Density-functional theory approaches to XAS in solids

Peter Blaha*

Institute of Materials Chemistry, TU Vienna, Getreidemarkt 9, 1060 Vienna, Austria. *Correspondence e-mail: pbblaha@theochem.tuwien.ac.at

An overview of density-functional theory (DFT) is presented, which provides the basis for subsequent calculations of X-ray absorption spectra in solids. The advantages of DFT, such as the possibility of obtaining reliable atomic and electronic structures of even fairly complicated materials, as well as its disadvantages, namely the limited accuracy for certain systems containing strongly correlated electrons or having important van der Waals bonding effects and its restriction to being a ground-state theory, are discussed. The most important band-structure methods and the corresponding codes which are currently in use are then briefly reviewed, followed by an introduction to the basic theory of X-ray absorption near-edge structure. Finally, the most common methods used to include the important electron–hole interactions, namely a static core-hole approach with a supercell of the solution of the Bethe–Salpeter equation, are reviewed and selected examples from the literature using various codes are given.

1. Introduction

X-ray absorption spectroscopy (XAS) is a spectroscopy in which a core electron is excited into an empty conduction-band state. When we want to simulate XAS, we need to know the detailed electronic structure of the corresponding system. In principle, this can be obtained by solving the quantum-mechanical many-body problem of the interacting electrons and the corresponding nuclei using the time-dependent Schrödinger equation. Unfortunately, this is an intractable task except for a few very small systems and we have to make a couple of approximations.

First of all, we deal only with stationary states, i.e. we change to the time-independent Schrödinger equation; this is probably not a large approximation for standard XAS, but for some time-resolved spectroscopies this would break down. Next, we usually apply the Born–Oppenheimer approximation, which means that we decouple the ionic and electronic motions and treat the nuclei at fixed positions, or eventually move the atoms just as classical particles following Newton’s law. This approximation is also usually valid, since the electrons have a much smaller mass than the nuclei and instantaneously follow the ionic cores. In addition, most nuclei are so heavy that a quantum-mechanical treatment of their movement is not necessary. This leads to the well known Schrödinger equation \( H \Psi = E \Psi \), where the Hamiltonian \( H \) contains the kinetic energy of all electrons and the Coulomb interaction of all electrons and nuclei. Due to the Coulomb repulsion between the electrons, this problem cannot be factorized into individual electrons, but \( \Psi(r_1, r_2, \ldots, r_n) \) is a multidimensional function of \( 3n \) coordinates, with \( n \) being the number of electrons. This is again an intractable problem.
Traditional quantum chemistry tackles this problem using the Hartree–Fock (HF) method, which is exact for the exchange interaction (Pauli principle) and leads to effective one-electron equations, but completely neglects the Coulomb correlation between electrons of different spins. The latter can be incorporated by various approximate schemes such as Møller–Plesset perturbation theory (MP2), various coupled cluster schemes (CC) and complete active-space (CAS-SCF) or (full) configuration interaction (CI) methods (see, for example, Shavitt & Bartlett, 2009). Unfortunately, these methods are computationally very expensive, are limited to relatively few electrons and have found very little application to solids.

An alternative to the wavefunction-based methods described above is density-functional theory (DFT), and the rest of this chapter will deal with the use of DFT to simulate XAS spectra.

1.1. DFT and beyond

An efficient and manageable, but still accurate, scheme for solving the many-electron problem of a crystal (with nuclei at fixed positions) is Kohn–Sham theory within DFT (Hohenberg & Kohn, 1964; Kohn & Sham, 1995). The key quantity therein is the spin density \( \rho_\sigma (r) \), in terms of which the total energy is

\[
E_{\text{tot}}(\varrho_r, \varrho_i) = T_s(\varrho_r, \varrho_i) + E_{\text{xc}}(\varrho_r, \varrho_i) + E_{\text{ee}}(\varrho_r, \varrho_i) + E_{\text{NN}}(\varrho_r, \varrho_i) + E_{\text{xc}}(\varrho_r, \varrho_i) + E_{\text{NN}}.
\]  

(1)

\( T_s(\varrho_r, \varrho_i) \) is the kinetic energy (of non-interacting particles), \( E_{\text{ee}}(\varrho_r, \varrho_i) \) and \( E_{\text{NN}} \) represent the repulsive electron–electron and nuclear–nuclear Coulomb energies, respectively, \( E_{\text{xc}}(\varrho_r, \varrho_i) \) is the nuclear–electron attraction and \( E_{\text{xc}}(\varrho_r, \varrho_i) \) is the exchange–correlation energy. While the equation above is formally still exact and we can nowadays calculate almost all of these terms virtually exactly by numerical methods, we do not know the exact form of \( E_{\text{xc}}(\varrho_r, \varrho_i) \). This exchange–correlation energy should account for the approximations made in the kinetic energy (non-interacting particles) and cancel the self-interaction of an electron with itself in the classical \( E_{\text{xc}}(\varrho_r, \varrho_i) \) term. The most effective way to calculate the ground-state total energy is by means of the variational principle and the introduction of single-particle orbitals (wavefunctions) \( \chi_{ik}^{\sigma} \), which are used to construct the total spin density \( \rho_{\sigma}(r) = \sum_{ik} \rho_{ik}^{\sigma} |\chi_{ik}^{\sigma}(r)|^2 \). The occupation numbers \( \rho_{ik}^{\sigma} \) of state \( i \) vary between 0 and 1/\( w_k \) depending on \( \varepsilon_r \), where \( w_k \) is the symmetry-related weight of \( k \)-point \( k \). The variation of \( E_{\text{tot}} \) with respect to the density leads to the Kohn–Sham equations, which must be solved self-consistently in an iterative process:

\[
\left( -\frac{1}{2} \nabla^2 + V_{\text{xc}} + V_{\text{ee}} + V_{\text{NN}} \right) \chi_{ik}^{\sigma}(r) = \varepsilon_{ik}^{\sigma} \chi_{ik}^{\sigma}(r).
\]

(2)

Many different methods exist to solve the Kohn–Sham equations using different basis sets and under various approximations and they will be discussed later.

Besides the accurate solution of the Kohn–Sham equations, the approximations for \( E_{\text{xc}} \) and the corresponding exchange–correlation potential \( V_{\text{xc}}^{\sigma} \) determine the quality of the results. Perdew & Schmidt (2001) introduced ‘Jacob’s ladder into DFT heaven’, which classifies the various approximations according to accuracy but also complexity. It starts with the simple local density approximation (LDA), which originates from the homogeneous electron gas. The exchange–correlation energy at point \( r \) depends only on the local electron density \( \rho(r) \).

\[
E_{\text{xc}} = \int \rho(r) \varepsilon_{\text{xc}}[\rho(r)] \, dr,
\]

(3)

where \( \varepsilon_{\text{xc}} \) is the exchange–correlation energy density. \( E_{\text{xc}} \) can be split into an exchange and a correlation part, where \( E_x \) is simply given by

\[
E_x = -\frac{3}{4} \left( \frac{3}{\pi} \right)^{1/3} \int \rho(r)^{4/3} \, dr,
\]

(4)

while the correlation energy has been calculated by quantum Monte Carlo methods and parametrized by Perdew & Wang (1992).

The so-called generalized gradient approximations (GGAs) represent a clear improvement, where besides the local density its gradient \( \nabla \rho(r) \) also enters:

\[
E_{\text{xc}} = \int \rho(r) \varepsilon_{\text{xc}}[\rho(r), \nabla \rho(r)] \, dr.
\]

(5)

Various GGAs have been developed which differ in the enhancement factor \( F_{\text{xc}} \), which modifies the LDA \( \varepsilon_{\text{xc}} \). These GGAs thus have different accuracies for different properties (total energies or equilibrium distances) or types of atoms (light versus heavy atoms). For solids the most common GGA is the PBE-GGA of Perdew et al. (1996). The next rung, called meta-GGA, additionally employs the kinetic energy density \( \tau(r) \), and the latest versions (for example Sun et al., 2015) allow good equilibrium distances and reasonable cohesive energies to be obtained at the same time. Unfortunately, meta-GGAs lead to orbital-dependent potentials and are more difficult to implement and use than GGAs. Further steps use the occupied orbitals to introduce a certain amount of exact exchange, mixing the Hartree–Fock (HF) world with DFT. Various versions of these hybrid functionals became quite popular, in particular in quantum chemistry (B3LYP; Becke, 1993) and also in solids (HSE06; Krukau et al., 2006). They differ mainly in the amount of HF admixture and their short/long-range contributions and may lead to a significant improvement for semiconducting and insulating systems. However, because of limited correlation they fail badly for metals, predicting palladium and platinum to be ferromagnetic metals (Tran et al., 2012). The highest rung of Perdew’s ladder uses the unoccupied orbitals to calculate density-response functions and includes correlation effects on a higher level, as in the ACFDT-RPA approximation (Harl & Kresse, 2009).

Most solids can be simulated by GGA calculations with acceptable, although not perfect, accuracy (Tran et al., 2016). Exceptions are systems with correlated 3d or 4f electrons and very weakly bound van der Waals systems, where GGAs may break down and obtain an incorrect ground state or no bonding at all. For instance, the undoped cuprate YBa\textsubscript{2}Cu\textsubscript{3}O\textsubscript{6} becomes a nonmagnetic metal instead of an antiferromagnetic insulator when using plain LDA. For such systems one often employs an empirical Hubbard \( U \) (the energy cost of placing
two electrons at the same site) selectively to the correlated 3d or 4f electrons (Anisimov et al., 1993). This very cheap approximation is often used for 3d transition-metal or lanthanide compounds. For van der Waals systems one can use the RPA mentioned above, but cheaper (empirical) approximations also exist (Klimis̆ & Michaelides, 2012).

As already mentioned, the Kohn–Sham (KS) eigenvalues are formally not valid excitation energies and the corresponding one-electron orbitals (wavefunctions) only serve the purpose of yielding the correct total electron density when summed over all occupied single-particle states. In other words, formally it is NOT justified to use these orbitals and the corresponding eigenvalues to calculate any spectroscopic data. Nevertheless it is common practice to perform this at least as a first step, but it is important to understand the expected loss of accuracy from such a procedure. For example, it is well known that energy band gaps in semiconductors and insulators are typically underestimated by about 50%. This is not a shortcut of the LDA/GGA approximation, but even the exact KS gap typically underestimated by about 50%. This is not a shortcut corresponding eigenvalues to calculate any spectroscopic data.

words, formally it is NOT justified to use these orbitals and the summed over all occupied single-particle states. In other purpose of yielding the correct total electron density when accuracy from such a procedure. For example, it is well known that energy band gaps in semiconductors and insulators are typically underestimated by about 50%. This is not a shortcut of the LDA/GGA approximation, but even the exact KS gap misses the integer discontinuity $\Delta_{\text{xc}}$ (Kraisler & Kronik, 2014). On the other hand, experience has shown that the k dispersion of the quasi-particle energies as measured by angle-resolved photoemission is pretty well described by the KS eigenvalues, and this is particularly true for metallic states in the vicinity of the Fermi energy (Schäfer, 2005).

The formally exact extension of DFT to its time-dependent version (TDDFT; Runge & Gross, 1984) in principle allows investigation of the properties in the presence of time-dependent potentials (such as X-ray radiation) and the calculation of excitation spectra. However, in practice TDDFT also suffers from the necessary approximation of the now time-dependent $V_{\text{xc}}(r, t)$ and TDDFT has not yet provided a breakthrough in theoretical spectroscopies of solids.

For these reasons, hybrid DFT is sometimes used for the calculation of band gaps which are often enlarged (depending on the fraction of HF mixing) and close to the experimental values (Koller et al., 2013). A less time-consuming alternative is to use the Tran–Blaha modified Becke–Johnson (TB-mBJ) potential (Tran & Blaha, 2009), which can give band gaps close to experiment or expensive GW calculations at the computational cost of a standard GGA calculation.

In most cases the best estimates of the quasiparticle spectrum can be obtained from many-body perturbation theory and the GW method (Hedin, 1965). It approximates the nonlocal and frequency-dependent self-energy $\Sigma$ by the first term of an expansion into the single-particle Green’s function $G$ and the screened Coulomb interaction $W$. Since $GW$ is a perturbative scheme, it may have problems when the starting point (typically a DFT calculation) is very far from reality. In particular, in systems with strongly correlated electrons (3d electrons in transition-metal oxides or 4f electrons in lanthanides) all of the methods mentioned above may fail. For these classes of systems dynamical mean field theory (DMFT) has been developed as an alternative, where the problem is mapped onto a many-body local problem, called an impurity model (Held, 2007), where again DFT provides the basis of the method.

### 1.2. Periodic reciprocal-space methods versus direct-space cluster approaches

A piece of a solid consists of an almost ‘infinite number’ of atoms and in order to obtain its electronic structure one has to make approximations. One way to go is to approximate a solid by a finite (small) number of atoms and hope that the electronic structure of the central atoms of this cluster converge quickly with the size of the cluster. For some properties, where only localized electrons are important, surprisingly small clusters (the central atom and its nearest neighbours) may suffice, while in other cases efficient embedding schemes into point charges or the use of some model potentials are necessary and one has to increase the cluster size until convergence is reached. On the other hand, it is rather clear that such methods have limits in general; for example, the optimization of atomic positions is quite difficult, which is often the mandatory first step of any theoretical simulation, and properties depending on the long-range order cannot be obtained. Such an approach yields a discrete and finite number of molecular orbitals which eventually can be used for further interpretations.

The alternative is to model a perfectly periodic, infinite solid using a (fairly small) unit cell utilizing 3D periodic boundary conditions. Of course, also this is an approximation of the real world because all of our solids are finite (albeit with ~10^23 atoms) and have a surface and vacancies, impurities, disorder and other imperfections. In order to model ‘real’ solids, one creates a supercell (for computational reasons as small as possible, but to be realistic as large as possible) which should be chosen such that the perturbation that we want to model (the surface or the impurity etc.) is contained in the supercell and the periodic image of this perturbation does not influence the calculated properties. For instance, a surface is calculated by a periodic slab with typically 5–20 atomic layers, which are separated by an empty region (‘vacuum’) of 15–20 Å. Such a model ensures that the centre of the slab is already bulk-like, but the two surfaces (on the top and bottom of the slab) do not interact with each other. This leads to a ‘band-structure’ approach with eigenvalues classified by the reciprocal-space vector $k$ and the concept of density of states, and is in general a more universal approach for a complete description of a real solid since we can cover (approximately) short- and long-range order.

As some of the methods using finite clusters (Green’s function and muffin-tin methods) are described in other chapters, I will restrict myself in the following to periodic approaches.

### 1.3. Methods and basis sets

We cannot solve the KS equations analytically, but have to define a basis set and use the variational principle to find the best wavefunctions for the given basis (Martin, 2008). One can therefore classify the methods according to their basis set, and there are basically three different kinds of basis set in use today: linear combination of atomic orbitals (LCAO), plane-
wave (PW) pseudopotential (PP) and space-partition or augmentation methods.

LCAO-based methods are very common in molecular quantum chemistry, but several codes using a localized atomic basis set also exist for solids. The total wavefunction is constructed as $\Psi = \sum c_i \varphi_i$, with atomic orbitals $\varphi_i = R_i Y_{lm}$ centred at the position of the nuclei, where $Y_{lm}$ is a spherical harmonic function and the radial functions $R_i$ could be Gaussians $r^l \exp(-ar^2)$ (CRYSTAL, https://www.crystal.unito.it; Gaussian, https://gaussian.com/; CP2k, https://www.cp2k.org/; Slater orbitals $r^l \exp(-ar)$ (ADF-BAND, https://www.scm.com/) or simple numerical orbitals (DMOL3, https://www.psi.ch/isms/cnt-group/dmol3; FPLO, https://www.fplo.de/; FH1-aims, https://fh1-aims.org; SIESTA, https://departments.icmab.es/leem/siesta/). Obviously, the accuracy critically depends on the number (double-zeta etc.) and specific choice (localized basis, adaptive basis) of the basis functions, but for most approaches it is very clear that for heavier atoms such basis sets will become very large (1s to 6s basis for gold) and difficult to optimize, except when one uses these LCAO methods together with a PP approximation.

A ‘pseudopotential’ (PP) is constructed such that the strong Coulomb potential of the nuclei is replaced by a much smoother effective potential, such that the resulting wavefunctions for the valence electrons of free atoms (or ions) agree with all-electron orbitals beyond a certain ‘core radius’ but vary smoothly and are nodeless inside this sphere (near the nucleus). Numerous PP construction schemes exist and some of the older schemes (norm-conserving PP) might lead to substantial errors in real applications. However, modern PP and in particular the most recent PAW (projector augmented-wave) potentials (Lejaeghere et al., 2016) can reach the accuracy of all-electron methods for certain properties. PP calculations in solids are most commonly used together with a plane-wave (PW) basis set. PWs in principle form a complete basis set, but except for ultralight atoms (H, Li) this cannot be converged in practice with an all-electron potential. The benefits of a PW–PP method are the technical ‘simplicity’, the easy and efficient computer implementation (fast Fourier techniques) and a single and simple controllable convergence parameter (PW cutoff energy). Thus, numerous implementations exist such as ABINIT (https://www.abinit.org/), CASTEP (http://www.castep.org/), CPMD (https://www.cpmd.org/), GPAW (https://wiki.fysik.dtu.dk/gpaw/), Quantum ESPRESSO (http://www.quantum-espresso.org/) and VASP (https://www.vasp.at/).

Another way to overcome the convergence problems of PWs while avoiding the PP approximation is to ‘augment’ the PW basis in some way (for instance with localized Gaussian orbitals). In space-partitioning methods such as augmented plane-wave (APW) methods the unit cell is decomposed into non-overlapping atomic spheres (similar to as in PP methods) and an interstitial region. Outside the spheres the wavefunctions vary fairly smoothly and the PWs are an efficient and unbiased basis set. Inside the atomic spheres of atom $\alpha$ of radius $R_\text{MT}^\alpha$ a partial wave expansion, i.e., a linear combination of radial functions $W_{lm}$ times spherical harmonics $Y_{lm}$, is used together with some matching coefficients $A_{lm,k_n}$:

$$\varphi_{k_n}(r) = \frac{1}{(\text{Vol})^{1/2}} \exp(ik_n r), \quad r \in I. \quad (6)$$

$$\varphi_{k_n}(r) = \sum_{lm} A_{lm,k_n} W_{lm}(r) Y_{lm}(\hat{r}) \quad r \in R_\text{MT}^\alpha, \quad (7)$$

where $k_n$ is $k$ (a vector in the first Brillouin zone) plus $K_n$ (a reciprocal-lattice vector).

The Kohn–Sham orbitals are expanded into these APWs according to the linear variation method,

$$\psi_k = \sum_{k_n} c_{k_n} \varphi_{k_n}, \quad (8)$$

and the coefficients $c_{k_n}$ are determined by solving the resulting generalized eigenvalue problem. The convergence of this basis is controlled by a cutoff parameter $R_\text{MT}K_{\text{max}} = 5$–10 and typically about 100 APWs per atom are necessary.

In the original APW method (Slater, 1937), the $W_{lm}$ functions are the numerically exact solutions of the radial Schrödinger equation $\psi(r, \varepsilon)$ within the self-consistent potential. Unfortunately, these functions depend on the unknown energy of the corresponding eigenvalue $\varepsilon$, leading to a non-linear eigenvalue problem, which is very expensive to solve. Subsequently, various linearization schemes were put forward (LAPW, LAPW+LO, APW+lo etc.) and transformed the original APW method into a highly accurate all-electron method (Singh & Nordström, 2006). Naturally, the complicated nature of the basis set makes efficient implementations more difficult and the computational cost is usually somewhat larger than in PW–PP approaches, but one does not depend on the quality of predefined PPs, can easily treat all atoms of the periodic table and has access to properties where core electrons are very important. The most common codes that use an APW-type basis set are ELK (https://elk.sourceforge.io/), exciting (http://exciting.wikidot.com/), FLEUR (http://www.flapw.de) and WIEN2k (http://www.wien2k.at/; Blaha et al., 2001).

Other space-partitioning methods are based on the Korringa, Kohn and Rostoker (KKR) method (https://www.ebert.cup.uni-muenchen.de/sprkkr) or various versions of the linear muffin-tin orbital (LMTO) method.

2. XANES calculations and the dipole selection rule

X-ray absorption spectroscopy is a spectroscopy in which one uses monochromatic X-ray radiation of varying energy to excite core electrons in the sample. With increasing energy, at some point a sudden absorption onset (edge) occurs, which is generally given by the corresponding atom and core state and usually labelled as a $K (1s)$, $L_{1,2,3}$ $(2s, 2p_{1/2}, 2p_{3/2})$, $M$ or $N$ edge. We are interested here in the near-edge (XANES) region up to about 30–50 eV above the edge and we can interpret this spectroscopy as the excitation of a core electron into an (unoccupied) conduction-band state, which also naturally defines the edge onset as the energy necessary to bring the core electron into the lowest unoccupied state. Using time-dependent perturbation theory, one can describe the absorption

$$\varphi_{k_n}(r) = \frac{1}{(\text{Vol})^{1/2}} \exp(ik_n r), \quad r \in I. \quad (6)$$

$$\varphi_{k_n}(r) = \sum_{lm} A_{lm,k_n} W_{lm}(r) Y_{lm}(\hat{r}) \quad r \in R_\text{MT}^\alpha, \quad (7)$$

where $k_n$ is $k$ (a vector in the first Brillouin zone) plus $K_n$ (a reciprocal-lattice vector).

The Kohn–Sham orbitals are expanded into these APWs according to the linear variation method,

$$\psi_k = \sum_{k_n} c_{k_n} \varphi_{k_n}, \quad (8)$$

and the coefficients $c_{k_n}$ are determined by solving the resulting generalized eigenvalue problem. The convergence of this basis is controlled by a cutoff parameter $R_\text{MT}K_{\text{max}} = 5$–10 and typically about 100 APWs per atom are necessary.

In the original APW method (Slater, 1937), the $W_{lm}$ functions are the numerically exact solutions of the radial Schrödinger equation $\psi(r, \varepsilon)$ within the self-consistent potential. Unfortunately, these functions depend on the unknown energy of the corresponding eigenvalue $\varepsilon$, leading to a non-linear eigenvalue problem, which is very expensive to solve. Subsequently, various linearization schemes were put forward (LAPW, LAPW+LO, APW+lo etc.) and transformed the original APW method into a highly accurate all-electron method (Singh & Nordström, 2006). Naturally, the complicated nature of the basis set makes efficient implementations more difficult and the computational cost is usually somewhat larger than in PW–PP approaches, but one does not depend on the quality of predefined PPs, can easily treat all atoms of the periodic table and has access to properties where core electrons are very important. The most common codes that use an APW-type basis set are ELK (https://elk.sourceforge.io/), exciting (http://exciting.wikidot.com/), FLEUR (http://www.flapw.de) and WIEN2k (http://www.wien2k.at/; Blaha et al., 2001).

Other space-partitioning methods are based on the Korringa, Kohn and Rostoker (KKR) method (https://www.ebert.cup.uni-muenchen.de/sprkkr) or various versions of the linear muffin-tin orbital (LMTO) method.

2. XANES calculations and the dipole selection rule

X-ray absorption spectroscopy is a spectroscopy in which one uses monochromatic X-ray radiation of varying energy to excite core electrons in the sample. With increasing energy, at some point a sudden absorption onset (edge) occurs, which is given by the corresponding atom and core state and usually labelled as a $K (1s)$, $L_{1,2,3}$ $(2s, 2p_{1/2}, 2p_{3/2})$, $M$ or $N$ edge. We are interested here in the near-edge (XANES) region up to about 30–50 eV above the edge and we can interpret this spectroscopy as the excitation of a core electron into an (unoccupied) conduction-band state, which also naturally defines the edge onset as the energy necessary to bring the core electron into the lowest unoccupied state. Using time-dependent perturbation theory, one can describe the absorption
cross section of the incident light with polarization $\varepsilon$ and frequency $\omega$ according to Fermi’s golden rule as

$$
\sigma(\omega) = \frac{4\pi^2}{\omega} \sum_l |\langle f | \varepsilon \cdot \mathbf{p} \exp(ikr) | i \rangle|^2 \delta(E_i - E_f + \hbar \omega),
$$

(9)

where $|i\rangle$, $|f\rangle$, $E_i$ and $E_f$ denote the initial (core) and final (conduction band) states and their energies, $k$ defines the wavenumber of the incident photon and $p$ is the momentum operator of the electron.

Since the momentum corresponding to the X-rays is quite small (and the corresponding wavelength is quite large compared with the intra-atomic dimensions), one can expand the exponential into a power series and truncate after the first term,

$$
\exp(ikr) = 1 + ikr - \frac{1}{2} k^2 r^2 + \ldots \approx 1.
$$

(10)

This is called the dipole approximation, which is usually reasonably well justified for excitation energies of up to a few keV and simplifies the matrix elements to a momentum matrix element, which can be transformed into the dipole matrix element by a commutator rule,

$$
\mathbf{e}(|f|\mathbf{p}|i\rangle) = \frac{im}{\hbar} (E_i - E_f) \mathbf{e}(|f|\mathbf{r}|i\rangle),
$$

(11)

The initial state $|i\rangle$ can be written as $\psi_i = R_l^f Y_{lm}$ and has a particular angular momentum value $l$. $\mathbf{r}$ can be expanded into spherical harmonics $Y_{lm}$ with $l = 1$ and $|f\rangle$ into $\psi_f = \sum_l R_l^r Y_{lm}$. The above integral can then be decomposed into a radial and an angular part. For the latter we can use the properties of ‘Gaunt numbers’ (or Clebsch–Gordan coefficients), which together with the required parity constraint yield the dipole selection rules that the $l$ value of the final states must differ from the initial core state by $\pm 1$ (while $\Delta m = 0, \pm 1$). We are then left with radial matrix elements $\int r^2 R_l^f R_l^r d\mathbf{r}$, which due to the locality of the core state restrict the final states to be located around the specific element where the XAS spectrum has been measured. Thus, XANES probes the unoccupied orbitals (conduction bands) of $\Delta l = \pm 1$ character of the probe atom and provides complementary information to X-ray emission spectroscopy, which gives information about the same states but from the occupied part of the electronic structure (valence bands).

Of course, it is not necessary to restrict the interpretation to the dipole approximation; quadrupolar contributions for high-energy edges can also be considered (Bună & Calandra, 2013).

### 2.1. Electron–hole interactions in single-particle theory

Using the concepts described above, it should now be possible to derive a scheme which allows accurate simulations of XANES spectra. One would calculate the electronic (ground-state) structure of the material using an accurate electronic structure method based on DFT, and using the dipole approximation the corresponding unoccupied partial density of states (DOS) weighted with the specific radial matrix elements gives the XANES spectrum. While such an approach may work for good metals with very efficient screening, it turns out that for semiconductors and insulators the spectra are in very poor agreement with experiment: in general, the intensity of the first peaks is too low and too broad. This has nothing to do with the shortcomings of DFT mentioned previously; using a DOS from a GW calculation would also give miserable results. It can be understood by the final-state rule (von Barth & Grossmann, 1979), which says that the electronic structure of the final state determines the shape of a spectrum. Our final state has a core hole on one of the atoms and an additional electron in the conduction band. Obviously, the missing core electron leads to a much smaller screening of the nuclear charge by the core (a larger ‘effective’ nuclear charge) and the valence/conduction-band electrons on that atom are more strongly attracted by the nucleus, leading to a downward shift in energy and localization in a much sharper peak. This excitonic effect (electron–hole interaction) can be simulated approximately by taking one electron out of the core on one of the probed atoms and adding this electron to the valence electrons. To avoid ‘every’ atom in the crystal being excited, we need to create a supercell (typically 32–128 atoms per cell) and then take out the core electron on one of the atoms only. The electrons of all of the other atoms in the supercell can then contribute to the screening. In solids this approach suffers from at least two problems: the size of the supercell is always limited and we have no influence on where to put the extra electron, i.e. the electron goes into the first empty states, even if they are of very different character to those in the experimental excitation. Take a transition-metal oxide such as NiO: for an O K edge the electron should go into O $p$ states, but in the supercell approach it would fill an empty Ni 3$d$ orbital. Thus, the screening of the core hole might be underestimated (in particular for metals) and thus lead to approaches of taking only a partial core hole (Luitz et al., 2001; Mizoguchi et al., 2004). Note that Slater’s transition-state theory (Slater, 1972), which is sometimes quoted as justification for using half a core hole, has nothing to do with XANES but is only justified as an approximation to core electron binding energies in X-ray photoelectron spectroscopy, where the electron would leave the solid (ionize). Instead of explicitly adding an electron to the conduction band, one can also add a constant charge background to make the supercell neutral again. This may eventually be a good approximation anyway when the excited electron sits in a very delocalized state. Besides the problems mentioned above and the fact that DFT eigenvalues should in principle not be taken as excitation energies, the ‘final-state rule’ is also just a ‘rule’ and not a fundamental principle. It assumes that a static, fully relaxed picture is sufficient, neglects any time (frequency) dependency of the excitation process and assumes that DFT can describe the exchange–correlation effects.

Although many question marks to such an approach have been raised here, experience shows that it works surprisingly well in many cases such as K edges (mainly of light elements); L or M edges of heavier elements have also been described reasonably well (see Section 2.1.1), and much more involved BSE calculations yield quite similar results. For these reasons this method has been implemented into several different...
codes and such simulations are possible even with pseudopotentials, which are generated with a (fractional) core hole.

2.1.1. Applications of core-hole single-particle calculations. In this section I will describe a couple of examples and applications, preferably using different band-structure codes. As mentioned before, most applications deal with K edges (preferentially of light elements), but L2,3 and higher edges have also been investigated using various codes. Note that this is by no means a full list of published calculations, just a very small selection.

In the early days of electronic structure calculations the available computer power restricted applications to very simple systems, mainly elemental metals or simple intermetallic compounds. For instance, in a seminal paper Müller & Wilkins (1984) formulated the basic theory and used a modified non-self-consistent APW method to calculate various edges of f.c.c. palladium. In these early applications to metals the ground-state electronic structure was usually used to interpret the spectra and excitonic effects were neglected (Muller et al., 1998). In 2001 Luitz and coworkers investigated the Cu L3 edge of f.c.c. copper and showed in particular that for such a metallic system a partial core hole (but not a full hole) also leads to better agreement with experiment (Luitz et al., 2001). Scott et al. (2001) investigated the C K edges of various transition-metal carbides, comparing multiple-scattering cluster calculations with WIEN97 band-structure calculations and experiment. Around this time, when band-structure codes and the available hardware became more mature, the closely related theories and the corresponding codes for EELS were also developed (Nelhiebel et al., 1999; Hébert, 2007). Subsequently, the effect of core holes as well as of anisotropy on single crystals or oriented surfaces was investigated in many papers, for example Mizoguchi et al. (2004) (using an LCAO band-structure code) and Stroev et al. (2005). Calculations became possible for quite complicated systems such as single monolayers of h-BN on various transition-metal (111) surfaces (Laskowski et al., 2009) or the O K and Nb M5 edges of Nb2O5(OH) (Khan et al., 2016). For strongly correlated materials such as NiO it turned out that better agreement with experiment can be achieved when the modified Becke–Johnson (Tran & Blaha, 2009) potential was used instead of standard GGA (or GGA+U) calculations (Hetaba et al., 2012).

As an illustration of how different exchange–correlation potentials and different approximations for the excitonic effects influence the resulting spectrum, Fig. 1 shows the C K edge of the C60 molecule. It was simulated using GGA and LDA, and for the full core-hole calculation a surprisingly large difference between the corresponding spectra can be seen. However, this is a quite indirect effect because it comes solely from the fact that geometry-optimized structures were used in these simulations and LDA/GGA led to smaller/larger C–C bond distances by about 0.01 Å. These small changes in bond length are responsible for the differences in the spectra, while the direct influence of LDA or GGA on the spectrum for fixed geometry is negligible. This example also nicely demonstrates the enhancement of the first peak with respect to higher energy features and some additional peak shifts in energy when a half or full core hole is taken into account.

In more recent years, pseudopotential codes also started to be used to investigate XANES spectra using pseudopotentials with core holes and a reconstruction of the all-electron wavefunction with the PAW approach. Gao et al. (2009) used the CASTEP PP code to calculate various K edges and a few L edges. They exploited the PAW approach and compared their results with published all-electron WIEN2k calculations (Scott et al., 2001; Hébert, 2007). Ljungberg et al. (2011) implemented this approach into the GPAW code. Bunau & Calandra (2013) used the PAW method as implemented in Quantum ESPRESSO to calculate Cu L2,3 edges in f.c.c. copper and Cu2O and S L2,3 edges in 2H MoS2. In agreement with Luitz et al. (2001) they found that for metallic copper no core hole or only a small core hole should be used, while for the other examples a core hole is essential to reproduce the experiment. An interesting side remark is that for S L edges there is also a significant contribution from S s states in addition to S d states, which is very different from transition-metal L edges, where the d contribution clearly dominates.

2.1.2. Limits of single-particle approaches: 3d transition-metal L2,3 edges. Unfortunately, there are a couple of cases in which the simple, DFT-based single-particle approach described above does not work so well. These are cases where strong correlation effects are present either between the core electrons or between the valence/conduction-band electrons. In the latter case this is not a great surprise, because DFT cannot even properly describe the ground-state electronic structure of some correlated transition-metal compounds such as the high-Tc cuprates. However, even for nominal d0 transition-metal compounds such as CaF2, SrTiO3 or TiO2 many details and in particular the nonstatistical branching ratio between the L2 and L3 parts of the spectrum cannot be explained at all by single-particle theories.
2.2. Electron–hole interactions beyond the single-particle approximation

2.2.1. Bethe–Salpeter equation (BSE). The Bethe–Salpeter equation (BSE) represents a many-body approach for the description of a general two-particle problem and has been used from particle physics to solid-state physics (Onida et al., 2002). Here, we use it to describe the excitonic effects, i.e. the bound state of an electron–hole pair. After a couple of simplifications (Tamm–Dancoff) the BSE represents a two-particle Schrödinger-like (eigenvalue) equation for the electron–hole (e-h) pair,

$$\sum_{h,e,k} H^{e-h}_{e,k} = E^h A^h_{e,k}.$$  (12)

where $H^{e-h}$ represents the effective e-h Hamiltonian and $h, e$ and $k$ represent the hole (core) and electron (conduction band) state at wavevector $k$. The eigenvalues $E^h$ are the exciton energies and the eigenvectors $A^h$ are interpreted as coupling coefficients between the hole and the electron. The ground-state DFT orbitals for core and conduction-band states form the basis for this problem, leading to large basis sets (matrix size) of dimensions $N_{core} \times N_{cond} \times k$ points. The e-h Hamiltonian consists of three terms: a diagonal term $H^{diag}$, represented by the DFT eigenvalue differences ($\epsilon_{ek} - \epsilon_{ek}$), an attractive, statically, but nonlocally screened Coulomb term $H^{int}$ between the electron and the hole, and a corresponding repulsive, unscreened exchange term $H^x$. In particular, the calculation of the nonlocal dielectric function used to screen the Coulomb interaction needs special care to ensure convergence due to the presence of the very localized core states. The resulting energies and eigenvectors $A^h$ can be used to calculate the imaginary part of the dielectric tensor $\varepsilon_2$.

$$\varepsilon_2^{ex}(\omega) = \sum_{k} \sum_{h,k} A^h_{e,k} \left( \epsilon_{ek} - i \nabla_{ek} \right)^2 \delta(E^h - \hbar \omega).$$  (13)

Obviously, for $K$ edges of light elements there is only one shallow core state and the calculations are thus not too expensive for an all-electron code and can be performed by standard BSE programs (which are usually designed for excitons in semiconductors). Olovsson et al. (2009) applied this to the Li $K$ edges of some lithium halides and the Mg $L_{2,3}$ edge in MgO and compared the results with supercell core-hole calculations. Even when the authors emphasized the differences between these two methods, overall the results differ only in small details, justifying the use of the core-hole approximation. In fact, another application to Be $K$ edges (Olovsson et al., 2013) even showed that for BeO the core-hole approach agrees even better with experiment than the BSE calculations. However, for $L_{2,3}$ spectra it is essentially to use all six core states and in addition it is absolutely necessary to include spin–orbit coupling, because the size of the spin–orbit splitting of the $2p_{1/2}$ and $2p_{3/2}$ states determines the branching ratio, as will be shown below.

We have applied this method to several transition-metal compounds (Laskowski et al., 2009) and I illustrate the results for rutile TiO$_2$ below. The experimental spectrum (Fig. 2) shows clearly separated $L_3$ and $L_2$ subspectra and each spectrum is split into two peaks by crystal field effects. However, the first hints towards a more complicated explanation arise from the fact that the intensity ratio between $L_3$ and $L_2$ is not 2:1 (similarly, the crystal field split peaks also do not show the 3:2 ratio expected from the 3$d$ $t_{2g}$ and $e_g$ degeneracy) and the line shapes of the $L_3$ and $L_2$ subspectra are very different. The spectrum calculated from ground-state DFT has very little in common with experiment and the spectral features are very broad since all excitonic features are missing. Calculations with a core hole show the expected localization into sharper peaks, but of course the $L_3/L_2$ intensity ratio is 2:1 as expected. Details of the line shape cannot be reproduced and while the peak originating from the $t_{2g}$ states has a reasonable width, the $e_g$ peak is much too narrow and does not show the characteristic shoulder on the left side. In addition, the size of the spin–orbit splitting as well as the crystal field splitting seems to be overestimated in theory. The actual amount of core hole has only a very small effect in this case. On the other hand, our fully relativistic BSE calculations agree very well with experiment. Obviously, the intensity ratio is about right, the line shapes of the $L_2$ and $L_3$ edges are very different and both spin–orbit and crystal field splitting is reduced. The effects can be explained as follows. The spin–orbit splitting of the initial Ti $2p$ states is relatively small (around 5 eV), and thus even when the Ti $2p_{1/2}$ character dominates in an exciton there is also a small admixture of $2p_{3/2}$ and vice versa. In the calculation of $\varepsilon_2$ the matrix elements are squared and thus the direct small contribution nearly vanishes, but the mixed cross terms can be large and in some cases can have opposite sign, reducing or enhancing the intensity of the ‘$L_3$’ or ‘$L_2$’ peaks. Similar interference terms are responsible for the unexpected ‘$e_g$’ and ‘$t_{2g}$’ peaks. The excitonic binding energies vary from 0.3 to 4 eV and thus change the ‘spin–orbit’ or ‘crystal field’ splitting that is actually observed.

To demonstrate the importance of the interference terms between $2p_{1/2}$ and $2p_{3/2}$ we also manually increased the spin–orbit splitting of these states. With increasing splitting, the
intensity ratio moves towards the expected 2:1 ratio according to the degeneracy of these states and interference effects reduce drastically.

It should be noted that the BSE may probably describe the electron–hole interactions very well, but naturally it does not account for the deficiencies of DFT, as manifested in an insufficient description of correlation effects within the 3d shell. Thus, for TM compounds with partly filled 3d shells such as MnO or CoO the BSE results are much better than static core-hole calculations, but are still not in perfect agreement with experiment.

3. Summary

Density-functional theory provides a powerful framework to study the geometric and electronic structure of solids, which is the basis for subsequent simulations of XANES. Several codes and band-structure methods now exist which can provide a reliable electronic structure. Usually, one cannot use the ground-state electronic structure obtained from DFT for the interpretation of XAS, but it is necessary to consider the electron–hole interaction in a more or less sophisticated way.

In the simple ‘core-hole’ approach a static core hole on a single probe atom is introduced within a large supercell. This modifies the corresponding electronic structure and shifts the spectral weight from higher energy down into the first peak, and the resulting XAS spectra contain localization effects close to the often observed sharp edge onsets in the experimental spectra. Sometimes one can obtain better agreement with experiment by variation of the size of the core hole, as the static screening in such calculations may not be perfect in all cases (metals).

On the other hand, one can also use these DFT orbitals as a basis for highly sophisticated perturbation theory for the description of the electron–hole interaction. The resulting BSE approach is expensive, but also allows XANES simulations in cases where the single-particle DFT core-hole approach fails because of important interference terms, which can drastically change the intensities due to a mixing of either the initial states (for example the $p_{1/2}$ and $p_{3/2}$ states in $L_{2,3}$ spectra) or the final states (for example mixing of the $e_g$ and $t_{2g}$ states). In addition, the resulting excitons may have very different binding energies, modifying some peak splittings.

Funding information

I acknowledge support by the SFB F41 (Vicom) of the Austrian Science Fund (FWF).

References


