density calculated for an increased atomic core-charge [27]. Due to those difficulties, we will discuss the issue of core-hole localization in detail in section 3.1.

An additional problem for calculations of systems with periodic boundary conditions (PBC) is the treatment of the charge introduced by the core-hole calculations. The unit cell of every periodic calculation has to be net-neutral to avoid failure of the Ewald summation. Therefore, a uniform negative charge background is introduced into the calculation of the positively charged unit cell with the core-hole, which might lead to unpredictable changes in the calculated energies. This charging issue in periodic systems has previously been recognized as problematic [66, 67]. The effect of the background charge can in principle be reduced by increasing the unit cell size. Unfortunately, the necessary unit cell sizes to reduce the introduced shift to a negligible value are commonly thought to be too large for practical uses. However, while we show in section 3.3 that the absolute energies are not easily converged, we also found that the relative shifts are quite robust with varying cell size.

## 2.2. *Ab initio simulation of NEXAFS*

*Principles of NEXAFS:* excellent accounts of the measurement and first-principles simulation of x-ray absorption have been given previously by Hähner [68], Frati *et al* [1], Tanaka and Mizoguchi [69–71], Norman and Dreuw [72], and Guda *et al* [73]. We do not attempt to cover the full breadth of literature on the subject here and will focus mostly on simulation with constraint-based DFT methods.

An x-ray absorption spectrum is experimentally recorded by varying the photon energy of incident electromagnetic radiation and measuring a detector signal, which is proportional to the absorption of x-rays by the sample, and therefore also proportional to the photon energy dependent cross section  $\sigma(h\nu)$ . This x-ray absorption cross section originating from excitation of core-state  $i$  as given by Fermi's golden rule in the single particle and dipole approximation is [2]:

$$\sigma_i(h\nu) \propto \sum_f |\langle \psi_f | \mathbf{p} \cdot \mathbf{A}(\mathbf{r}) | \psi_i \rangle|^2 \delta(h\nu + E_i - E_f), \quad (2.7)$$

where  $\psi_f$  and  $\psi_i$  correspond to the wave functions of the final valence and the initial core electronic states, respectively.  $\mathbf{p}$  is the dipole operator and  $\mathbf{A}(\mathbf{r})$  is the vector potential of the incident electromagnetic field. This can be further simplified by assuming a classical polarization wave with polarization along a unit vector  $\mathbf{e}$ , leading to  $\mathbf{A}(\mathbf{r}) = \mathbf{e}A_0 e^{i\mathbf{k}\cdot\mathbf{r}}$  [74]. For local core-excitations, the dipole approximation is valid ( $e^{i\mathbf{k}\cdot\mathbf{r}} = 1$ ) and yields:

$$\sigma_i(h\nu) \propto \sum_f |\langle \psi_f | \mathbf{e} \cdot \mathbf{p} | \psi_i \rangle|^2 \delta(h\nu - \underbrace{(E_f - E_i)}_{\Delta E_{fi}}). \quad (2.8)$$

Therefore, the intensity of a NEXAFS spectrum is defined by the nature and overlap of initial and final states. The  $\delta$ -function ensures energy conservation between the energy of the incident light and the transition energy  $\Delta E_{fi}$ . To obtain the spectral shape, the discrete  $\delta$ -functions of each excitation are typically replaced by a model of the finite line shape to account for the life-time of final states and other broadening effects, in analogy to the XPS line shapes (*vide infra*).

*Simulation of NEXAFS spectra:* a large variety of methods to simulate x-ray absorption spectra have been proposed including linear response based on CASSCF calculations [18], TD-DFT [15, 75], and Bethe–Salpeter equation calculations [76–79]. Here we discuss NEXAFS in the context of molecular bulk and metal–organic surface systems, where the choice of method is much more limited due to the intrinsic size of and computational cost associated with such systems.

A variety of more approximate DFT-based methods to calculate NEXAFS transition energies  $\Delta E_{fi}$  have sprung from the  $\Delta$ SCF method. These include methods that explicitly simulate each transition between the core-hole and all unoccupied states, such as the  $\Delta$ SCF method and the Slater–Janak transition-state method [57], where one or half an electron are excited from the core-level  $p_c$  to the relevant valence state  $p_v$ , respectively. The transition energy is given by the difference in total energy of the core-excited calculation ( $p_c = 0.0, p_v = 1.0$ ) and the DFT ground state ( $p_c = 1.0, p_v = 0.0$ ):

$$\Delta E^{\Delta\text{SCF}} = E(p_c = 0.0, p_v = 1.0) - E(p_c = 1.0, p_v = 0.0). \quad (2.9)$$

In this case,  $M$  calculations are required, where  $M$  is the number of valence states  $v$  that will be considered. This procedure needs to be repeated for all relevant core-levels  $N$ , on which the core-hole is to be localized. For  $K$ -shell excitations,  $N$  equals the number of symmetry-inequivalent atoms. The total number of DFT calculations for a single NEXAFS spectrum is

therefore  $(N \cdot M) + 1$  (including the ground-state). Using this approach, the sheer amount of calculations can be prohibitive if the system contains a large number of atoms.

The transition potential (TP) method [80, 81], is an approximation based on Slater's transition state approach [57], which circumvents such a large number of calculations. Rather than calculating each excitation explicitly as a difference between the total energy of an excited state and the total ground-state DFT energy, the transition energy is defined as the difference between the KS energies of valence ( $\varepsilon_v$ ) and core-states ( $\varepsilon_c$ ). These KS energies are calculated for a system, where half an electron is removed from the core-state to model the effect of the core-hole on the transition:

$$\Delta E^{\text{TP}} = \varepsilon_v(p_c = 0.5, p_v = 0.0) - \varepsilon_c(p_c = 0.5, p_v = 0.0). \quad (2.10)$$

A single TP DFT calculation provides all KS states and therefore yields all NEXAFS transitions and the number of calculations for a spectrum is reduced to the number of atomic species  $N$ . A variant of the TP method has been proposed that introduces a full core-hole (FCH), instead of a half core-hole [82, 83]:

$$\Delta E^{\text{FCH}} = \varepsilon_v(p_c = 0.0, p_v = 0.0) - \varepsilon_c(p_c = 0.0, p_v = 0.0). \quad (2.11)$$

Note that in both cases the (partial) electron removed from the core-state is not placed into the valence states and the system is not net neutral. The lack of charge neutrality, as is also the case for  $\Delta$ SCF calculations of XPS binding energies, can provide a challenge for simulations of bulk and surface systems using PBC. While charge neutrality in the unit cell can be restored by the introduction of a homogeneous charge background [82], this charge background itself can lead to unphysical artefacts in the calculations [67]. Several solutions to this problem have been proposed [84, 85], for example by placing the removed charge into the lowest unoccupied molecular orbital (LUMO) of the system as done in the XCH method which introduces a full core-hole [26, 30]:

$$\Delta E^{\text{XCH}} = \varepsilon_v(p_c = 0.0, p_{\text{LUMO}} = 1.0) - \varepsilon_c(p_c = 0.0, p_{\text{LUMO}} = 1.0). \quad (2.12)$$

A comprehensive summary of different approximate core-hole constraint methods has recently been given by Michelitsch and Reuter, where the authors also introduced other previously not considered variants of the TP method [30]. For a small set of organic molecules in gas-phase, the authors compare these different approaches and find that carbon and nitrogen  $K$ -edge NEXAFS spectra (intensities and excitation energies) are best represented by the XCH and XTP methods. The latter is a variation of TP where the removed half core-hole is neutralized with half an electron placed into the LUMO:

$$\Delta E^{\text{XTP}} = \varepsilon_v(p_c = 0.5, p_{\text{LUMO}} = 0.5) - \varepsilon_c(p_c = 0.5, p_{\text{LUMO}} = 0.5). \quad (2.13)$$

While the TP and XTP methods seem to provide reliable and robust predictions of the relative peak positions and intensities in NEXAFS spectra [30], the XCH and  $\Delta$ SCF methods,

which both introduce a full core-hole, seem to provide a better description of absolute energies [24]. Whereas the half core-hole constraint in TP and XTP appears to accurately account for core-hole relaxation effects in valence states, the chemical shift and core-hole screening is more accurately represented with a full core-hole in  $\Delta$ SCF and XCH. This conundrum has previously already been recognized by Triguero *et al* [24]. They proposed a hybrid approach, where the TP (half core-hole) is corrected by the ionization potential of the core state calculated with a full core-hole. Following Triguero *et al* [24], we call this ionization-potential-corrected TP method  $\Delta$ IP-TP to distinguish it from the original TP method [80, 81]. The energy of a transition between a core-level and an unoccupied valence state is calculated in  $\Delta$ IP-TP by combining the results of a  $\Delta$ SCF and a TP calculation (see also figure 1(b), right):

$$\Delta E^{\Delta\text{IP-TP}} = \varepsilon_v(p_c = 0.5, p_v = 0.0) + E_B(1s \rightarrow \text{free}). \quad (2.14)$$

In (2.14),  $\varepsilon_v$  are the KS energies of the unoccupied states from the half core-hole TP calculation, whereas  $E_B(1s \rightarrow \text{free})$  is the total binding energy of the core-level from a full core-hole  $\Delta$ SCF calculation as laid out in (2.6). This approach retains the accurate description of relative transition energies for each atomic center, but additionally improves not only the description of the chemical shift of each atom but also the absolute transition energies. We have employed such a hybrid approach in all our previous work on metal–organic interfaces and organic molecules, where we have achieved good agreement with experimental data [34–37]. In principle this method can also be employed on the basis of XTP instead of TP to calculate the KS-energies of the unoccupied states, as we will show below.

TP core-hole calculations are regularly performed in codes that use all-electron atomic orbital basis sets [30] or codes that use plane waves and pseudopotentials [31]. In the former case, the binding energy from  $\Delta$ SCF can yield highly accurate absolute energies if well-designed basis sets with sufficiently flexible core-functions are used [27, 32, 60, 86]. In the case of pseudopotential plane-wave codes, in order to calculate dipole matrix elements, the full all-electron wave function needs to be reconstructed using the projector-augmented-wave method [31, 70]. The frozen core adds to already existing errors that stem from the self-interaction error of approximate exchange–correlation functionals and the resulting absolute binding energies are typically not close to experiment. However, the application of a rigid global shift can still allow a fruitful comparison with the experiment.

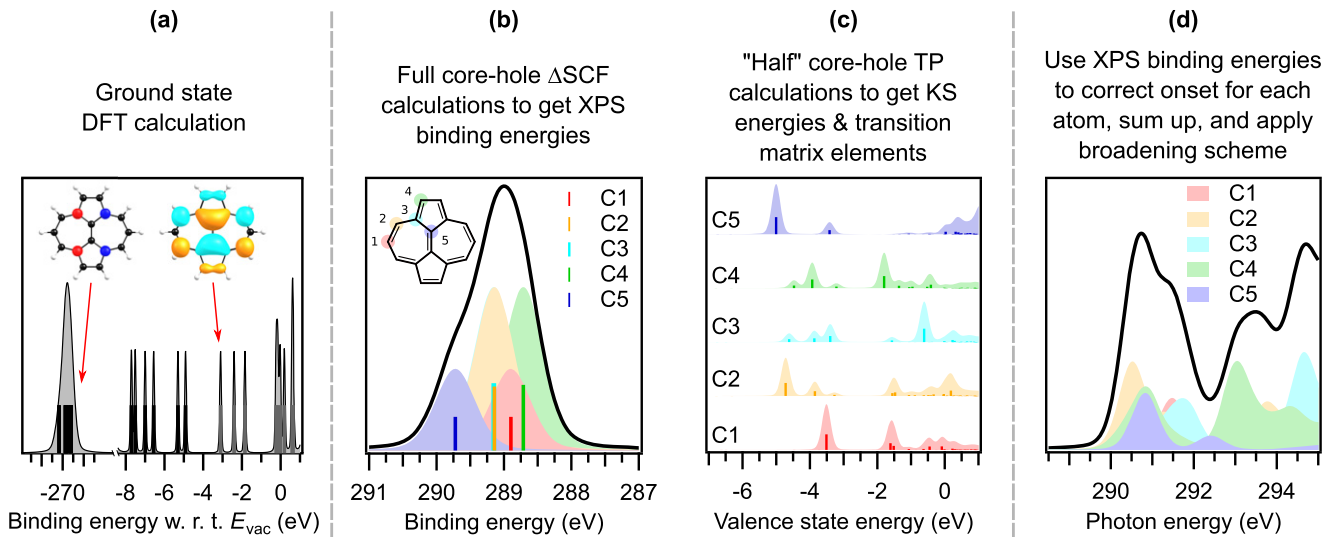
Figure 2 summarizes the practical steps involved in performing a  $\Delta$ IP-TP (or  $\Delta$ IP-XTP) simulation of a *K*-edge NEXAFS spectrum: a ground-state DFT calculation (figure 2(a)) is followed by  $N$   $\Delta$ SCF calculations (figure 2(b)) with a full core-hole to determine the binding energy for each inequivalent core-level [ $E_B(1s \rightarrow \text{free})$  in (2.14)] and  $N$  TP calculations (figure 2(c)) with a half core-hole to determine the KS-energies of the unoccupied states [ $\varepsilon_v$  in (2.14)]. These calculations each have to be performed  $N$  times, because the full/half core-hole has to be introduced into each of the  $N$  symmetry-inequivalent 1s states in the system. Each TP

calculation first provides a self-consistently converged density and then works to accurately converge the unoccupied valence states. This approach is necessary to achieve the same convergence accuracy for the unoccupied states as for the occupied ones. Therefore, care needs to be taken that convergence thresholds are set appropriately for the former as their KS energies directly control the transition energies and the respective transition dipole matrix elements with the core-state.

Finally, the  $\Delta$ SCF binding energies are added to the TP eigenvalue spectra to align the contributions originating from different atomic species onto a single energy axis and the spectra are summed up (figure 2(d)). The result is the final *ab initio* predicted spectrum. This spectrum can be plotted with the contributions from each core-state (as shown) or further analysed by projecting to the contributions from valence states or molecular orbitals (see section 3.4).

Both XPS and NEXAFS spectra are broadened to simulate typical experimental resolutions and ease the comparison to measured data. We chose to use a bare-bone scheme employing a minimal number of parameters and a mathematically simple implementation. While physically more meaningful peak functions exist [2], their proper application would be much more complex. The scheme used in this publication is still able to produce good agreement with experimental spectra and is superior for most practical purposes due to its simplicity of use. For the XPS spectrum, a fixed peak width and Gaussian/Lorentzian (*G/L*) ratio can normally be chosen for all peaks stemming from the same core-level. The spectrum is yielded as a summation over the pseudo-Voigt-functions [47, 48] for each symmetry equivalent atom weighted by their abundance in the system. While the broadening function can also be chosen to be the same for the NEXAFS transitions of all atomic species, it has to change with photon energy in the NEXAFS spectrum. This approach is necessary to account for the difference in lifetime of states below and above the ionization threshold. Therefore, two different broadening schemes are employed for low and high photon energies. In the high photon energy regime, the width is larger and the *G/L* ratio is turned more Lorentzian, while the reverse is true for the low photon energy regime. Both regimes are linked by a linear gradient for width and *G/L* ratio, which crosses the ionization threshold, in line with common approaches in literature [87, 88]. Systematic errors in the method and the self-interaction errors in the exchange–correlation functional make it furthermore necessary to globally shift the simulated spectra to align with the experimental energy axis both in the XPS and NEXAFS spectrum.

*Angular- and polarization-resolved NEXAFS simulations:* NEXAFS can not only probe the character of electronic states, but also their orientation with respect to the incident light polarization [89]. As can be seen in (2.7), the intensity depends on the orientational alignment of the direction of polarization of incident light  $\mathbf{e}$  with respect to the vectorial change in dipole moment due to the electronic transition  $\mu_{fi}$ . Assuming linear polarized light, we can simplify the matrix elements in (2.7)



**Figure 2.** Computational workflow to generate a  $K$ -shell NEXAFS spectrum with the  $\Delta$ IP-TP approach. (a) Ground state DFT calculation. (b)  $N$  full core-hole  $\Delta$ SCF calculations to establish the chemical shifts of each atomic species. These calculations also provide the basis for the XPS spectrum. (c)  $N$  half core-hole TP calculations to obtain the absorption spectra for each species. (d) Post-processing: each atomic species is shifted according to its  $\Delta$ SCF binding energy and finally all atomic contributions are summed up. Additionally a suitable broadening scheme is included. The inset in (a) shows two exemplary core and valence states. The inset in (b) contains the structural formula of our example molecule azupyrene and the peak assignment for the inequivalent carbon atoms.

as follows:

$$\mu_{fi} = |\langle \psi_f | \mathbf{e} \cdot \mathbf{p} | \psi_i \rangle|^2 \approx |\mathbf{e} \cdot \underbrace{\langle \psi_f | \mathbf{p} | \psi_i \rangle}_{\mathbf{D}_{fi}}|^2 = \left( \sum_{r=x,y,z} e_r \cdot D_{fi,r} \right)^2. \quad (2.15)$$

Herein,  $\mathbf{D}_{fi}$  are the dipole matrix elements between initial and final states  $i$  and  $f$ , respectively. Therefore, the intensity is largest if the direction of the electric field vector  $\mathbf{e}$  and the transition dipole vector are parallel. In case of a spherical  $1s$  core-state, this means that the intensity is largest if the light polarization aligns with the electronic dipole of the final electronic state, see figure 3(a).

For some systems, all orientational dependence of the NEXAFS transitions is likely to be averaged out in the experimental data due to random orientation within the sample, for example in the gas phase, in liquids, or in amorphous or disordered polycrystalline solids. If this is the case for the experimental data we want to compare our simulations to, the orientational dependence can also be neglected in the calculations:

$$\mu_{fi}^{\text{total}} = D_{fi,x}^2 + D_{fi,y}^2 + D_{fi,z}^2. \quad (2.16)$$

For molecules adsorbed on metal surfaces, however, the polarization dependence of the NEXAFS signal (the so-called dichroism) is often present in the experiment and quite useful, because it enables the extraction of structural information about the studied system [10]. To describe the polarization dependence, we can define the polarization direction  $\mathbf{e} = (e_x, e_y, e_z)$  of the incident light in terms of polar  $\vartheta$  and azimuthal  $\varphi$  angles with respect to the surface (see figure 3(b)):

$$\mathbf{e} = (e_x, e_y, e_z) = (\sin \vartheta \cdot \cos \varphi, \sin \vartheta \cdot \sin \varphi, \cos \vartheta). \quad (2.17)$$

The resulting intensity for a transition between state  $i$  and state  $f$  excited by polarized light defined by the angles  $\vartheta$  and  $\varphi$  would be:

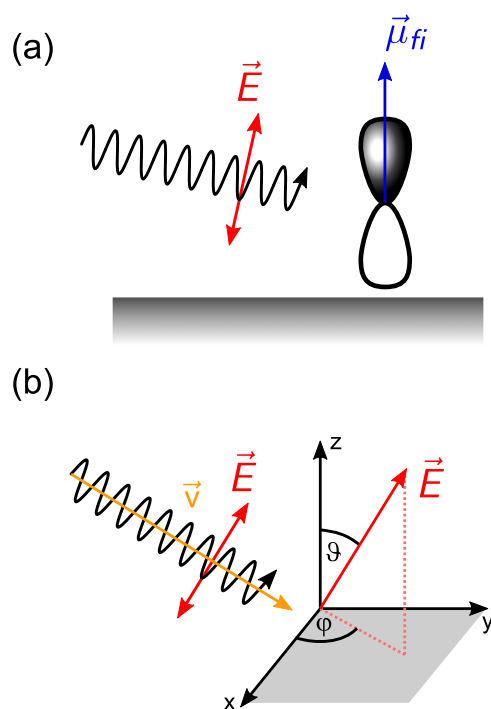
$$\mu_{fi}^{\text{p}}(\vartheta, \varphi) = (D_{fi,x} \cdot e_x + D_{fi,y} \cdot e_y + D_{fi,z} \cdot e_z)^2. \quad (2.18)$$

For some systems, the variation of the azimuthal orientation  $\varphi$  of the incident radiation relative to the substrate can also be used to extract structural information [90]. Most systems, however, and this includes most metal–organic interfaces, possess a random azimuthal orientation, or many rotational domains which cancel out the  $\varphi$ -dependence due to the symmetry of the underlying substrate [2]. For such systems, it is therefore best practice to average over the azimuthal  $\varphi$  dependence of the signal (in the  $x, y$  plane) for the corresponding simulations, yielding (2.19):

$$\mu_{fi}^{\text{p,av}}(\vartheta) = (D_{fi,x}^2 + D_{fi,y}^2) \cdot \sin^2 \vartheta + D_{fi,z}^2 \cdot \cos^2 \vartheta. \quad (2.19)$$

### 2.3. Computational details

The here presented DFT calculations have been performed with two different electronic structure software packages: the pseudopotential plane-wave code CASTEP [91] and the all-electron atomic-orbital code FHI-aims [92]. CASTEP was used to calculate NEXAFS spectra with the  $\Delta$ IP-TP method, as laid out in the method section and figure 2. This approach includes  $\Delta$ SCF total energy calculations and TP calculations for each atom, all of which were performed in CASTEP. The all-electron code FHI-aims was used to calculate XPS energies with the  $\Delta$ SCF method to obtain comparable values free from the potential errors brought by the forced periodicity and the use of pseudopotentials in CASTEP. The raw inputs and outputs of all electronic structure calculations have been deposited



**Figure 3.** Schematic representation of the angular dependence in NEXAFS. (a) The intensity of a NEXAFS transition depends on the relative orientation of transition dipole moment  $\mu_{fi}$  and electrical field vector  $\mathbf{E}$  ( $=A_0 \cdot \mathbf{e}$  in the dipole approximation). Note that  $\mu_{fi}$  is fixed relative to the axes by the orientation of the calculated system. For  $K$ -shell excitations, the  $\mu_{fi}$  orientation is also only determined by the final state. (b) The orientation of the electrical field vector can be described by two angles, the polar angle  $\vartheta$  and the azimuth  $\varphi$ ,  $\mathbf{v}$  is the propagation direction of the radiation.

in the NOMAD repository and are freely available online (DOI: <https://dx.doi.org/10.17172/NOMAD/2020.10.15-1>).

*All electron calculations in FHI-aims:* to localize the core-hole in the  $\Delta$ SCF calculation, several approaches were tested and are compared in section 3.1. These approaches employ the *force\_occupation\_basis* (FOB) and the *force\_occupation\_projector* (FOP) keywords implemented in FHI-aims by Matthias Gramzow (Fritz-Haber-Institute of the Max Planck Society, Berlin). Both keywords apply occupation constraints on KS eigenstates and use variants of the MOM to ensure that the constraint is satisfied [60]. With the FOP keyword, we define the occupation constraint directly onto a KS eigenstate. If eigenstates swap between consecutive SCF steps, the constraint moves with the eigenstate following the maximum overlap with the constrained state of the previous SCF step. With the FOB keyword, we define the core-hole occupation constraint in terms of a localized atomic orbital basis function. The occupation constraint will be enforced on the KS eigenstate that has the highest contribution from the selected core atomic orbital. During successive SCF steps, the constraint will follow the eigenstate with the highest contribution from the atomic orbital, which ensures that the population of the core-level associated with the right atom is constrained. This approach works well for localized core-states such as 1s levels of most elements. The more delocalized the target core-state is, the less reliable this approach will be as basis

functions on adjacent orbitals will complement each other. We find that the FOB approach is in most cases preferable as it correctly distinguishes between symmetry-equivalent atoms and therefore eliminates the necessity of an additional calculation to break the symmetry. We tested our approach against calculations with the method proposed by Kahk and Lischner [27], where the core-hole calculation using the FOP functionality is enabled by the use of an initial wave function with reduced symmetry. This symmetry-broken wave function is created from a single SCF step with slightly increased core-charge on the relevant atom, resulting in a symmetry breaking for the desired core-state. This approach requires more effort than the FOB method, but proved more resilient when working with hybrid functionals. Alternatively, the issues of the FOB approach with hybrid functionals can be solved by combining it with a Boys localization that is implemented in FHI-aims [93] and applied before the core-hole calculation [30, 94].

The basis set convergence of the atom-centered numerical orbitals native to FHI-aims was extensively tested, as described in detail in the supplementary material (<https://stacks.iop.org/JPCM/33/154005/mmedia>) (see tables S5–S8). To achieve a better description of the core-states, augmented near-core basis functions were introduced according to the literature [27]. These additional basis functions are s-type hydrogen-like functions with high effective nuclear charges. Taking the results of the convergence series into account, the ‘tier2’ basis with internal default tight settings for the numerical integration and the mentioned core-hole augmentation was chosen for the remaining calculations. The cluster calculations of the molecular crystal (in section 3.5) employed the ‘tight-tier2’ basis set only for the atoms in the central molecule itself, while the surrounding 16 molecules were supplied with a ‘light-tier2’ basis to reduce computational load. The same was true for the lower metal layers in the copper cluster calculations.

In section 3.2, different exchange correlation functionals were tested, these include PBE [95], PW91 [96], HSE06 [97], PBE0 [98], TPSS [99], SCAN [100], and xDH-PBE0 [101]. The calculations were performed with the FHI-aims versions 200408 and 200511, both of which were checked to give the same results. All structures were optimized in FHI-aims with the PBE functional, a ‘tight-tier2’ basis and a force threshold of  $0.01 \text{ eV } \text{\AA}^{-1}$ . The electronic convergence settings were set to  $1 \times 10^{-4} \text{ e } \text{\AA}^{-3}$  for the electron density,  $1 \times 10^{-2} \text{ eV}$  for the KS-eigenvalues and  $1 \times 10^{-6} \text{ eV}$  for the total energy. Relativistic effects were taken into account by employing the atomic ZORA functionality [92]. Higher order relativistic effects such as spin-orbit coupling could be neglected for the studied  $K$ -shell excitations and binding energies. XPS spectra were generated from the calculation results by representing each carbon 1s binding energy with a pseudo-Voigt function [47, 48]. The application of this broadening scheme allows a better comparison to experimental data and enables a realistic judgement of the magnitude of differences in the XPS binding energies. The parameters of the pseudo-Voigt function were chosen to reflect typical experimental resolutions achieved in surface science experiments (FWHM = 0.7 eV, 70%/30% Gaussian/Lorentzian ratio).



*Plane-wave calculations in CASTEP:* the plane-wave code CASTEP was used to perform both  $\Delta$ SCF and TP calculations within the  $\Delta$ IP-TP approach. All calculations were carried out in PBC using on-the-fly generated pseudopotentials. The stability of the calculation results with regards to the additional parameters introduced by PBC and pseudopotentials was extensively tested. In particular the choice of pseudopotentials, the energy cut-off for the pseudopotentials, and the vacuum box size around isolated molecules (necessary due to the PBC), was investigated in detail, as presented in section 3.3. Calculations were performed using either CASTEP version 18.1 or 19.11 [91], while maintaining the same pseudopotential definitions and with a total energy per atom convergence tolerance of at least  $1 \times 10^{-6}$  eV/atom. For the calculation of XPS binding energies, the  $\Delta$ SCF method was used. Here, for each carbon atom an excited core-hole calculation was performed, where the electron configuration of the respective pseudopotential was modified to  $[1s^1, 2s^2, 2p^3]$ , effectively localising the core-hole at this atom while moving the electron to the valence region. However, the new electron in the valence region is then removed by the introduction of a total charge of +1.0 e. Due to the PBC, this charge is compensated by a homogeneous background charge of -1.0 e, yielding a net-neutral unit cell. The core-level binding energies obtained for each carbon atom have furthermore to be corrected for the energy change of the core-electrons contained in the modified pseudopotential [70]. NEXAFS simulations were carried out using the ELNES module in CASTEP [70], the half core-hole was included by changing the electron configuration of the pseudopotential to  $[1s^{1.5}, 2s^2, 2p^{2.5}]$ . Each (half) core-hole calculation includes a total energy SCF calculation followed by a band structure calculation to converge the unoccupied states, and the subsequent evaluation of the dipole matrix elements using the projector augmented wave reconstruction method to reconstruct the core-electron wave functions [31]. A self-written post processing tool called MolPDOS [102] was used to postprocess the data and generate the spectra. This tool is contained in CASTEP version 18.1 and later and exists alongside other useful tools for this purpose such as OptaDOS [103]. The total NEXAFS spectrum is produced from summing all single carbon species contributions and shifting them with their corresponding core binding energies, resulting in the so-called  $\Delta$ IP-TP approach.

To make the NEXAFS spectra obtained by the calculations comparable to experimental data, a photon energy dependent broadening scheme using pseudo-Voigt functions [47, 48] was included. As laid out in section 2, the change in peak shape and width is meant to simulate the different life-times of excited states below and above the ionization potential. After the correction according to the XPS energies, the NEXAFS spectrum is divided into three ranges. The first range starts with the leading edge, spans the first 5 eV of the spectrum and is assigned a pseudo-Voigt FWHM of 0.75 eV and a 80%/20% Gaussian/Lorentzian ratio, while the third range starts from 15 eV above the leading edge and is assigned a FWHM of 2.0 eV and a 20%/80% G/L ratio. Both ranges are connected by an intermediate range, in which the FWHM and the G/L ratio change linearly. Every NEXAFS transition is represented by a

pseudo-Voigt peak with parameters according to its position on the photon energy axis. The values mentioned here were consistently used in this publication. If comparison to experiment is desired, these broadening parameters need to be adjusted accordingly, e.g. as done in figure S6 of the supplementary material.

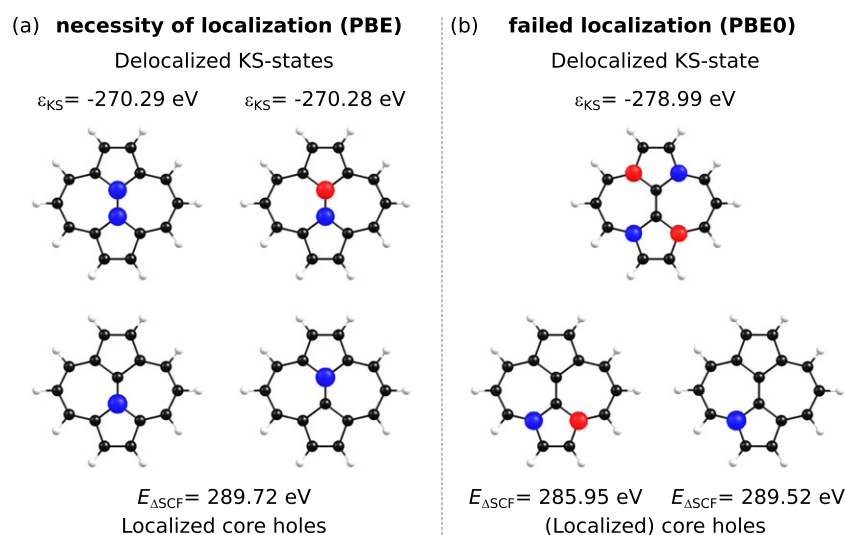
### 3. Results and discussion

In the following, we will discuss the calculation of x-ray absorption and photoelectron spectra for a small set of typical systems to exemplify the technical and numerical aspects, as well as the pitfalls involved in performing such calculations. We performed calculations for azupyrene (see figure 2(b)) and ethyl trifluoroacetate (ETFA, see figure 5(a)), as well as for the gas-phase azulene molecule, the azulene bulk crystal, and azulene adsorbed on a Cu(111) surface (see figure 8) [35]. ETFA shows extreme chemical shifts in its carbon 1s binding energies and has been an important reference system since the dawn of photoelectron spectroscopy [104–106]. Azupyrene is a polycyclic aromatic hydrocarbon, which was chosen because of its high symmetry, yielding a large number of equivalent carbon atoms. The three different systems involving the azulene molecule are used to show the versatility of our approach. A comprehensive experimental data set exists for azulene on Cu(111) which was previously used to show the feasibility of our DFT-based structures, energies and spectroscopic simulations [35, 36, 107].

#### 3.1. The choice of core-hole localization method for XPS simulations

The most straightforward way to calculate XPS energies via DFT is the use of the negative KS orbital energy values  $\epsilon_{KS}$  according to (2.5). In addition to neglecting relaxation effects, this approach carries an additional problem: the KS-states may be delocalized, especially for highly symmetric molecules, e.g. azupyrene, which possesses the symmetry point group  $D_{2h}$ . In such cases, the KS states for symmetrically equivalent atoms are linear combinations of the atomic core-states (see figure 4). If the respective atoms are close enough to produce appreciable overlap, an issue is encountered: the energy of the bonding combination of the core-states is lowered while the energy of the anti-bonding combination is raised. Therefore, two different binding energies exist for two symmetrically equivalent carbon atoms (see figure 4(a)).

The delocalization of the KS-states is not only an issue for the direct use of the KS energies to model binding energies, but also in  $\Delta$ SCF calculations that account for core-hole screening effects, (2.6). If the occupation of a certain core-level is changed to calculate the ionized state for the  $\Delta$ SCF calculation, for example by using the FOP functionality in FHI-aims to remove an electron from a KS state, it might distribute the core-hole over multiple atoms. The result is an unphysical system with a calculated binding energy between the values of KS-states and the properly localized core-holes (see figure 4(b)).



**Figure 4.** Core-hole localization for azupyrene. (a) The KS states of the two central C 1s orbitals are delocalized over both atoms, resulting in two different KS energies, which cannot be assigned to one carbon atom (upper part). The correctly localized core-holes in the  $\Delta\text{SCF}$  calculation yield the same binding energy for both symmetry equivalent carbons (lower part). (b) The core-hole localization, if not strictly enforced, often fails for hybrid functionals (e.g. PBE0). In this example, the localization worked for one case and produced a core-state properly localized on one carbon atom with a realistic binding energy (289.52 eV), while for the second case the core-state is still delocalized over two of the four symmetry equivalent atoms and the binding energy is unrealistically low (285.95 eV). The iso-value for all states shown is 0.1.

To avoid these issues, additional steps have to be taken to assure the proper localization of the core-hole in the calculations of the ionized state for the  $\Delta\text{SCF}$  approach. In particular, it is necessary to break the symmetry of the electronic wave function of the system. In a pseudopotential code this step is straightforward, because the use of a pseudopotential with a built-in core-hole for each atom automatically breaks the symmetry and localizes the core-hole [31]. The same is true for frozen core calculations, where the ionic occupation for the core-orbitals is frozen out with a core-hole localized at the desired atom [56].

For all-electron calculations, the problem is more difficult to solve. One approach suggested by Kahk and Lischner [27, 108] is to introduce a third step in addition to the ground state and the core-hole calculation. In this intermediate step, the symmetry of the wave function is broken, e.g. by introducing an increased core-charge to the atom in question. The (unphysical) wave function with broken symmetry and increased core-charge is then used as starting point for a core-hole calculation with normal core-charge at all atoms [27, 108]. The FOP functionality in FHI-aims is able to implement this approach. Alternatively, localization can also be achieved by using the FOB functionality in FHI-aims. In the latter, the occupation constraint on the KS state is defined by its overlap with a 1s atomic orbital basis function localized at the relevant atom. This approach in most cases automatically breaks the symmetry of the wave function during the self-consistent core-hole calculation.

It should be noted that both approaches, if the core-hole localization is successful, yield numerically identical results for the systems studied here. For most cases, we have found the FOB approach to be more stable with regard to electronic convergence and more convenient, because it does not require an additional initialization step. However, in the case of the highly

symmetric azupyrene molecule, the direct use of the FOB approach failed for all hybrid functionals (see figure 4(b)). To obtain the correct binding energies when using hybrid functionals it was necessary to first break symmetry using the Boys localization [94], as proposed in the literature [30]. This approach was successful for all molecule and XC functional combinations. It appears that hybrid functionals, contrary to GGA and meta-GGA functionals, suffer much more from variational collapse during  $\Delta\text{SCF}$  calculations and require a constraint that explicitly breaks symmetry to ensure core-hole localization. An additional way for solving this problem would be the use of a modified version of the  $\Delta\text{SCF}$  method, such as the local SCF method [109, 110] or the linear-expansion  $\Delta\text{SCF}$  [102] approach. Additionally, one of the tested meta-GGA functionals (SCAN) showed unphysical behaviour with great differences in the valence spin channels caused by the core-hole. For all other functionals the overall spin density is determined by the core-hole, showing a localised difference in the core-state spin channels on the excited atom, while the valence spin states only show minute differences yielding a vanishing spin density at the other atoms. But for SCAN the wave functions of the valence spin channels differ by a lot, leading to a larger difference in the up and down electron densities and therefore the spin-density distributed over the whole molecule. This finding is consistent with the known ‘over-magnetization’ of SCAN reported in the literature [111].

For the unsymmetrical molecule ETFA (symmetry point group  $C_1$  in the chosen rotational conformation), the localization poses less of a problem. Here, both tested meta-GGA functionals (SCAN and TPSS) failed when using the FOB approach for some atoms, but results could be obtained using the FOP.

The localization of the core-hole is also important for the calculation of NEXAFS spectra with all-electron DFT and was

recently discussed in detail in the literature [30]. In this work, we only perform NEXAFS simulations in the plane-wave code CASTEP with core-hole excited pseudopotentials. Here, the description of the (half) core-hole within the pseudopotential automatically ensures localization and symmetry breaking. The additional effects this core-hole excited pseudopotential exerts on XPS and NEXAFS signatures is discussed separately below.

### 3.2. The choice of the exchange correlation functional

To judge the influence of the exchange correlation functional on the absolute and relative C 1s XPS energies, a comprehensive study sampling 8 functionals of various types was carried out. The tested functionals included GGAs (PBE and PW91), meta-GGAs (SCAN and TPSS), hybrid functionals (PBE0, HSE06 and B3LYP), as well as the double hybrid functional xDH-PBE0. The test calculations were performed on the two gas phase molecules ETFA and azupyrene. These two molecules were chosen because they represent two different challenges when calculating XPS binding energies. ETFA (see figure 5(a)) is commonly used as example for XPS measurements and calculations due to the extreme chemical shifts of its four different carbon atoms [104–106]. On the other hand, azupyrene (see figure 2(b)) represents an alternative challenge in XPS calculations. It is a common organic molecule with carbon atoms in similar environments that produce slight but possibly measurable relative shifts in their C 1s binding energies. For both molecules, the calculations regarding the performance of the XC functionals were carried out using the all-electron code FHI-aims with a ‘tight-tier2’ core-augmented basis set.

The ETFA molecule possesses extreme chemical shifts for the C1s binding energies of up to 8 eV, leading to clearly separated peaks in a spectrum with typical experimental resolution. There is high quality experimental reference data available obtained by measurements on the gas-phase ETFA molecule [105, 106]. This experimental data is compared to the spectra calculated with different XC functionals in figure 5(a) on an absolute binding energy axis. For reference, all calculated binding energies are also compiled in table S1 of the supplementary material.

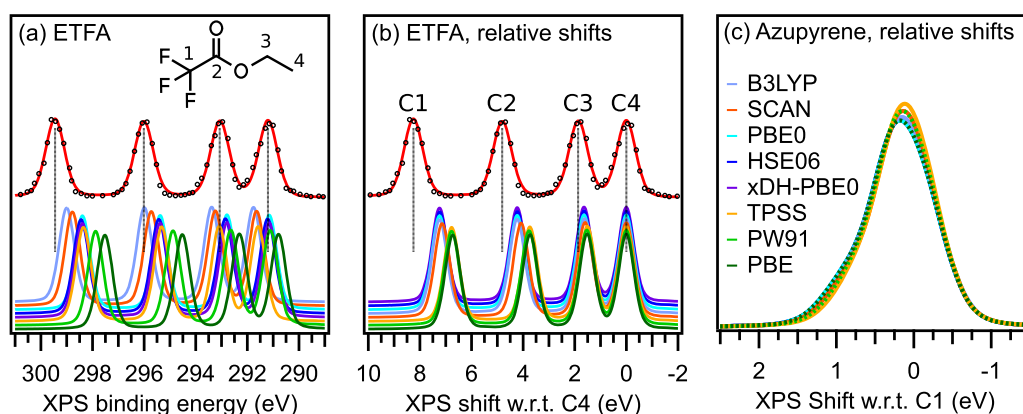
On first sight, the overall performance in terms of absolute binding energies is not straightforward to extract (figure 5(a), data tabulated in table S1). However, a clearer trend emerges when the spectra are plotted as relative shifts with respect to the lowest binding energy (figure 5(b), data tabulated in table S2). The performance of the different functionals regarding the relative shifts in the binding energies can be divided into three groups: (1) PBE, PW91, TPSS with average deviations of 0.74, 0.72, 0.73 eV; (2) SCAN with 0.53 eV; (3) PBE0, HSE06, B3LYP, xDH-PBE0 with 0.45, 0.46, 0.46, 0.45 eV, respectively. It is apparent that the hybrid functionals perform best here, and the GGAs perform worst, while TPSS behaves more like the GGAs and SCAN is somewhere in between.

When discussing the performance on the absolute binding energy scale, things are more complex. Due to an additional global displacement of all calculated energies, the average

deviation in the absolute binding energies would seem to be smallest for B3LYP and SCAN. But according to the relative shifts, the case can be made that all functionals have similar problems describing the binding energies of the carbon atoms with extreme binding energy shifts. In fact, the functionals PW91, PBE0, HSE06 and xD-PBE0 have an excellent agreement with the absolute binding energy of carbon C4, which possesses the lowest binding energy. Following this argument, the best performance for the absolute binding energies is provided by the hybrid functionals PBE0, HSE06 and xD-PBE0 with an average deviation of 0.50, 0.46, 0.50 eV and the GGA PW91 with the deviation of 0.81 eV.

By comparing absolute and relative deviations, it is obvious that the main error in these calculations is not only in the absolute energies, but already in the relative energies. Generally the hybrid functionals perform slightly better than the meta-GGAs and GGAs but their results are still quite poor compared to experimental data. Apparently all functionals have problems calculating the high binding energies resulting if the core-state is extremely descreened, e.g. for the carbon atom C1 in the CF<sub>3</sub> group of ETFA. For carbon atoms in more common chemical environments, e.g. C4 in ETFA, all functionals perform satisfactorily, i.e. yielding a deviation below or around 0.5 eV. It should also be noted that the use of the double hybrid functional xD-PBE0 brings no improvement relative to the regular hybrids. The better performance of hybrid functionals for *K*-shell XPS energies was also reported in the literature [112], while other studies found a better performance of regular GGAs [113]. However, many test sets used to judge the performance of binding energy calculations include only molecules with more moderate shifts [24, 63, 112, 113].

To further test the performance of the XC functionals for less extreme chemical shifts, we turn to the molecule azupyrene, which is an aromatic hydrocarbon without heteroatoms. The chemical shifts exhibited by its carbon atoms are therefore more subtle and only caused by its non-alternant topology, similar to the azulene molecule already discussed in the literature [35]. To compare the performance of the functionals with regard to these more subtle shifts, only relative differences in the binding energies are analysed. These relative shifts with respect to carbon C1 (see figure 2) are compared in table 1 for the different XC functionals. For reference, the absolute binding energies can be found in table S3. Within each class of functional used, GGA, meta-GGA and hybrid, the values obtained for the relative XPS shifts are similar with maximal deviations of 0.06 eV within the hybrids and 0.09 eV in the meta-GGAs. As a caveat, the meta-GGA SCAN showed problems with the response of the valence electron states to the core-holes and produced an unphysical spin density. However, the C1s shifts are still in line with the other functionals. It should be noted that the simple approach using Koopmans’ theorem also gives similar relative shifts in the binding energies (see table S4 in the supplementary material). This indicates that for the azupyrene molecule the shifts are mainly determined by the chemical environment of the initial state and that the relaxation of the core-hole is very similar for all chemical species.



**Figure 5.** Comparison of the C 1s XPS spectra of ETFA and azupyrene calculated with different XC functionals. (a) XPS spectrum of ETFA on the absolute binding energy scale with respect to the vacuum level. Upper part: experimental spectrum from reference [105] (open circles, data points; red line, fit function). Lower part: calculated spectra. Vertical lines as guide-for-the-eye. (b) XPS spectrum of ETFA on a relative binding energy scale w.r.t. the atom C4. (c) XPS spectrum of azupyrene on a relative binding energy scale w.r.t. C1 (see figure 2(b)). Both spectra were obtained by applying the same pseudo-Voigt broadening on the calculated C 1s binding energies to mimic a typical experimental resolution. Inset: structural formula of ETFA with numbering. The color scheme for the various functionals in (c) is valid for all panels.

**Table 1.** Relative shifts of the carbon 1s binding energies of the azupyrene molecule in dependence of the XC functional, calculated using the  $\Delta$ SCF approach in the all-electron code FHI-aims. All shifts are in eV and w.r.t. the binding energy of C1 (see figure 2(b)).

Carbon	PBE	PW91	HSE06	PBE0	B3LYP	SCAN	TPSS	XDH-PBE0
C1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C2	0.25	0.25	0.31	0.32	0.31	0.26	0.25	0.32
C3	0.25	0.25	0.26	0.26	0.22	0.23	0.19	0.26
C4	-0.18	-0.19	-0.19	-0.19	-0.20	-0.19	-0.18	-0.19
C5	0.82	0.82	0.90	0.91	0.85	0.86	0.77	0.91

Figure 5(b) shows the XPS spectrum of azupyrene obtained by applying the same pseudo-Voigt broadening as used in figure 5(a) to all relative binding energies in table 1. When the results obtained with the different functionals are thereby compared against the backdrop of a typical experimental resolution, they provide almost indistinguishable results. For this reason, we chose the functional PBE for all subsequent calculations.

### 3.3. The choice of basis set in XPS and NEXAFS simulations

In sections 3.1 and 3.2 we have discussed the role of core-hole localization and exchange–correlation functional when performing all-electron atomic orbital-based core-hole simulations of XPS binding energies. Another important numerical choice is the type and size of basis set that is used for such calculations. For atomic orbital basis sets, the importance of core augmentation functions has been extensively discussed in literature [27, 32, 60, 86]. We also find that basis set convergence for XPS binding energies in FHI-aims is very fast and, once core augmentation functions (according to reference [27]) are added, does not require to go beyond the standard ‘tight’ basis set definitions. The full data for the convergence series is shown in tables S5 to S8 in the supplementary material.

Moving from an all-electron basis to a pseudopotential plane wave formalism has the benefit of eliminating the issue of core-hole localization that was discussed in section 3.1, because core-hole excited pseudopotentials account for the core-excitation only on the relevant atom in question. However, this introduces other challenges. For example, the frozen core described by the pseudopotential does not relax and therefore absolute binding energies are typically much worse than if all-electron descriptions are used. Similarly, as the core-states are not explicitly treated, the core-level wave function needs to be rebuilt to enable the calculation of transition dipole moments between core and valence states in NEXAFS [31]. Lastly, it is not clear how the choice of pseudopotential affects core-level binding energies and XAS signatures.

Different types of pseudopotentials exist varying in their degree of ‘hardness’, i.e. the number of states that are pseudoized and the constraints that are placed on the core region [114]. Examples include non-local norm-conserving pseudopotentials [115] and ultra-soft pseudopotentials [116]. The latter require much lower energy cutoffs and fewer plane waves to accurately represent observables, which they achieve by lifting the constraint of charge conservation in the core region such that the wave functions satisfy the generalized orthonormality condition. We compared XPS and NEXAFS spectra calculated with three different pseudopotentials for

the example molecule azupyrene. Those three are based on the default settings of on-the-fly generated pseudopotentials in CASTEP and only differ in the type of projector, namely they use (for each angular momentum channel): (1) a norm-conserving projector, (2) a single ultrasoft projector, or (3) two ultrasoft projectors. All other parameters that define the pseudopotentials were left unchanged from the defaults. The detailed definitions are available in the raw data deposited on the NOMAD repository.

For the XPS calculations, we compared the spectra obtained with all three pseudopotential definitions with the spectrum generated by the all electron code FHI-aims. The XPS spectra of the different pseudopotentials had to be rigidly shifted by  $-6.1$  eV to align them to the all-electron results, because the lack of electron relaxation imposed by the frozen core approximation and the introduced charge background discussed below. The peak shape, however, is almost identical for all types of projectors and the all-electron method. The visualization of this result is shown in figure S1(a) and the relative peak positions of the different carbon species are compiled in table S9 of the supplementary material. In the case of NEXAFS simulations, the same is true and the different pseudopotentials provide almost identical spectra, with the exception of slight differences in intensities at high excitation energies (see figure S1(b) of the supplementary material). We therefore find that the main features of core-level spectra are robust with respect to the type of pseudopotential projection functions used. For all further calculations, we select the ultrasoft pseudopotential with two projectors per angular momentum channel, which is the default in CASTEP.

Another consequence of using a plane-wave pseudopotential basis set for core-hole simulations is the limitation that all calculations have to be performed using PBCs. The use of PBCs with a final state core-hole simulation approach such as  $\Delta$ SCF or TP can introduce errors in the calculation of XPS binding energies. Errors can arise due to finite size effects, when unit cells are too small and core-holes are interacting across periodic images. Another issue is when the positive charge of the core-hole is not compensated.  $\Delta$ SCF simulations of XPS binding energies and  $\Delta$ IP-TP simulations of NEXAFS transitions introduce a net charge of  $+1$  or  $+0.5$  e, respectively. Unit cells in PBCs must be charge neutral, otherwise the electrostatic potential would diverge. Therefore, a net charge is typically compensated by the introduction of a uniform (negative) background charge over the whole unit cell. This is a fail safe mechanism of an Ewald summation rather than an artefact, although it does introduce erroneous physics, such as a strong dependence of the total energy on the size of the unit cell.

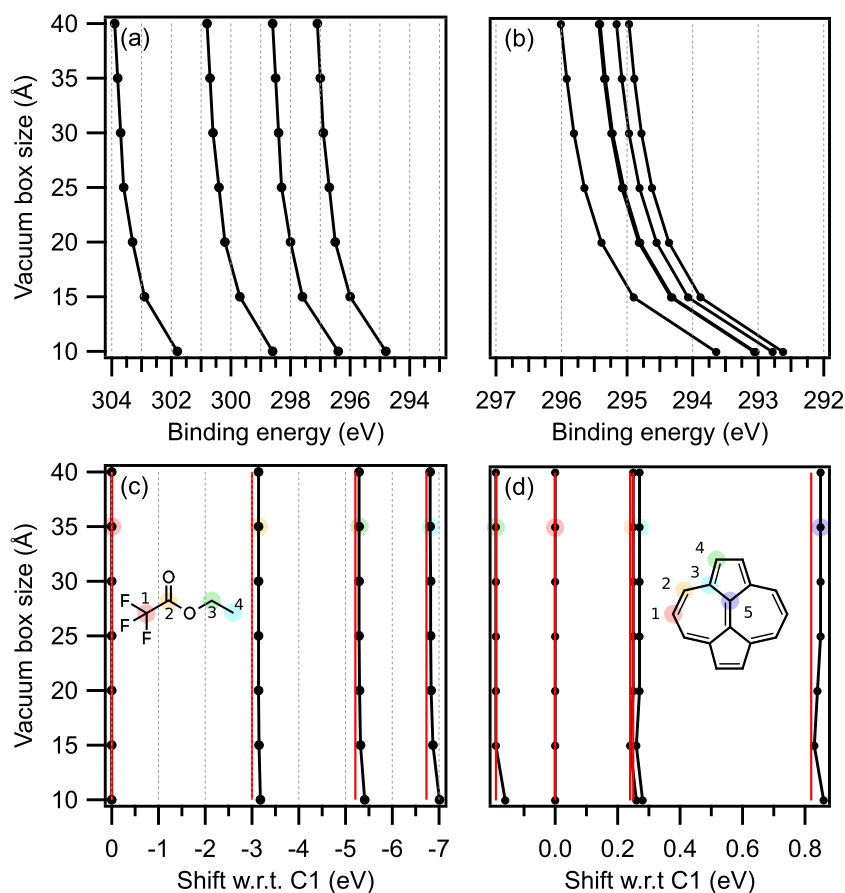
To calculate gas-phase molecules with PBC, the molecule is placed in a large vacuum box. Figure 6 shows the convergence behaviour of the C 1s binding energies for the gas-phase ETFA and azupyrene molecules as a function of the size of this vacuum box. The box is of cubic shape with increasing size from 10 to 40 Å in all directions. It can be seen that the absolute values of core-level binding energies (figures 6(a) and (b)) vary by 2–3 eV as the box size increases and are clearly not converged even at very large

vacuum box sizes. This effect can be attributed to the change in core-hole density and electrostatic interaction with the background charge and was recently studied in detail by Taucher *et al* [67]. However, this dependence on the box size of the absolute energies is not seen in the relative chemical shifts between the single carbon atoms within the molecules. In figures 6(c) and (d) the behaviour of the relative shifts with the vacuum box size is plotted and compared to the relative shifts obtained from the all-electron code FHI-aims (in red). Here it is clear that the relative shifts are much less affected by the background charge. While very small box sizes lead to slight deviations in relative shifts, the peaks converge very quickly and after about 20 Å box size, no change is seen. The converged relative shifts agree well with the all-electron calculations with average deviations of 0.1 eV for the extreme shifts in ETFA and 0.01 eV for the more subtle shifts in azupyrene. This result shows that absolute binding energies are affected by a homogeneous background compensation charge, but the issue is much less severe for relative energies where careful convergence can reproduce all-electron results. In addition to this data, we compare the performance of the  $\Delta$ SCF and  $\Delta$ IP-TP methods with their charge neutral variants for azulene adsorbed on Cu(111) in section 3.5.

The influence of the box size was also studied for the NEXAFS spectra calculated with the  $\Delta$ IP-TP method and shown in figure S2(a) of the supplementary material. As seen with the XPS binding energies, the NEXAFS spectra are converged at a box size of 20 Å where little to no change is seen afterwards. Regarding basis set convergence, the spectra show little change in the general shape and features with increased cut-off energy (see figure S2(b)), while the overall intensity increases and is converged after 450 eV, which is a plane wave cut-off energy similar to what is typically required for intermolecular interaction energies or adsorption energies with the employed ultrasoft pseudopotentials. We complete our discussion of numerical and technical parameters of core-hole simulations by studying the influence of the exchange–correlation functional on the simulated NEXAFS spectrum. We have calculated the NEXAFS spectrum of azupyrene with three different GGA functionals: PBE, BLYP and PW91 (see figure S3 of the supplemental materials). While there are slight differences for the intensities of higher energy transitions, the spectra show little variation overall and are almost identical for the lowest energy peaks. The deviations found by us are slightly smaller than what has previously been reported [24].

#### 3.4. Analysis of core-level spectra in terms of initial and final state contributions

Once XPS and XAS spectra are simulated, a comparison with experiment often requires a deeper analysis of the individual signatures that constitute the spectra. The decomposition in terms of the core-levels from which photoelectrons and excited electrons originate is trivial, as our approach uses separate core-hole calculations for each carbon atom (see figure 2). Therefore initial state decomposition is achieved by only summing up the transitions which originate from the 1s orbital localized at a specific carbon atom as shown in figure 7(a) for the azupyrene molecule.



**Figure 6.** Graphs comparing the influence of the vacuum box size on the XPS binding energies of gas-phase molecules in periodic calculations. ETFA on the left and azupyrene on right. (a) and (b) show the changes in the absolute binding energy of C 1s electrons. (c) and (d) show the relative shifts in C1s binding energy with respect to C1. Lines in red represent the relative shifts obtained from the all-electron FHI-aims calculations without PBC and the grey gridlines highlight 1 eV energy steps. All calculations were performed with the PBE functional.

An analysis in terms of final states in NEXAFS simulations requires further consideration, but can be particularly appealing for organic molecules adsorbed at surfaces and in thin-films, as spectra are often interpreted in terms of the symmetry of these states ( $\pi^*$ ,  $\sigma^*$ ). The final state decomposition for azupyrene in figure 7(b) shows the spectrum projected onto molecular orbital (MO) reference states according to their overlap with the final states  $f$  of each transition [102].

$$c_f^{\text{MO}} = |\langle \psi^{\text{MO}} | \psi_f \rangle|^2. \quad (3.1)$$

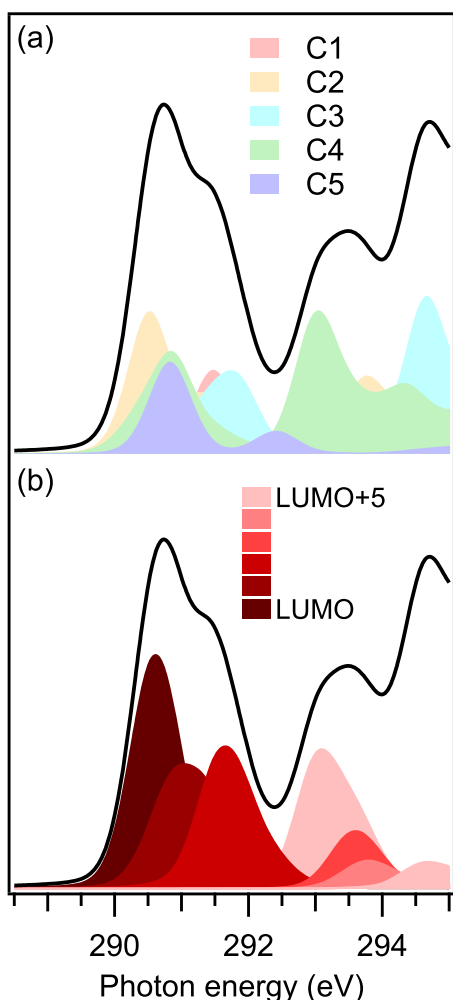
The projection coefficient  $c_f^{\text{MO}}$  is then multiplied with the cross section in (2.8) to select core-level transitions that lead to an excitation of a final state that has a contribution from the reference state MO:

$$\sigma_i^{\text{MO}}(h\nu) \propto \sum_f c_f^{\text{MO}} |\mu_{fi}|^2 \delta(h\nu - \Delta E_{fi}). \quad (3.2)$$

This approach is implemented in CASTEP via the MolPDOS tool [102], which enables such projections onto densities-of-states. For the projection, the reference states MO have to be

provided for each TP calculation in the form of a reference wave function. In the simplest case, this reference wave function is generated by the ground state calculation of exactly the same system as used for the TP calculations (as shown in figure 7(b)). In this case the projection scheme directly provides contributions of each unoccupied state to the overall NEXAFS spectrum.

However, it might be beneficial to choose a different reference system. For example, if the overall system in question is a molecule adsorbed on a metal surface, the reference system can be chosen as the free standing molecular overlayer (in exactly the same structure but without the surface). The projection of the NEXAFS transitions of the whole adsorbate system on the unoccupied states of the molecular overlayer now yields information about how the molecular states are changed by the adsorption and how they still contribute in a distinct manner to the overall spectrum. This approach was already used multiple times and proved its utility for molecules adsorbed on metal surfaces both with regard to NEXAFS transitions and for the analysis of the ground-state density of states [34–37, 117].



**Figure 7.** Different decomposition schemes for the simulated NEXAFS spectrum of azupyrene. (a) The initial state decomposition reveals which part of the spectrum is related to which carbon 1s orbital. For the presented case of the highly symmetric azupyrene molecule, the groups of symmetry equivalent atoms were summed up to improve visibility. (b) The final state decomposition shows the contributions of the transitions to each unoccupied molecular orbital up to LUMO+5.

### 3.5. XPS and NEXAFS simulation of organic crystals and metal–organic interfaces

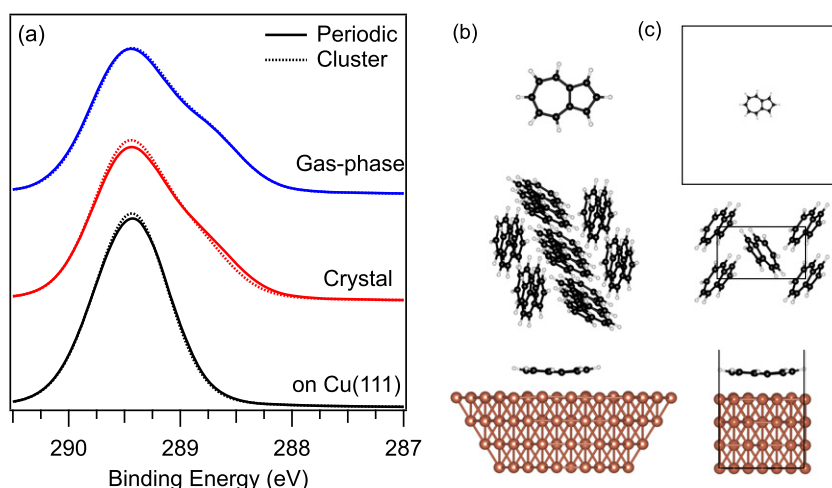
Up to now, we have only discussed core-level simulations of XPS and NEXAFS spectra for isolated molecules. In the following, the three different systems involving the azulene molecule are used to show the versatility of our approach and pinpoint special challenges for spectroscopic calculations of bulk and surface systems. As exemplar systems, we chose the azulene molecule both in a molecular crystal and adsorbed on the Cu(111) surface, depicted in figure 8, and compare both to the gas-phase molecule.

One additional issue of condensed systems compared to gas-phase calculations is the sampling of electronic states across the periodic crystal via Brillouin zone integration (or  $k$ -space sampling). Therefore, the convergence of both the XPS and NEXAFS calculations has to be ensured with respect to the number of  $k$ -points included. In the supplementary

material (see figures S4 and S5(a)), we provide a detailed discussion of the convergence properties of the XPS and NEXAFS spectra for the azulene molecular crystal and azulene adsorbed on Cu(111) when simulated with CASTEP. In short, both systems show rapid convergence with respect to the density of the employed Monkhorst–Pack  $k$ -grid [118]. In the case of azulene on Cu(111), we have additionally tested convergence with respect to the size of the vacuum slab that separates the surface model from its periodic image perpendicular to the surface and find that this parameter shows virtually no influence on the spectra (see figure S5(c)). Moreover, we also repeated the convergence series with regard to the cutoff energy of the plane wave basis and found that the convergence of the metal-adsorbed molecule is similar to the gas-phase molecule (see figure S5(b)).

The comparison between gas-phase azulene, the azulene molecular crystal and azulene adsorbed on Cu(111) allows us to show that all three systems can be treated with approximate core-hole simulation methods such as  $\Delta$ SCF and  $\Delta$ IP-TP, but also allows us to compare aperiodic simulations with FHI-aims and calculations under PBCs with CASTEP. First, XP spectra were calculated with the  $\Delta$ SCF approach in both FHI-aims and CASTEP. In the case of FHI-aims, we employ an aperiodic approach (also called cluster approach) where we correctly account for the positive charge in all three systems after photoionization. On the other hand, the plane wave nature of CASTEP made a 3D periodic cell necessary, even for those systems showing a lower periodicity. The cluster approach in FHI-aims can easily simulate the isolated molecule, while the inherently 3D periodic plane-wave approach has to simulate the molecule in a large periodic vacuum box (see section 3.3 for a discussion) at significant computational overhead. For the molecular crystal and the molecule adsorbed on the surface, the plane wave code is more suitable, because it already includes the 3D/2D periodicity. In the cluster approach for the all electron calculation of the crystal and the metal–organic interface, the choice of the cluster is not straightforward. For the molecular crystal, the 16 molecules surrounding a central molecule were cut out of the periodic crystal structure and the XPS shifts of the central molecule were calculated. For the surface-adsorbed molecule, a cluster was cut-out of the 2D periodic slab. The choice of the cluster was guided by the desire to preserve the hexagonal symmetry of the surface while simultaneously providing a sufficient surface area for adsorption. To reduce the number of subsurface atoms, the cluster was truncated into a conical shape, which becomes narrower in lower-lying layers. In the literature, no strong dependence of the XPS calculations on the specific cluster termination was found [27], which we can confirm. Still, the choice of both cluster models for the molecular crystal and the surface system is somewhat arbitrary, forming a draw-back of this method.

Figure 8(a) shows a comparison of the XPS calculation results for both approaches and all three systems. The overall spectra were shifted (to account for the difference in absolute binding energies due to the charge background issue within the PBCs) and a broadening scheme was applied to simulate experimental resolution. Both methods agree very well for all systems, be it the free molecule, the molecular crystal



**Figure 8.** XPS calculations of azulene, as a free molecule, in the molecular crystal, and when adsorbed on Cu(111). (a) Comparison of the XPS spectra, calculated both with the cluster approach in the all-electron code FHI-aims and in a periodic system with the plane-wave code CASTEP. The XPS spectra were shifted to align the peak maxima and to better show the difference in the peak shape; broadening was introduced to simulate a typical experimental resolution. Right side: structures of the used clusters (b) and periodic unit cells (c).

or the molecule adsorbed on the metal surface. This agreement is a very important result, because it proves that cluster and periodic approaches can yield compatible results, despite their respective challenges. For the systems investigated here, both approaches appear to provide excellent agreement for the obtained relative binding energies. We note, however, that the cluster calculations performed with FHI-aims, both for the molecular crystal and the molecule adsorbed on the surface were much more computationally demanding than the periodic calculations. The all-electron calculations required vastly more computational resources and proved to be less stable with regards to electronic convergence than the periodic calculations, which are straightforward as implemented in CASTEP.

The calculated spectra also provide insight into the interactions present in each system. It is apparent that the interaction between the azulene molecules in the molecular crystal exerts only a minor influence on the XPS peak, which retains its gas-phase shape. The interaction between the molecule and the Cu(111) surface, on the other hand, changes the peak shape a lot, which is an indication of the increased molecule–metal interaction and was already discussed in connection to the experimental data elsewhere [35, 36]. For the adsorbed azulene, we further note that coverage as modelled within PBCs is different to the low-coverage situation captured by the cluster. However, the coverage only exerts a minor influence on the XPS binding energies for this system, as was also observed in experiment.

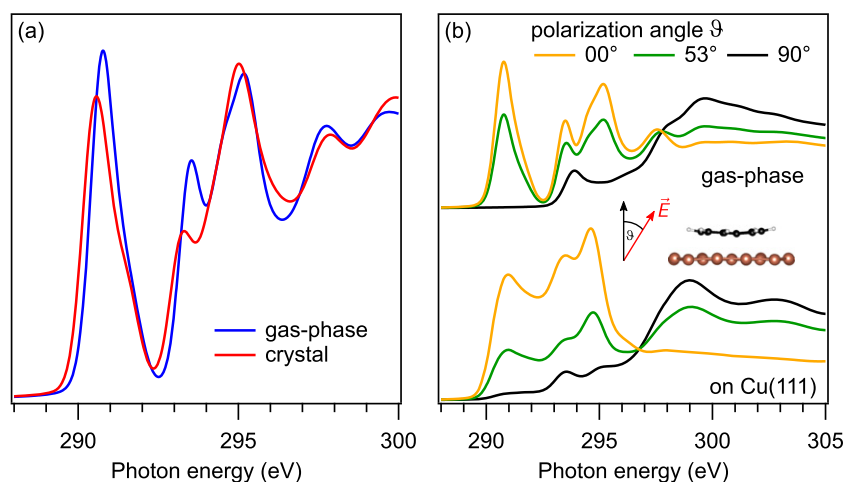
The NEXAFS spectra of all three systems, calculated with the  $\Delta$ IP-TP approach in the plane-wave code CASTEP, are compared in figure 9. The spectra were not shifted with respect to each other. The good agreement, despite the presence of the space-charge effect, is due to the ionization potential correction ( $\Delta$ IP) of the NEXAFS transitions according to the C1s binding energy of each carbon. In figure 9(a) it can be seen that the spectra of free molecule and molecular crystal are very similar. For the NEXAFS spectrum of the adsorbed molecule

(figure 9(b)), the angular dependence of the NEXAFS transitions with regard to the polarization vector is important. Therefore we chose to simulate three different polar angles  $\vartheta$  for the polarization direction. Two of those correspond to the polarization direction parallel ( $90^\circ$ ) and orthogonal to the surface ( $0^\circ$ ), while the third is the so called *magic angle* ( $53^\circ$ ). The *magic angle* is the polarization angle, at which the orientation of the transition dipole moment  $\mu_{fi}$  relative to the substrate does not influence the transition intensity. It should be noted that the ideal magic angle (for our threefold-symmetric substrate) is in fact  $54.7^\circ$  [2]. However, the actual magic angle encountered in an experiment is dependent on the degree of polarization of the used radiation [2, 10]. The value for the degree of polarization of 0.9 used in the corresponding experiments [35] leads to a magic angle of  $53^\circ$ . For all polar angles, the spectra are averaged over the azimuthal angle, due to the symmetry of the substrate (see (2.19)). The comparison between NEXAFS of gas-phase and adsorbed molecule reveals stark differences in the spectra, giving insight into the mechanism of the molecule–surface interaction. While the overall dichroism reveals that the molecule lies approximately flat on the Cu(111) surface, there is another telling detail in the spectra. The non-vanishing intensity of the C 1s to LUMO transition (291 eV) for the  $90^\circ$  polar angle shows that the clear separation of  $\sigma^*$  and  $\pi^*$  states (as present for the free molecule) is now broken due to the hybridization between the electronic states of adsorbed molecule and metal surface [34, 119]. The corresponding experimental spectra, the simulated data, and the subsequent interpretation were already reported in previous publications and show a good agreement between experiment and theory [35, 36].

### 3.6. The effect of charge in periodic core-hole simulations

We use this opportunity of having reliable experimental data for azulene thin films and azulene monolayers adsorbed on the Cu(111) surface to show a systematic comparison of the





**Figure 9.** NEXAFS simulations of azulene in the periodic systems with the plane-wave code CASTEP. (a) Comparison of the spectra of the gas-phase molecule and the molecular crystal. (b) Comparison of the spectra of the gas-phase molecule and the molecule adsorbed on Cu(111) including angular dependence on the polarization direction of the incident radiation relative to the surface or molecular plane, respectively. The NEXAFS spectra include broadening to simulate a typical experimental resolution.

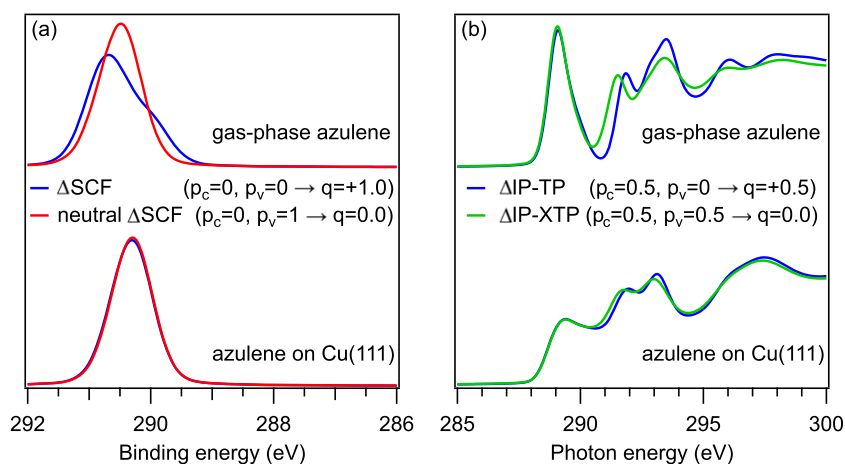
$\Delta$ IP-TP method with the similar  $\Delta$ IP-XTP calculations. The advantage of the XTP method is that it creates a net-neutral unit cell which does not require a spurious uniform background counter charge, which is otherwise present in the periodic  $\Delta$ SCF and ( $\Delta$ IP-)TP calculations and can significantly influence electrostatic properties [67]. The array of calculations making up a  $\Delta$ IP-XTP simulation differs only in the third step (as shown in figure 2(c)) from the  $\Delta$ IP-TP method: we perform  $N$  half core-hole (but zero charge) XTP calculations where we introduce half an electron into the lowest unoccupied state to compensate for the core-hole.

The results of the  $\Delta$ IP-TP and  $\Delta$ IP-XTP simulations are compared for the azulene molecule in the gas phase and adsorbed on the copper surface in figure 10. In this figure all spectra were rigidly shifted to compensate for the global shift due to the different charge state of the systems and therefore enable a better inspection of their spectral shape. Both methods produce virtually indistinguishable XPS peak shapes for the molecule adsorbed on the metal surface, while stark differences are present for the gas-phase molecule (see figure 10(a)). This fact is unsurprising, because the peculiar peak shape of the gas-phase azulene molecule is due to the localized charge distribution, as already discussed in the literature on the backdrop of experimental XPS data for thin multicrystalline films of azulene [35]. The neutral  $\Delta$ SCF calculations do not remove the electron taken from the core-state, but instead promote it into the lowest unoccupied state. The occupation of the former LUMO has massive consequences for the electronic structure of the gas-phase molecule and therefore the resulting spectra are negatively influenced. On the other hand, in the case of azulene adsorbed on Cu(111), the continuous density of states of the metal surface around the Fermi level has the effect that the additional electron does not present a significant change in the electronic structure and the resulting spectra remain similar.

For the proper simulation of the NEXAFS spectra in figure 10(b), the obvious error in the neutralized  $\Delta$ SCF calculations for the free molecule made it necessary to perform the  $\Delta$ IP shift both for the TP and XTP simulation on the basis of the charged  $\Delta$ SCF calculations. The results show that again the spectra for the molecule adsorbed on the metal surface are almost identical for  $\Delta$ IP-TP and  $\Delta$ IP-XTP, while there are some differences for the gas-phase molecule.

In comparison to various core-hole constraining methods, Michelitsch and Reuter have previously reported that the XTP method (note that no  $\Delta$ IP correction is applied) on balance provides the best agreement with experiment for absolute energies and intensities of first and second transition energies of NEXAFS spectra for a range of isolated organic molecules [30]. We also compared the  $\Delta$ IP-TP and  $\Delta$ IP-XTP spectra with the experimental data for the multicrystalline azulene films from reference [37] (see figure S6 in the supplemental material). The quantitative comparison extracted from these spectra yields an average deviation of 0.13 eV for  $\Delta$ IP-TP and 0.05 eV for  $\Delta$ IP-XTP, when the relative shifts between the four lowest-energy peaks are compared to the experiment. The relative peak intensities for those four peaks deviate by average factors of 1.8 and 1.5 for  $\Delta$ IP-TP and  $\Delta$ IP-XTP, respectively. Therefore we found both  $\Delta$ IP-TP and  $\Delta$ IP-XTP to be in good agreement with the experiment, with  $\Delta$ IP-XTP providing slightly smaller deviations both in transition energies and intensities. As can be seen by this quantitative comparison, the approximate treatment with the  $\Delta$ IP-TP and  $\Delta$ IP-XTP methods based on transitions with infinite lifetime and empirical lifetime broadening already provides a reasonable agreement with experiment in the near-edge region, even as many-body effects such as inelastic scattering, Auger effects, and vibronic coupling are neglected.

In the future we will provide a more systematic study of method differences on a larger set of benchmark systems, to test if the performance of the  $\Delta$ IP-TP and  $\Delta$ IP-XTP approaches is robust across a variety of cases including



**Figure 10.** Core-level spectra calculated with and without background charge compensation for azulene molecule in the gas-phase and adsorbed on the Cu(111) surface. (a) Comparison of  $\Delta$ SCF XP spectra calculations with a homogeneous compensation charge (blue) and with a compensation charge introduced in the lowest unoccupied state (red). (b) Comparison of the  $\Delta$ IP-TP (blue) and  $\Delta$ IP-XTP (green) NEXAFS spectra. Both spectra share a  $0.5 e^-$  core-hole. The TP calculation includes  $-0.5 e^-$  background charge (blue) while the XTP calculations compensate charge by adding half an electron into the lowest unoccupied state (green). All spectra include broadening to simulate a typical experimental resolution and were shifted for better comparison.

more challenging systems, e.g. containing coherent dipolar arrangements [67].

#### 4. Conclusion and future outlook

We have presented the numerical and technical details of DFT-based core-hole simulation methods to calculate 1s XPS and *K*-edge NEXAFS signatures of organic molecules, organic molecular crystals, and metal–organic interfaces. The introduced methods are variants of the  $\Delta$ SCF method and the TP approach ( $\Delta$ IP-TP). For the benefit of practitioners in the field, we have compared and contrasted key parameter choices in performing such simulations including the choice of exchange–correlation functional, basis set, and if the model is to be aperiodic or treated within PBCs. We find that all tested functionals perform well for absolute and relative XPS binding energies, as long as the shifts are not too extreme due to strong descreening. For such extreme shifts, as encountered in the ETFA molecule, hybrid functionals perform better than GGAs and meta-GGAs. For periodic calculations, the calculation of absolute energies is more difficult, however, we find that periodic and aperiodic calculations yield comparable results for relative shifts, even in the presence of the inherent space charge problem due to the use of a compensating background charge.

The choice of periodic or aperiodic description for the simulation of XPS and NEXAFS spectra is strongly dependent on the system under investigation and comes with a complex trade-off regarding advantages and disadvantages. For example, aperiodic atomic-orbital based calculations, as we have performed with the FHI-aims package, can provide accurate core-level binding energies [27] and NEXAFS transitions [30] for an isolated molecule, a molecular crystal, and a metal-adsorbed molecule, but lack robustness with respect to core-hole localization and provide substantial computational overhead for the latter two systems. More advanced

approaches to core-hole localization based on excited-state orbital optimization [120] or more sophisticated constraint definition [102, 109, 110] may be necessary to resolve this generally. In the case of a periodic pseudopotential plane wave description, as provided within CASTEP, the core-hole localization is in-built and the computations are efficient and numerically robust. However, the employed PBC can introduce an artificial charge background and spurious electrostatics with some core-hole constraining approaches [67]. We show that both, aperiodic and periodic approaches yield virtually identical relative core-level binding energies and that the relative peak positions in XPS and NEXAFS spectra do not appear to be strongly affected, even when different charge compensation schemes are used. However, we note that other systems with strong dipoles and no Fermi level pinning may show more significant artefacts in spectra when a homogeneous charge background is present. Furthermore, the comparison of absolute binding energies between simulation and experimental data for condensed matter systems carries additional pitfalls not addressed in this publication. We will systematically assess both issues in future work.

While the here presented constraint-based core-hole simulation methods are efficient and robust, they fail to describe many critical features of core-hole spectra such as Auger effects and energy-dependent lifetime broadening. Highly accurate first-principles methods based on many-body perturbation theory [79, 121–123] have recently emerged to address this. Nevertheless, approximate core-hole simulation methods remain essential workhorses for highly complex spectroscopy simulations with applications ranging from in-operando electrocatalysis [124] to ultrafast dynamics [125] and pump–probe spectroscopy [126]. The efficiency of these methods enables the creation of high-volume data [112, 127, 128] which can be used for machine-learning-assisted spectral assignment [129–131]. The here presented  $\Delta$ SCF and  $\Delta$ IP-TP approaches as implemented in CASTEP are versatile and reliable for

organic molecules in different environments and represent important tools in this context.

## Acknowledgments

SJH and RJM acknowledge funding for a PhD studentship through the EPSRC Centre for Doctoral Training in Molecular Analytical Science (EP/L015307/1) and computing resources via the EPSRC-funded HPC Midlands+ computing centre (EP/P020232/1) and the EPSRC-funded Materials Chemistry Consortium for the ARCHER UK National Supercomputing Service (EP/R029431/1). RJM acknowledges support via a UKRI Future Leaders Fellowship (MR/S016023/1). The authors acknowledge many fruitful and enjoyable discussions with Katharina Diller and David Duncan.

## Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://doi.org/10.17172/NOMAD/2020.10.15-1>.

## Conflict of interest

The authors declare no conflict of interest.

## ORCID iDs

Benedikt P Klein  <https://orcid.org/0000-0002-6205-8879>  
 Samuel J Hall  <https://orcid.org/0000-0003-3765-828X>  
 Reinhard J Maurer  <https://orcid.org/0000-0002-3004-785X>

## References

- [1] Frati F, Hunault M O J Y and De Groot F M F 2020 *Chem. Rev.* **120** 4056–110
- [2] Stöhr J 1992 *NEXAFS Spectroscopy* (Berlin: Springer)
- [3] Bracco G and Holst B (ed) 2013 *Surface Science Techniques* (Berlin: Springer)
- [4] Bagus P S et al 2019 *Phys. Rev. B* **100** 115419
- [5] Presel F, Jabeen N, Pozzo M, Curcio D, Omiciuolo L, Lacovig P, Lizzit S, Alfè D and Baraldi A 2015 *Carbon* **93** 187–98
- [6] Presel F, Tache C A, Tetlow H, Curcio D, Lacovig P, Kantorovich L, Lizzit S and Baraldi A 2017 *J. Phys. Chem. C* **121** 11335–45
- [7] Netzer F P and Ramsey M G 1992 *Crit. Rev. Solid State Mater. Sci.* **17** 397–475
- [8] Gao Y 1999 *Acc. Chem. Res.* **32** 247–55
- [9] Papp C and Steinrück H-P 2013 *Surf. Sci. Rep.* **68** 446–87
- [10] Breuer T, Klues M and Witte G 2015 *J. Electron Spectrosc. Relat. Phenom.* **204** 102–15
- [11] Haase J 1985 *Appl. Phys. A* **38** 181–90
- [12] Niskanen J, Sahle C J, Gilmore K, Uhlig F, Smiatek J and Föhlisch A 2017 *Phys. Rev. E* **96** 013319
- [13] Powell C J and Jablonski A 2009 *Nucl. Instrum. Methods Phys. Res. A* **601** 54–65
- [14] Besley N A 2020 *Acc. Chem. Res.* **53** 1306–15
- [15] Tenorio B N C, Oliveira R R, Nascimento M A C and Rocha A B 2018 *J. Chem. Theory Comput.* **14** 5324–38
- [16] Golze D, Wilhelm J, van Setten M J and Rinke P 2018 *J. Chem. Theory Comput.* **14** 4856–69
- [17] Golze D, Dvorak M and Rinke P 2019 *Front. Chem.* **7** 377
- [18] Helmich-Paris B 2020 posted online 13.04.2020 accessed: 01.02.2021 arXiv:2004.05845
- [19] Liu J, Matthews D, Coriani S and Cheng L 2019 *J. Chem. Theory Comput.* **15** 1642–51
- [20] Vidal M L, Pokhilko P, Krylov A I and Coriani S 2020 *J. Phys. Chem. Lett.* **11** 8314–21
- [21] Gunnarsson O and Lundqvist B I 1976 *Phys. Rev. B* **13** 4274–98
- [22] von Barth U 1979 *Phys. Rev. A* **20** 1693–703
- [23] von Barth U and Grossmann G 1980 *Phys. Scr.* **21** 580–4
- [24] Triguero L, Pettersson L G M and Ågren H 1998 *Phys. Rev. B* **58** 8097–110
- [25] Takahata Y and Chong D P 2003 *J. Electron Spectrosc. Relat. Phenom.* **133** 69–76
- [26] Prendergast D and Galli G 2006 *Phys. Rev. Lett.* **96** 215502
- [27] Kahk J M and Lischner J 2019 *Phys. Rev. Mater.* **3** 100801
- [28] Bunău O and Calandra M 2013 *Phys. Rev. B* **87** 205105
- [29] Donval G, Moreau P, Danet J, Larbi S J-S, Bayle-Guillemaud P and Boucher F 2017 *Phys. Chem. Chem. Phys.* **19** 1320–7
- [30] Micheli G S and Reuter K 2019 *J. Chem. Phys.* **150** 074104
- [31] Gao S-P, Pickard C J, Perlov A and Milman V 2009 *J. Phys.: Condens. Matter* **21** 104203
- [32] Ambroise M A and Jensen F 2019 *J. Chem. Theory Comput.* **15** 325–37
- [33] Blobner F, Abufager P N, Han R, Bauer J, Duncan D A, Maurer R J, Reuter K, Feulner P and Allegretti F 2015 *J. Phys. Chem. C* **119** 15455–68
- [34] Diller K, Maurer R J, Müller M and Reuter K 2017 *J. Chem. Phys.* **146** 214701
- [35] Klein B P et al 2019 *Phys. Rev. X* **9** 011030
- [36] Klein B P et al 2019 *J. Phys. Chem. C* **123** 29219–30
- [37] Klein B P et al 2020 *Chem. Mater.* **32** 1041–53
- [38] Reinert F and Hüfner S 2005 *New J. Phys.* **7** 97
- [39] Osterwalder J 2012 *Surface and Interface Science* vol 1 (New York: Wiley) pp 151–214
- [40] Diller K et al 2014 *J. Chem. Phys.* **141** 144703
- [41] Taucher T C, Hehn I, Hofmann O T, Zharnikov M and Zojer E 2016 *J. Phys. Chem. C* **120** 3428–37
- [42] Hedin L 1999 *J. Phys.: Condens. Matter* **11** 489–528
- [43] Doniach S and Sunjic M 1970 *J. Phys. C: Solid State Phys.* **3** 7
- [44] Green J C and Decleva P 2005 *Coord. Chem. Rev.* **249** 209–28
- [45] Puschign P et al 2011 *Phys. Rev. B* **84** 235427
- [46] Shang M-H, Nagaosa M, Nagamatsu S-i, Hosoumi S, Kera S, Fujikawa T and Ueno N 2011 *J. Electron Spectrosc. Relat. Phenom.* **184** 261–4
- [47] Schmid M, Steinrück H-P and Gottfried J M 2014 *Surf. Interface Anal.* **46** 505–11
- [48] Schmid M, Steinrück H P and Gottfried J M 2015 *Surf. Interface Anal.* **47** 1080
- [49] Zheng X and Cheng L 2019 *J. Chem. Theory Comput.* **15** 4945–55
- [50] Kas J J, Vila F D, Rehr J J and Chambers S A 2015 *Phys. Rev. B* **91** 121112
- [51] Koopmans T 1934 *Physica* **1** 104–13
- [52] Janak J F 1978 *Phys. Rev. B* **18** 7165–8
- [53] Bellafont N P, Illas F and Bagus P S 2015 *Phys. Chem. Chem. Phys.* **17** 4015–9
- [54] Williams A R and Lang N D 1978 *Phys. Rev. Lett.* **40** 954–7
- [55] Pueyo Bellafont N, Bagus P S and Illas F 2015 *J. Chem. Phys.* **142** 214102
- [56] Chong D P, Gritsenko O V and Baerends E J 2002 *J. Chem. Phys.* **116** 1760–72
- [57] Slater J C 1972 *Adv. Quantum Chem.* **6** 1–92
- [58] Bagus P S 1965 *Phys. Rev.* **139** A619
- [59] Johansson Å and Stafström S 1999 *J. Chem. Phys.* **111** 3203–8

- [60] Besley N A, Gilbert A T B and Gill P M W 2009 *J. Chem. Phys.* **130** 124308
- [61] Viñes F, Sousa C and Illas F 2018 *Phys. Chem. Chem. Phys.* **20** 8403–10
- [62] Williams A R, DeGroot R A and Sommers C B 1975 *J. Chem. Phys.* **63** 628–31
- [63] Chong D P 2007 *J. Electron Spectrosc. Relat. Phenom.* **159** 94–6
- [64] Noodleman L, Post D and Baerends E J 1982 *Chem. Phys.* **64** 159–66
- [65] Chong D P 2010 *Can. J. Chem.* **88** 787–96
- [66] Okazaki T and Laali K K 2004 *Org. Biomol. Chem.* **2** 2214–9
- [67] Taucher T C, Hofmann O T and Zojer E 2020 *ACS Omega* **5** 25868–81
- [68] Hähner G 2006 *Chem. Soc. Rev.* **35** 1244–55
- [69] Tanaka I and Mizoguchi T 2009 *J. Phys.: Condens. Matter* **21** 104201
- [70] Mizoguchi T, Tanaka I, Gao S-P and Pickard C J 2009 *J. Phys.: Condens. Matter* **21** 104204
- [71] Mizoguchi T, Olovsson W, Ikeno H and Tanaka I 2010 *Micron* **41** 695–709
- [72] Norman P and Dreuw A 2018 *Chem. Rev.* **118** 7208–48
- [73] Guda A A *et al* 2019 *Catal. Today* **336** 3–21
- [74] Rehr J J and Albers R C 2000 *Rev. Mod. Phys.* **72** 621–54
- [75] Lopata K, Van Kuiken B E, Khalil M and Govind N 2012 *J. Chem. Theory Comput.* **8** 3284–92
- [76] Rehr J J, Soininen J A and Shirley E L 2005 *Phys. Scr. T* **115** 207–11
- [77] Olovsson W, Tanaka I, Mizoguchi T, Puschnig P and Ambrosch-Draxl C 2009 *Phys. Rev. B* **79** 041102
- [78] Vinson J, Rehr J J, Kas J J and Shirley E L 2011 *Phys. Rev. B* **83** 115106
- [79] Gilmore K, Vinson J, Shirley E L, Prendergast D, Pemmaraju C D, Kas J J, Vila F D and Rehr J J 2015 *Comput. Phys. Commun.* **197** 109–17
- [80] Stener M, Lisini A and Decleva P 1995 *Chem. Phys.* **191** 141–54
- [81] Hu C-H and Chong D P 1996 *Chem. Phys. Lett.* **262** 729–32
- [82] Hetényi B, De Angelis F, Giannozzi P and Car R 2004 *J. Chem. Phys.* **120** 8632–7
- [83] Tanaka I, Araki H, Yoshiya M, Mizoguchi T, Ogasawara K and Adachi H 1999 *Phys. Rev. B* **60** 4944–51
- [84] Ozaki T and Lee C C 2017 *Phys. Rev. Lett.* **118** 026401
- [85] Freysoldt C, Mishra A, Ashton M and Neugebauer J 2020 *Phys. Rev. B* **102** 04 5403
- [86] Sarangi R, Vidal M L, Coriani S and Krylov A I 2020 *Mol. Phys.* **118** e1769872
- [87] Kolczewski C *et al* 2001 *J. Chem. Phys.* **115** 6426–37
- [88] Püttner R, Kolczewski C, Martins M, Schlachter A S, Snell G, Sant’Anna M, Viehhaus J, Hermann K and Kaindl G 2004 *Chem. Phys. Lett.* **393** 361–6
- [89] Hébert-Souche C, Louf P-H, Blaha P, Nelhiebel M, Luitz J, Schattschneider P, Schwarz K and Jouffrey B 2000 *Ultra-microscopy* **83** 9–16
- [90] Duncan D A, Pfisterer J H K, Deimel P S, Acres R G, Fritton M, Feulner P, Barth J V and Allegretti F 2016 *Phys. Chem. Chem. Phys.* **18** 20433–42
- [91] Clark S J, Segall M D, Pickard C J, Hasnip P J, Probert M I J, Refson K and Payne M C 2005 *Z. Kristallogr.* **220** 567–70
- [92] Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K and Scheffler M 2009 *Comput. Phys. Commun.* **180** 2175–96
- [93] Rossi M, Ceriotti M and Manolopoulos D E 2016 *J. Phys. Chem. Lett.* **7** 3001–7
- [94] Foster J M and Boys S F 1960 *Rev. Mod. Phys.* **32** 300–2
- [95] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865–8
- [96] Perdew J P, Chevary J A, Vosko S H, Jackson K A, Pederson M R, Singh D J and Fiolhais C 1992 *Phys. Rev. B* **46** 6671–87
- [97] Heyd J, Scuseria G E and Ernzerhof M 2003 *J. Chem. Phys.* **118** 8207–15
- [98] Adamo C and Barone V 1999 *J. Chem. Phys.* **110** 6158–70
- [99] Tao J, Perdew J P, Staroverov V N and Scuseria G E 2003 *Phys. Rev. Lett.* **91** 146401
- [100] Sun J, Ruzsinszky A and Perdew J P 2015 *Phys. Rev. Lett.* **115** 036402
- [101] Zhang I Y, Su N Q, Brémond É A G, Adamo C and Xu X 2012 *J. Chem. Phys.* **136** 174103
- [102] Maurer R J and Reuter K 2013 *J. Chem. Phys.* **139** 014708
- [103] Nicholls R J, Morris A J, Pickard C J and Yates J R 2012 *J. Phys.: Conf. Ser.* **371** 012062 (Institute of Physics Publishing)
- [104] Siegbahn K 1973 *Atomic Physics 3* ed W G Smith S J (Berlin: Springer) pp 137–91
- [105] Gelius U, Basilier E, Svensson S, Bergmark T and Siegbahn K 1973 *J. Electron Spectrosc. Relat. Phenom.* **2** 405–34
- [106] Travnikova O, Børve K J, Patanen M, Söderström J, Miron C, Sæthre L J, Mårtensson N and Svensson S 2012 *J. Electron Spectrosc. Relat. Phenom.* **185** 191–7 (Special Issue in honor of Professor T Darrah Thomas: High-Resolution Spectroscopy of Isolated Species)
- [107] Kachel S R *et al* 2020 *J. Phys. Chem. C* **124** 8257–68
- [108] Kahk J M and Lischner J 2018 *Phys. Chem. Chem. Phys.* **20** 30403–11
- [109] Ferré N and Assfeld X 2002 *J. Chem. Phys.* **117** 4119–25
- [110] Loos P-F and Assfeld X 2007 *Int. J. Quantum Chem.* **107** 2243–52
- [111] Mejia-Rodriguez D and Trickey S B 2019 *Phys. Rev. B* **100** 041113
- [112] Pi J M, Stella M, Fernando N K, Lam A Y, Regoutz A and Ratcliff L E 2020 *J. Phys. Chem. Lett.* **11** 2256–62
- [113] Tolbatov I, Chipman D M and Shavitt I 2016 *Highlights in Theoretical Chemistry* vol 9 ed D T Shepard and R Pitzer (Berlin: Springer)
- [114] Payne M C, Teter M P, Allan D C, Arias T A and Joannopoulos J D 1992 *Rev. Mod. Phys.* **64** 1045–97
- [115] Kleinman L and Bylander D M 1982 *Phys. Rev. Lett.* **48** 1425–8
- [116] Vanderbilt D 1990 *Phys. Rev. B* **41** 7892–5
- [117] Müller M, Diller K, Maurer R J and Reuter K 2016 *J. Chem. Phys.* **144** 024701
- [118] Monkhorst H J and Pack J D 1976 *Phys. Rev. B* **13** 5188–92
- [119] Mainka C, Bagus P S, Schertel A, Strunskus T, Grunze M and Wöll C 1995 *Surf. Sci.* **341** L1055–60
- [120] Hait D, Haugen E A, Yang Z, Oosterbaan K J, Leone S R and Head-Gordon M 2020 *J. Chem. Phys.* **153** 134108
- [121] Liang Y, Vinson J, Pemmaraju S, Drisdell W S, Shirley E L and Prendergast D 2017 *Phys. Rev. Lett.* **118** 096402
- [122] Liang Y and Prendergast D 2018 *Phys. Rev. B* **97** 205127
- [123] Liang Y and Prendergast D 2019 *Phys. Rev. B* **100** 075121
- [124] Nattino F and Marzari N 2020 *Phys. Chem. Chem. Phys.* **22** 10807–18
- [125] Northey T, Norell J, Fouda A E A, Besley N A, Odelius M and Penfold T J 2020 *Phys. Chem. Chem. Phys.* **22** 2667–76
- [126] Ehlert C, Gühr M and Saalfrank P 2018 *J. Chem. Phys.* **149** 144112
- [127] Mathew K, Zheng C, Winston D, Chen C, Dozier A, Rehr J J, Ong S P and Persson K A 2018 *Scientific Data* **5** 1–8
- [128] Aarva A, Deringer V L, Sainio S, Laurila T and Caro M A 2019 *Chem. Mater.* **31** 9243–55
- [129] Aarva A, Deringer V L, Sainio S, Laurila T and Caro M A 2019 *Chem. Mater.* **31** 9256–67
- [130] Zheng C *et al* 2018 *npj Comput. Mater.* **4** 12
- [131] Rankine C D, Madkhali M M M, Penfold T J and Penfold T J 2020 *J. Phys. Chem. A* **124** 4263–70