

Multi-code Benchmark on Simulated Ti K-edge X-ray Absorption Spectra of Ti-O Compounds

Fanchen Meng,¹ Benedikt Maurer,² Fabian Peschel,² Sencer Selcuk,¹ Mark Hybertsen,¹
Xiaohui Qu,¹ Christian Vorwerk,^{3,*} Claudia Draxl,^{2,†} John Vinson,^{4,‡} and Deyu Lu^{1,§}

¹*Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973, United States*

²*Institut für Physik and IRIS Adlershof, Humboldt-Universität zu Berlin, Berlin Germany*

³*Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States*

⁴*Material Measurement Laboratory, National Institute of Standards
and Technology, Gaithersburg, Maryland 20899, United States*

(Dated: January 31, 2024)

X-ray absorption spectroscopy (XAS) is an element-specific materials characterization technique that is sensitive to structural and electronic properties. First-principles simulated XAS has been widely used as a powerful tool to interpret experimental spectra and draw physical insights. Recently, there has also been growing interest in building computational XAS databases to enable data analytics and machine learning applications. However, there are non-trivial differences among commonly used XAS simulation codes, both in underlying theoretical formalism and in technical implementation. Reliable and reproducible computational XAS databases require systematic benchmark studies. In this work, we benchmarked Ti K-edge XAS simulations of ten representative Ti-O binary compounds, which we refer to as the Ti-O-10 dataset, using three state-of-the-art codes: XSPECTRA, OCEAN and **exciting**. We systematically studied the convergence behavior with respect to the input parameters and developed a workflow to automate and standardize the calculations to ensure converged spectra. Our benchmark comparison considers a 35 eV spectral range starting from the K-edge onset, representative of widely used near-edge spectra. Quantitative comparison over this range is based on Spearman’s rank correlation score (r_{sp}). Our results show: (1) the two Bethe-Salpeter equation (BSE) codes (OCEAN and **exciting**) have excellent agreement with an average r_{sp} of 0.998; (2) good agreement is obtained between the core-hole potential code (XSPECTRA) and BSE codes (OCEAN and **exciting**) with an average r_{sp} of 0.990, and this smaller r_{sp} reflects the noticeable differences in the main edge spectral shape that can be primarily attributed to the difference in the strength of the screened core-hole potential; (3) simulations from both methods overall reproduce well the main experimental spectral features of rutile and anatase, and the different treatments of the screened core-hole potential have visible impact on pre-edge intensities and the peak ratio of the main edge; (4) there exist moderate differences in the relative edge alignment of the three codes with a standard deviation of about 0.2 eV, which arise from multiple contributions including the frozen core approximation, final state effects, and different approximations used for the self-energy correction. Our benchmark study provides important standards for first-principles XAS simulations with broad impact in data-driven XAS analysis.

I. INTRODUCTION

X-ray absorption spectroscopy (XAS) probes the excitations that promote deeply bound core electrons into the unoccupied states. Due to the energy separation between core levels of different elements and the small spatial extent of core orbitals, XAS is element-specific and sensitive to the local chemical environment around the absorbing atoms. As a first approximation the spectra reflect details of the unoccupied density of states localized on the site of the absorbing atom, although spectra can be heavily modified by the presence of the positively charged core hole. Particularly, the low-lying excitation region extending about 30 eV above the onset of absorption from a specific core level, known as x-ray absorption near-edge structure (XANES)¹, has been routinely used to extract the local chemical characteristics of the absorption site, such as local symmetry, type of hybridization, charge state, spin state, and bond distortions². Physically, local structural and chemical characteristics that modify the local unoccupied orbitals can result in measurable changes to the XANES.

The rich information contained in XANES leads to its extensive use for materials characterization in condensed matter physics,^{3,4} materials science,⁵ chemistry,⁶ and biology.⁷ Despite these broad applications, interpreting XANES spectra is non-trivial because the observed spectral features represent a complex convolution of the atomic and electronic structures. Standard XANES analysis relies on fitting the measured spectra with empirical fingerprints collected from experimental standards with known chemical formulae and atomic arrangements.² However, the empirical fingerprint approach is limited by the chemical and configuration space spanned by available experimental standards. Typically the available data for a target element are restricted to simple crystals and small molecules. Therefore, it remains challenging to analyze XANES spectra of complex materials (e.g., surfaces, interfaces, nanostructures and amorphous materials) and structural evolution during physical and chemical processes (e.g., phase transitions and chemical reactions).

First-principles XANES simulations have made remarkable progress in the past few decades. Due to their predictive power, first-principles calculations provide a concrete connection between XANES spectra and the underlying atomic and electronic structures. This makes them a powerful tool for spectral interpretation.^{1,8} In many cases, simulated XANES spectra yield excellent agreement with measurements on key spectral features, e.g., number of peaks, peak positions and peak height ratios, thus supporting quantitative assignment of the spectra.^{9–11} Recently, there is a growing interest in data-driven XANES analysis leveraging machine learning (ML).^{12,13} First-principles simulations of XANES play a significant role in this new paradigm. In practice, the first step in a data-driven approach is to construct a database containing atomic structures and the corresponding simulated XANES spectra for either curated configuration spaces (e.g., metal clusters¹⁴, bi-metallic clusters^{15,16}, metal oxide clusters¹⁷, small molecules with structural distortions¹⁸, transition metal compounds¹⁹, amorphous materials^{20,21}, catalysts^{22,23}, or interfaces²⁴) or chemical spaces derived from public structure databases (e.g., a wide range of small molecules at their equilibrium structures^{25,26} or transition metal oxides^{27–31}). Multiple data-analytics methods can be applied subsequently, including computational spectral fingerprints^{18,20,21,31–33}, ML surrogates to predict spectrum from structure (i.e., the forward problem)^{25,30,34,35} and ML classifiers to extract physical descriptors from spectra (i.e., the inverse problem)^{14–19,22–24,26–29}. Bridging from models trained on simulated datasets to experiment can be challenging. Although promising results were reported in special cases when ML classifiers trained on synthetic data were directly applied to experimental data^{14–16,18}, in general the systematic error between theory and experiment needs to be carefully addressed.

In comparison to the traditional first-principles modeling approach, the data-driven approach has the advantage that the generated data are findable, accessible, interoperable, reusable (known as the FAIR data principles) and expandable. This allows researchers to identify important trends from the full energy range of XANES of diverse materials spaces in an unbiased way. As a result, the data-driven approach is playing a more and more important role in XANES analysis, especially for high-throughput structure identification and real-time spectral interpretation pertinent to autonomous experimentation. A critical first step in this paradigm is building high-quality XANES databases that expand a wide energy range of 30 to 40 eV, in contrast to many existing studies that focus on features in a narrow energy range. Despite several existing simulated XANES databases^{36–38}, many more need to be constructed to cover the vast chemical space and materials diversity in practical applications. However, there are several caveats in the simulated XANES database construction.

First, one needs to choose from many available XANES simulation codes. Density functional theory (DFT)-based or many-body perturbation theory-based XANES simulation methods are a favorable choice due to the interpretability associated with band structure theory. It is also feasible to use these methods to build a database of XANES spectra. Because of different theoretical methods (e.g., treatment of the core-hole final-state effects) and numerical implementations (e.g., basis sets) employed, it is important to systematically understand the performance of different codes on representative benchmark systems both among themselves and against experiment. In addition, the continued improvement of x-ray measurements (e.g., increased resolution and decreased noise and sample damage) highlights the need for quantitative theoretical predictions. Systematic and quantitative comparisons between codes are necessary

to elucidate the nature of discrepancies, which can originate from different approximations in the theory and different numerical implementations^{39,40}. Such systematic comparison has, to our knowledge, not previously been attempted, though comparisons with limited scope have been carried out for the O K edge^{10,41}, C K edge⁴², Mg K edge⁴¹ and Ti K edge^{5,31}. In part, the lack of such benchmark studies is due to the complexity of accounting for different approximations and settings in multiple codes or methods, which can be non-trivial even for domain experts, making it difficult to establish reliable comparisons. The computational cost associated with multiple codes and multiple choices of parameters is another limiting factor.

Second, the quality of the simulated XANES spectra strongly depends on the numerical convergence of multiple key parameters. While some parameters are generic to the excited state calculations, many are code specific, such as the choice of pseudopotentials. We emphasize that spectral data quality is essential to data-driven XANES analysis, because it directly affects the performance of the downstream data analytics applications. In other words, robust control for parameter choice is essential to the utility of spectral databases, in particular to avoid misleading or unphysical trends in subsequent analysis. Establishing the know-how to achieve the convergence in multiple codes requires a collective effort from multiple research groups. Significant benchmark studies have been performed for ground state DFT codes⁴³ and *GW* codes^{44,45}. However, similar cross-code benchmark studies are still needed in the computational x-ray spectroscopy field.

Last, in order to generate large spectral databases from high-throughput calculations, it is mandatory to have an automated workflow that requires little human oversight. This workflow needs to provide system-independent default parameters and, more importantly, educated guesses of system-dependent parameters based on well-established trends from benchmark studies. Such a workflow not only provides a common standard to obtain reliable XANES simulation data, but also lowers the barrier of XANES simulation for non-experts. In addition, a well tested workflow avoids repetitive work in determining input parameters on similar systems and thus saves research time and computational resources. Due to the complexities associated with the first two issues discussed above, a multi-code XANES workflow does not exist.

A meaningful benchmark against experimental data requires a set of highly reliable measurements with quantified estimations of errors and uncertainties as well as a quantification of the often uncontrolled approximations used in the theoretical approaches. Comparison to measured data from the literature is confounded by differences between published measurements arising from such factors as calibration, instrumental resolution, sample validation and preparation, self-absorption corrections, and other issues. Theoretical approaches adopt a number of widely used approximations, the effects of which can be large compared to the differences between codes as well as significant for comparison to measured spectral features. Case studies illustrate the impact of many common approximations, including neglect of the role of vibrations⁴⁶⁻⁵⁰ and different technical approximations in first-principles computational approaches⁵¹⁻⁵⁶. In view of these well-known issues, we have phased our study. In this paper, we focus on first establishing the level of agreement and uniformity between standard XANES simulation methods with only a brief comparison with experiment. A benchmark against a broader set of measured data will be the focus of a future study.

Specifically, in this study, we compare simulated XANES spectra among multiple codes that implement state-of-the-art first-principles theoretical approaches. We carry out a quantitative comparison among three popular codes, OCEAN^{57,58}, `exciting`⁵⁹⁻⁶¹, and XSPECTRA^{62,63}, for calculating XANES spectra. These codes follow two differing theoretical approaches and rely on different implementations and approximations, as will be discussed in the next section. We benchmark Ti K-edge XANES for ten representative Ti_xO_y compounds, which we refer to as the Ti-O-10 dataset. We have developed scripts to automatically generate consistent input files from the crystal structure, establish defaults for general input parameters, and carry out convergence tests on several key parameters. We find overall good agreement in the calculated spectra among the three codes over the spectral range typically considered in XANES, despite differences in theoretical approximations and technical implementations. We identify the main reasons for the differences that are observed, and we discuss the comparison to experiment for the widely studied examples of TiO_2 in the rutile and anatase structure.

II. THEORETICAL FRAMEWORK

A. X-ray Absorption Spectroscopy

The x-ray absorption cross section can be calculated from Fermi's golden rule according to⁸

$$\sigma(\omega) = 4\pi^2\alpha\omega \sum_f |M_{0,f}|^2 \delta(E_f - E_0 - \omega), \quad (1)$$

where E_0 and E_f are the total energies of the many-body initial state $|\Psi_0\rangle$ and final state $|\Psi_f\rangle$. α is the fine structure constant, and ω is the x-ray energy. Unless otherwise specified, we use atomic units throughout the rest

of the paper. $M_{0,f} = \langle \Psi_f | \hat{O} | \Psi_0 \rangle$ is the transition matrix element with \hat{O} the transition operator. Under the electric field of the photon beam, the dipole and quadrupole terms are given by $\hat{O} = \mathbf{e} \cdot \mathbf{r} + i/2(\mathbf{e} \cdot \mathbf{r})(\mathbf{q} \cdot \mathbf{r})$, where \mathbf{e} and \mathbf{q} are the polarization vector and the wave vector of the photon beam, and \mathbf{r} is the position of the electron. Within the scope of this work, we will only consider the dominating contribution from the dipole term except where noted below. Evaluating XANES spectra at the ground state or the final state Hamiltonian are referred to as the initial or final state rule, respectively. Under the initial state rule, XANES spectra correspond to excitation of independent particles, while under the final state rule, the many-body effects of interacting particles (such as core hole screening) are included.

Practical calculations build on specific approximations. In practice, approximations need to be made to evaluate Eq. 1. Here we start with the single particle picture based on the Kohn-Sham density-functional theory (DFT) and discuss different approximations to treat electron correlation effects. Under the final state rule, we consider two different treatments of the core-hole final state effects. In the core-hole pseudopotential (CHP) method^{4,62,63}, we use a DFT-based approach and consider the static response to a core hole in the final state. In the linear response method^{57,58,60,64}, we use the many-body perturbation framework to describe the core electron excitation.

In the CHP method, the core hole on the absorber atom is treated explicitly by a core-hole pseudopotential, and the final state is solved self-consistently by allowing valence electrons to relax. The dipole contribution to Eq. 1 can be approximated by

$$\begin{aligned} \sigma_{CHP}(\omega) &= 4\pi^2 \alpha \omega \sum_f \left| \mathbf{e} \cdot \langle \tilde{\psi}_f | \mathbf{r} | \psi_\alpha \rangle \right|^2 \delta(\tilde{\epsilon}_f - \epsilon_\alpha - \omega) \\ &= -4\pi \alpha \omega \text{Im} \left[\mathbf{e} \cdot \langle \psi_\alpha | \mathbf{r} (\omega - \tilde{H}_{DFT} + i\eta)^{-1} \mathbf{r} | \psi_\alpha \rangle \cdot \mathbf{e} \right], \end{aligned} \quad (2)$$

where $|\psi_\alpha\rangle$ is the core-hole state before the excitation and $|\tilde{\psi}_f\rangle$ are empty states at the presence of the core hole, with ϵ_α and $\tilde{\epsilon}_f$ the corresponding Kohn-Sham energy levels.

Alternatively, within many-body perturbation theory, neutral electronic excitations are described under the *GW-Bethe-Salpeter Equation* (BSE) framework, as a post-DFT treatment. Under the *GW* approximation, the DFT exchange-correlation potential is replaced by the energy-dependent self-energy operator, which contains a screened exchange term and a Coulomb hole term. Except for a few cases^{11,52,65}, the self-energy correction to empty states is often neglected in practical XANES calculations and existing studies primarily focus on correcting the core-hole energy level^{66,67}. In this work, self-energy corrections to the DFT valence or conduction orbitals are not considered, and the correlation effects between the excited electron and core hole are described by the BSE. Specifically, $|\Psi_f\rangle$ in Eq. 1 are approximated by the correlated electron-hole excitations $|S\rangle$ with excitation energies Ω_S , which are the eigenvectors and eigenvalues of the BSE Hamiltonian^{57,60,64}, \hat{H}^{BSE} . This yields

$$\sigma_{BSE}(\omega) = 4\pi^2 \alpha \omega \sum_S |\mathbf{e} \cdot \langle S | \mathbf{r} | 0 \rangle|^2 \delta(\Omega_S - \omega), \quad (3)$$

where $|0\rangle$ denotes the DFT ground state. For XANES, only transitions from the localized core-level orbitals are considered, and the transition matrix elements can be evaluated in real-space despite the periodic boundary conditions. By introducing the single-particle velocity operator $\mathbf{v} = i[\hat{H}_{BSE}, \mathbf{r}]$ and making use of the identity $\langle S | \mathbf{r} | 0 \rangle = -i \langle S | \mathbf{v} | 0 \rangle / \Omega_S$, one can prove that

$$\sigma_{BSE}(\omega) = 4\pi^2 \alpha / \omega \sum_S |\mathbf{e} \cdot \langle S | \mathbf{v} | 0 \rangle|^2 \delta(\Omega_S - \omega), \quad (4)$$

which is proportional to the imaginary part of the macroscopic dielectric constant⁶⁸,

$$\begin{aligned} \epsilon_2 &\propto 4\pi \sum_S |\mathbf{e} \cdot \langle S | \mathbf{v} | 0 \rangle|^2 \delta(\Omega_S - \omega) \\ &\propto -4\pi \text{Im} \sum_{\Phi, \Phi'} \left[\mathbf{e} \cdot \langle 0 | \mathbf{v}^\dagger | \Phi \rangle \langle \Phi | (\omega - \hat{H}_{BSE} + i\eta)^{-1} | \Phi' \rangle \langle \Phi' | \mathbf{v} | 0 \rangle \cdot \mathbf{e} \right]. \end{aligned} \quad (5)$$

Here we have expanded $|S\rangle$ into the complete basis of single excitations $|\Phi_{c\mathbf{k}\alpha}\rangle = \hat{c}_{c\mathbf{k}}^\dagger \hat{c}_{\alpha\mathbf{k}} |0\rangle$ of a core hole (α) and excited electron (c) with wavevector \mathbf{k} . Matrix elements of \hat{H}_{BSE} are given by⁶⁰

$$H_{c\alpha\mathbf{k},c'\alpha'\mathbf{k}'}^{BSE} = \Delta E_{c\alpha\mathbf{k},c'\alpha'\mathbf{k}'} + 2V_{c\alpha\mathbf{k},c'\alpha'\mathbf{k}'} + W_{c\alpha\mathbf{k},c'\alpha'\mathbf{k}'}, \quad (6)$$

where $\Delta E_{c\alpha\mathbf{k},c'\alpha'\mathbf{k}'} = (\epsilon_{c\mathbf{k}} - \epsilon_{\alpha\mathbf{k}}) \delta_{cc'} \delta_{\alpha\alpha'} \delta_{\mathbf{k}\mathbf{k}'}$ is the energy difference between the excited electron and the core hole. Because the core-hole state is localized in real space, the core-hole energy is dispersionless. Therefore we drop

the momentum dependence in ϵ_α . The matrix elements of the bare electron-hole exchange and the screened direct interaction are given by

$$V_{c\alpha\mathbf{k},c'\alpha'\mathbf{k}'} = \int d^3r d^3r' \psi_{\alpha\mathbf{k}}^*(\mathbf{r}) \psi_{c\mathbf{k}}(\mathbf{r}) v(\mathbf{r}, \mathbf{r}') \psi_{c'\mathbf{k}'}^*(\mathbf{r}') \psi_{\alpha'\mathbf{k}'}(\mathbf{r}'), \quad (7)$$

$$W_{c\alpha\mathbf{k},c'\alpha'\mathbf{k}'} = - \int d^3r d^3r' \psi_{\alpha\mathbf{k}}^*(\mathbf{r}) \psi_{\alpha'\mathbf{k}'}(\mathbf{r}) w(\mathbf{r}, \mathbf{r}') \psi_{c\mathbf{k}}(\mathbf{r}') \psi_{c'\mathbf{k}'}^*(\mathbf{r}'), \quad (8)$$

where v is the bare Coulomb kernel and $w = \epsilon^{-1}v$ is the screened Coulomb kernel under the matrix notation, i.e., $w(\mathbf{r}, \mathbf{r}') = \int d^3\mathbf{r}'' \epsilon^{-1}(\mathbf{r}, \mathbf{r}'') v(\mathbf{r}'', \mathbf{r}')$. We follow the standard implementation of the BSE method and take the static approximation in the direct term, dropping the frequency dependence in the dielectric tensor.

Practically, it is often more convenient to evaluate the matrix elements in Eqs. 7 and 8 in reciprocal space, which leads to

$$V_{c\alpha\mathbf{k},c'\alpha'\mathbf{k}'} = \frac{1}{V} \sum_{\mathbf{G}} M_{\alpha\mathbf{k}}^*(\mathbf{G}, \mathbf{q} = 0) v_{\mathbf{G}}(0) M_{\alpha'\mathbf{k}'}(\mathbf{G}, \mathbf{q} = 0), \quad (9)$$

$$W_{c\alpha\mathbf{k},c'\alpha'\mathbf{k}'} = - \frac{1}{V} \sum_{\mathbf{G}\mathbf{G}'} M_{\alpha\mathbf{k}}^*(\mathbf{G}, \mathbf{q}) w_{\mathbf{G},\mathbf{G}'}(\mathbf{q}) M_{c'\mathbf{k}'}(\mathbf{G}', \mathbf{q}) \delta_{\mathbf{q},\mathbf{k}-\mathbf{k}'}, \quad (10)$$

where V is the volume of the crystal and $M_{i\mathbf{j}\mathbf{k}}(\mathbf{G}, \mathbf{q}) = \langle i\mathbf{k} | e^{-i(\mathbf{G}+\mathbf{q})\mathbf{r}} | j\mathbf{k} + \mathbf{q} \rangle$ with \mathbf{G} being reciprocal lattice vectors.

B. XAS Implementation in OCEAN, exciting, and XSPECTRA

In this work, we compare XAS spectra calculated from three popular *ab initio* codes: OCEAN^{57,58}, **exciting**⁵⁹, and XSPECTRA^{62,63}. Key aspects of the methodology and implementations of the three codes are summarized in Table I. While **exciting** and OCEAN compute the spectra within the BSE formalism, XSPECTRA employs the CHP formalism. Additionally, five important aspects in the numerical implementations are worth mentioning: 1) the boundary condition, 2) the treatment of the core state, 3) the basis set, 4) the size of simulation cell, and 5) the treatment of empty states.

All the three codes considered here employ periodic boundary conditions. Both OCEAN and XSPECTRA are plane-wave pseudopotential codes. XSPECTRA is a module of the QUANTUM ESPRESSO code,^{69,70} while OCEAN interfaces with QUANTUM ESPRESSO to generate the necessary DFT eigenstates. In the pseudopotential method, the bare Coulomb potential of the nuclei and the effects of tightly bound core electrons are replaced by a smooth effective potential only acting on valence electrons. Thus the core-hole wavefunction is not directly accessible in OCEAN and XSPECTRA, and is obtained from a separate calculation on an isolated atom. **exciting**, on the other hand, is a full-potential all-electron code that treats both core electrons and valence electrons explicitly. **exciting** employs the augmented linearized plane-wave plus local orbital (LAPW+LO) basis set to account for both the strong variation of the wavefunction inside the core region and the smooth wavefunction in the interstitial region. As such, the core wavefunction is directly accessible in **exciting**.

In XSPECTRA, due to the presence of the core hole at the absorbing atomic site, a large enough supercell is required to avoid spurious interaction between core-hole sites. In this work, we construct a supercell with the condition that the supercell size is no smaller than 9 Å along each principle axis. Our own testing and prior work³¹ indicate that this is sufficient to converge the XAS spectra relative to the supercell size. In the CHP simulations, the excited electron can be either neglected (i.e., a charged final state) or explicitly included in the system (i.e., a neutral final state). In our XSPECTRA calculations, the excited electron is treated as part of the system and placed at the bottom of the conduction band, which is known as the excited-electron and core-hole approach^{4,71,72}. XSPECTRA evaluates the resolvent in Eq. 2 using the Lanczos algorithm^{73,74} to avoid an explicit summation over final states. Because the Hamiltonian in Eq. 2 is based on DFT, only the density of valence electrons is required to construct the Hamiltonian, and therefore XSPECTRA calculations do not require empty bands explicitly.

Both **exciting** and OCEAN require an explicit summation over the unoccupied states to construct \hat{H}_{BSE} of the unit cell, and the number of unoccupied bands have to be converged carefully, because states with higher energies can contribute to the spectral weight at much lower excitation energies through the off-diagonal terms of \hat{H}_{BSE} . In OCEAN, $(\omega - \hat{H}_{BSE} + i\eta)^{-1}$ is calculated iteratively using the Lanczos method. This term in **exciting** is computed by diagonalizing \hat{H}_{BSE} explicitly. The **exciting** code expands BSE matrix elements in reciprocal space following Eqs. 9 and 10, thus introducing an additional convergence with respect to the cut-off of $|\mathbf{G} + \mathbf{q}|$ for the summation over reciprocal lattice vectors \mathbf{G} , referred to as $|\mathbf{G} + \mathbf{q}|_{max}$. The random phase approximation (RPA) is used to compute the screened Coulomb interaction in Eq. 10, which involves another truncated summation over unoccupied states.

TABLE I. Summary of the main features of the three codes used in the XAS simulations.

	exciting	OCEAN	XSPECTRA
Method	Bethe-Salpeter Equation	Bethe-Salpeter Equation	Core-hole Pseudopotential
Boundary condition	Periodic	Periodic	Periodic
Treatment of core	Explicit, all-electron	Pseudopotential	Pseudopotential
Basis	LAPW + LO	Planewave	Planewave
Simulation cell	Unit cell	Unit cell	Supercell
Explicit empty states	Yes	Yes	No

The OCEAN code calculates the core-hole screening in real space, using a hybrid RPA and model RPA approach⁷⁵. This requires an estimation of ϵ_∞ , the electronic contribution to the static dielectric constant, and this value is taken from the Materials Project^{76,77} for insulating systems or set to 10000 for metallic systems.

It should be noted that OCEAN and **exciting** output the imaginary part of the macroscopic dielectric constant ϵ_2 , while XSPECTRA outputs the absorption cross section σ . In all three codes it is assumed that the core-level response can be treated independently for each atom and x-ray edge. To have a direct cross-code spectral comparison, we convert the XSPECTRA output to ϵ_2 using $\epsilon_2(\omega) = \sigma(\omega)/(\alpha\omega\Omega)$, where Ω is the volume of the unit cell. All the cross-code spectral comparison is performed based on ϵ_2 .

C. Edge Alignment

The energy onset of an XAS edge encodes important information about the chemical environment around the absorbing atom. In particular, the edge shift can be correlated to chemical composition, as first observed by Bergengren in the study of phosphorus K-edge XAS⁷⁸. In any system with symmetrically inequivalent sites, including heterogeneous samples with mixtures of phases or stoichiometries, a proper accounting for core-level shifts is required to accurately compare simulated XAS to experiment. Based on the studies of K-edge and L-edge XAS^{79–81}, Kunzl’s law⁸¹ states that the edge shifts are governed by the valency. The absorbing atom exhibits a positive (negative) edge shift when it is oxidized (reduced), which can be understood as the shielding effects of valence electrons on the ionic potential. In the case of most $3d$ transition metal K edges, the oxidation state of the metal site can be deduced from the edge shift². However, there are also exceptions, such as Cr, where the position of the main absorption alone is an ambiguous measure of the oxidation state^{82,83}. In early studies, the *effective charge* of the absorbing atom was determined empirically according to its chemical environment⁸⁴. The three codes we investigate here are all capable of calculating relative edge shifts, or core-level shifts, from first principles.

Several physical effects contribute to the XAS edge alignment. We can treat XAS as two sequential processes as shown in Fig. 1a. In the core electron removal step, a core electron is removed from the system leaving a core hole behind in an $N - 1$ electron system. This step corresponds to the physical process in x-ray photoemission spectroscopy (XPS). The energy needed to remove a core-level electron, i.e. the energy difference between the ground state E_N and core-excited state $E_{N-1,\alpha}^*$, includes both the impact of the local potential on the core electron and the final state relaxation in the presence of the core hole. In the electron insertion step, an electron is put back to the conduction band. The electron–core-hole interaction affects the insertion energy. In particular, low energy excitations can involve the localization of the electron, binding to the core hole to form an exciton in insulators, with energies smaller than found in the independent particle approximation. These physical effects are treated under quite different approximations in the three codes. We first compare the methods for computing the electron removal energy, providing the basis for comparison to XPS shifts and probing the approximate treatment of the core-hole relaxation energy. Then, instead of considering the electron insertion process separately, we compare the methods based on the full excitation energy calculation.

The electron removal energy of the core level α , in the quasi-particle picture, is given by its eigenenergy (ϵ_α) with the opposite sign. The leading contributions are the single particle term and electrostatics, which are well captured within DFT. In the case of **exciting**, the DFT eigenvalue of the core level ϵ_α^e is readily available. A complication arises for pseudopotential codes. In OCEAN, the relative energy of the core level can be calculated by $\epsilon_\alpha^o = X_\alpha^o + \langle \psi_\alpha | V_{\text{KS}}^{ps} | \psi_\alpha \rangle$, where V_{KS}^{ps} is the system-dependent total Kohn-Sham potential, ψ_α is the core-level wavefunction under the frozen core approximation, and X_α^o is a system-independent unknown that cancels out when comparing relative shifts. The DFT eigenenergies from **exciting** or OCEAN should give identical shifts up to the limit of the frozen-core approximation. The many-body exchange correlation effects capture the relaxation of the rest of the system in the presence of the core hole^{85,86}. The state-of-the-art is to perform a self-energy (Σ) correction to ϵ_α

under the GW approximation. An approximate self-energy correction for the core level is implemented in OCEAN. ϵ_α^o is corrected by $\Sigma_\alpha = -\langle \psi_\alpha | 1/2W_c + V_{xc} | \psi_\alpha \rangle$, where W_c is the static screened Coulomb potential of the core hole and V_{xc} is the exchange-corelation potential^{58,87}. Since this approximation uses the same screened potential needed for the BSE calculation, there is very little cost to include it. In XSPECTRA, the core electron removal energy, including initial and final state effects, is calculated using the total energy formulation within a Δ SCF procedure, $\epsilon_\alpha^x = X_\alpha^x + E_{N-1,\alpha}^* - E_N$, where X_α^x is an unknown constant due to the use of different pseudopotentials for the absorber atom in the ground state and core-hole excited state. However, for every calculation of a given core level the same change in pseudopotentials can be used, making X_α^x system-independent. Consequently, X_α^x cancels out when comparing relative shifts. Note that both XSPECTRA and OCEAN make use of the frozen-core approximation, since neither the core-hole pseudopotential nor ψ_α , respectively, are allowed to change from system to system.

For the full calculation of the excitation energies, the CHP and BSE codes have very different implementations detailed in Fig. 1b. In this study, we consider the excited core hole approximation in CHP, where the core electron is placed at the bottom of the conduction band. The electron-core-hole interaction is described through the self-consistently screened core-hole potential that takes into account the relaxation of valence electrons. Clearly, the Kohn-Sham energy levels and wavefunctions of the CHP final state are different from those of the ground state, as the attractive core-hole potential pulls the valence electrons down in energy. The relative alignment of the lowest excitation in XSPECTRA is carried out within the Δ SCF procedure as $E_{N,\alpha}^* - E_N$, where $E_{N,\alpha}^*$ is the total energy of the CHP final state. In the two BSE codes (OCEAN and `exciting`), the final state effects are treated by the direct and exchange interactions of the electron-hole pairs, where the basis functions are the ground-state orbitals as indicated in Fig. 1b. In a full GW -BSE implementation, the self-energy correction is applied to both the core level and to the conduction bands. However, this is often neglected in practical studies due to the high computational cost. As noted, in OCEAN the self-energy correction for the core level is approximated by the static core-hole relaxation energy.

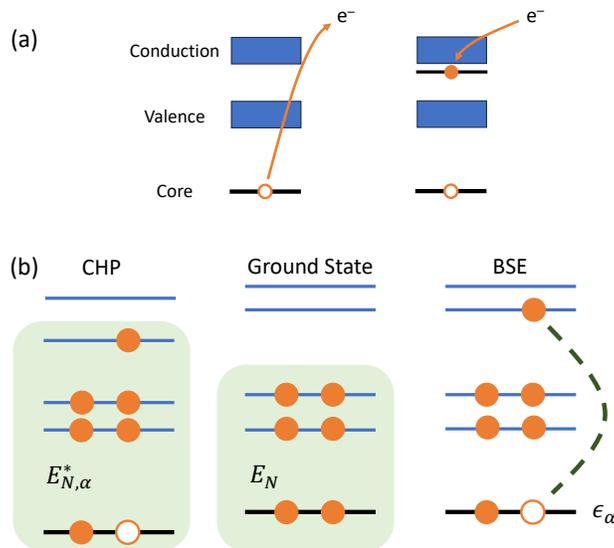


FIG. 1. Physical pictures of excitation energies that determine the edge alignment in XAS simulations. a) Separation of the neutral excitation induced by XAS into a sequence of two processes. Left: core electron removal; Right: electron insertion at the presence of the core hole. b) Excitation energies including final state effects as implemented in CHP (left) and BSE (right) codes. The dashed line in the BSE diagram indicates the electron - core hole coupling terms.

III. WORKFLOW

In this study, we focus on the XANES region of the XAS spectra, which is more sensitive to the electronic structure than the extended x-ray absorption fine structure (EXAFS) region at the higher energy range. To ensure a rigorous XANES cross-code validation, we have developed a workflow that automates the generation of the input files with a set of carefully tested input parameters. From this set of parameters, we can calculate fully converged XANES spectra using different codes. Details of this workflow are described below and summarized in Fig. 2, which includes input parameters for ground state and spectral calculations. Based on this workflow, a Python package called *Lightshow* has been developed to automate the XANES simulation input file generation for multiple codes⁸⁸.

A. Workflow for XANES spectra calculations

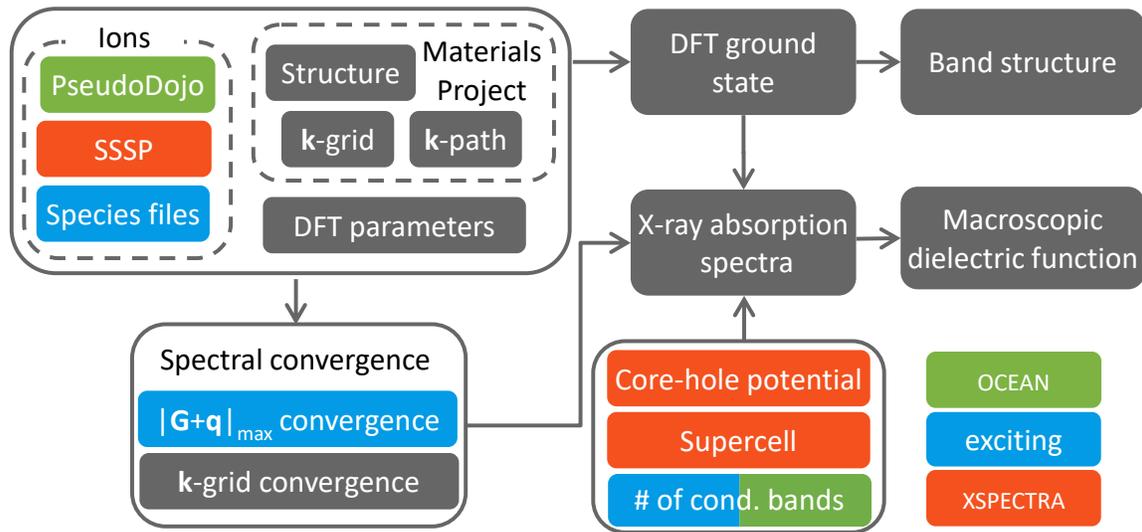


FIG. 2. Workflow of the XANES benchmark calculations. Computation steps common to all codes are shown in grey, OCEAN-specific ones in green, `exciting`-specific ones in blue, and `XSPECTRA`-specific ones in orange.

Ground state input parameters (on the upper left of Fig. 2) contain three groups: parameters to treat core electrons, structural parameters, and DFT parameters.

Core treatment. In OCEAN and XSPECTRA, the core electrons are not explicitly treated, and pseudopotentials are taken from standardized pseudopotential libraries. Norm-conserving pseudopotentials used in OCEAN calculations are taken from the PseudoDojo library⁸⁹. Pseudopotentials in XSPECTRA are taken from the precision version of the Standard solid-state pseudopotentials (SSSP) library^{43,90}, where Ti and O potentials are treated by the ultrasoft and the projector augmented wave (PAW) schemes, respectively. In the `exciting` code, the Kohn-Sham equation is solved for all electrons explicitly. The atomic basis functions, local orbitals and the respective trial energies are defined for each atomic species in the so-called *species files*. In order to obtain accurate band structures, we augment the default LAPW basis for O with one *s* and one *p* orbital, both with reference energies of 30 eV. For Ti, we augment the default settings, which contains one *s* orbital with a reference energy of -62.238 eV and one *p* orbital with a reference energy of -39.209 eV, with one *d* orbital and one *f* orbital, both with reference energies of 30 eV. These local orbitals at high energies are crucial to improve the band structure in the range of 25 eV to 40 eV above the CBM.

Structural parameters. Lattice parameters and atomic positions of the materials in the benchmark study are taken from the optimized structures obtained using the Perdew-Burke-Ernzerhof (PBE) functional⁹¹ reported in the Materials Project⁷⁶. The set of \mathbf{k} -points required for a converged DFT ground state calculation and the high-symmetry \mathbf{k} -path for the visualization of the band structure are also taken from the Materials Project. From the ground state calculation, the workflow generates the band structure.

DFT parameters. We choose the PBE exchange-correlation functional in all calculations for its general applicability and the consistency with the PBE optimized crystal structures from the Materials Project. For OCEAN and XSPECTRA, the plane-wave energy cutoffs for wavefunction and electron density employ the recommended values from the corresponding pseudopotential libraries. The total energy convergence threshold in the self-consistent field calculation is set to 10^{-10} Rydberg per atom. A Gaussian smearing of 0.002 Rydberg is used in XSPECTRA and `exciting` calculations. In OCEAN calculations, a Gaussian smearing of 0.02 Rydberg is included only for metallic systems. This small difference in the choice of smearing between XSPECTRA/`exciting` and OCEAN does not affect our results, as supported by the comparison of the band structure from different codes in the *Results* Section. The DFT wavefunctions and orbital energies are the input of Eqs. 2 and 5 for the subsequent XANES calculations using OCEAN. `exciting` adopts most of its default parameters for ground state properties.

Spectral parameters. For all the three codes, a Lorentzian broadening was applied to the Ti K-edge spectra with a half-width at half maximum of 0.89 eV. This value was chosen to give good resolution of the spectral features for measuring the similarity across different codes. For a quantitative comparison between simulation and experiment, multiple broadening mechanisms need to be considered, such as core-hole lifetime, the photo-electron lifetime in the final state and the resolution of the instrument⁹².

XSPECTRA calculations require a core-hole pseudopotential for the absorber atom. We use the same Ti core-hole PAW pseudopotential as in the Ref. 31, which was optimized for the $4p$ scattering states generated using the *atomic* module in the QUANTUM ESPRESSO code. During the reconstruction of the all electron wavefunction in the XSPECTRA calculations, we set the PAW radius (r_{paw}) the same as the cutoff radius in the core-hole PAW pseudopotential. XSPECTRA uses the Lanczos iterative method to solve Eq. 2. The maximum number of iterations is set to 5000, and the early exit threshold is set to 0.01 for spectra 200 iterations apart.

Both OCEAN and `exciting` use electron-hole pairs as a basis to expand the core electron excitations, which involves an explicit summation over unoccupied bands. The number of empty bands (n_c) needed to span a given energy range can be estimated from the density of states of the non-interacting electron gas: $n_c = (\sqrt{2}/3\pi^2)(\Delta E)^{3/2}\Omega$, where Ω is the volume of the unit cell and ΔE is the desired energy range in Ha. In practice the energy range of a given n_c is less than this formula estimates, and an input of $\Delta E \approx 50$ eV yielded a conduction band count sufficient to span approximately 40 eV in the systems we studied. While `exciting` explicitly diagonalizes the BSE Hamiltonian, OCEAN uses the Lanczos iterative method. The maximum number of iterations is set to 1000, with an early exit criterion of 0.001 for spectra 5 iterations apart. The total number of empty bands for computing the RPA dielectric response function is determined by setting the energy range to 100 eV in both OCEAN and `exciting`.

Finally, there are a set of spectral parameters that require careful convergence tests. Common to the three codes, the k -grid used in the Brillouin zone (BZ) sampling needs to be converged, and generally XAS calculations require denser sampling of the reciprocal space than ground state calculations. In `exciting` and OCEAN calculations, we employ the same k -grid shift of (0.125, 0.250, 0.375) in the reciprocal lattice vector coordinates. We use Γ -centered k -grids to sample the BZ of the supercell in XSPECTRA calculations. Specific to the `exciting` code, it requires the convergence with respect to the cutoff $|\mathbf{G} + \mathbf{q}|_{max}$ in Eqs. 9 and 10. Details on the convergence study are discussed in the next section.

B. Convergence of Spectral Parameters

1. Spectral similarity metric

In order to define the threshold for the convergence of spectral parameters, we introduce a spectral similarity metric. This definition is not unique and there are several possible choices, such as the Cosine similarity, L_2 -normalized Euclidean distance, the Pearson correlation coefficient, and the Spearman's rank correlation score⁹³. Without losing the generality, in this study we choose the Spearman's rank correlation score (r_{sp}) as our spectral similarity metric, and we have verified that the overall trend in the convergence behavior is similar when using the Cosine similarity or the Pearson correlation coefficient. Given two XANES spectra on the same grid, μ_1 and μ_2 , r_{sp} is defined as the Pearson correlation coefficient of the corresponding rank variables $R(\mu_1)$ and $R(\mu_2)$,

$$r_{sp} = \frac{\text{cov}(R(\mu_1), R(\mu_2))}{\sigma_R(\mu_1)\sigma_R(\mu_2)}, \quad (11)$$

where the rank function returns the rank of a given value when compared with the rest of the values in a list, and cov and σ denote the covariance and standard deviation of $R(\mu_1)$ and $R(\mu_2)$, respectively. If there are no repeated data points, a perfect Spearman's rank correlation yields +1 or -1 when the two spectra are perfectly monotonically correlated or anti-correlated, whether linear or not. In this sense, the Spearman's rank correlation score is more sensitive to the spectral shape than the Pearson correlation coefficient that focuses on the linear correlation. Further, we define the spectral similarity as

$$s = \log_{10}(1 - r_{sp}). \quad (12)$$

By visual inspection, we consider a spectrum to be converged with respect to the reference, when $r_{sp} \geq 0.999$ or equivalently $s \leq -3$. Under this threshold, two XANES spectra show negligible differences (see Fig. S1b in the Supplementary Material⁹⁴). At an even better similarity score of $s \leq -4$, two spectra are on top of each other as shown in Fig. S1c.

2. K -point grid resolution metric

Because all the three codes use periodic boundary conditions, the implicit integral in Eqs. 2 and 5 is evaluated as a spatial integral within the unit cell over crystal momentum within the first BZ. In practice, the integral over the BZ is carried out as a finite, discrete sum over k points on a regular mesh. It is important that this sampling is performed

on a k -grid with fine enough resolution, such that the numerical integration over a finite number of k points yields converged values. Although this convergence test can be performed trivially on a single system according to the size of the k -mesh (n_1, n_2, n_3), a system-independent metric needs to be introduced, because the resolution of the k -grid depends on both the mesh size and the size/shape of the BZ. For this purpose, we recognize that an $n_1 \times n_2 \times n_3$ k -point mesh of the unit cell is equivalent to a Γ -point sampling of an $n_1 \times n_2 \times n_3$ supercell. The k -grid resolution can be quantified by the effective crystal size (R_{eff}) defined as the shortest spacing between parallel faces of this supercell,

$$R_{\text{eff}} = 2\pi \min_i \left\{ \frac{n_i}{|b_i|} \right\}, \quad (13)$$

$$|b_i| = 2\pi \epsilon_{ijk} \left| \frac{a_j \times a_k}{a_i \cdot (a_j \times a_k)} \right|, \quad (14)$$

where a and b are the lattice vectors and reciprocal lattice vectors, respectively, and ϵ_{ijk} is the Levi-Civita symbol. Note that the magnitude of b_i depends inversely on the magnitude of a_i and directly on the angle between a_i and $a_j \times a_k$.

C. Ti-O-10 Dataset

In this work, we choose a dataset consisting of ten representative titanium oxide (Ti_xO_y) compounds referred to as the Ti-O-10 dataset, such that they cover a reasonably wide range of the materials space characterized by: insulator (5) and metal (5), oxidation state, coordination number, as well as the number of symmetrically nonequivalent absorber sites. This list includes common polymorphs observed experimentally: rutile (mp-2657), anatase (mp-390), and brookite (mp-1840). Details of the material attributes are listed in Table II.

TABLE II. Key attributes of the systems in the Ti-O-10 dataset: Materials Project id (mpid), chemical formula, space group, band gap (E_g), oxidation state (OS), coordination number (n_{coord}), and number of nonequivalent sites (n_{site}).

mpid	Formula	Space group	E_g (eV) ^a	OS	n_{coord}	n_{site}
mp-10734	Ti ₄ O ₅	I4/m	0.0	2.5+	6	1
mp-1203	TiO	C2/m	0.0	2+	4, 5	3
mp-1215	Ti ₂ O	P $\bar{3}$ m1	0.0	1+	3	1
mp-1840	TiO ₂ (brookite)	Pbca	2.29	4+	6	1
mp-2657	TiO ₂ (rutile)	P4 ₂ /mnm	1.77	4+	6	1
mp-2664	TiO	Fm $\bar{3}$ m	0.0	2+	6	1
mp-390	TiO ₂ (anatase)	I4 ₁ /amd	2.06	4+	6	1
mp-430	TiO ₂	P2 ₁ /c	2.23	4+	7	1
mp-458	Ti ₂ O ₃	R $\bar{3}$ c	0.0	3+	6	1
mvc-11115	TiO ₂	R3m	2.46	4+	4, 6	2

^a Band gap values were extracted from the Materials Project⁷⁶, which were calculated with the PBE functional.

IV. RESULTS

A. Band Structure

As the first step of the cross-code validation, we compare the band structure obtained from the three codes in order to assess the quality of pseudopotentials used for OCEAN and XSPECTRA as well as the settings of the *species* files for `exciting`. Band gaps of the insulators calculated from the three codes are very close to each other, with OCEAN / XSPECTRA and `exciting` within 30 meV as shown in Table III. This difference is the same as that within the pseudopotential codes. Note that these band gap values are calculated from a rough k -grid used in the ground state calculations, as the main purpose is to compare different codes and not to get accurate band gaps. Therefore small differences are expected between values reported in Table III and those from the Materials Project (Table II). Our results suggest that the error introduced by the pseudopotentials is negligible for the band gap.

Next we perform a more comprehensive comparison of the DFT energy levels on the k -grids of the ground state as shown in Table IV. Since the absolute energy eigenvalues from different codes cannot be compared directly, prior to

TABLE III. Comparison of the DFT-PBE band gap (in eV) of the insulators from three codes.

mpid	XSPECTRA	OCEAN	exciting
brookite (mp-1840)	2.35	2.37	2.36
rutile (mp-2657)	1.82	1.84	1.82
anatase (mp-390)	2.12	2.14	2.12
TiO ₂ (mp-430)	2.29	2.32	2.29
TiO ₂ (mvc-11115)	2.50	2.51	2.50

TABLE IV. Comparison of the DFT energy levels. The first, second and third columns in each section correspond to energy differences (in meV) between **exciting** and OCEAN (E-O), XSPECTRA and OCEAN (X-O), and **exciting** and XSPECTRA (E-X), respectively. The largest RMSD and Max D for each energy range are highlighted in bold.

mpid	comparison	Val. band			CBM + 10 eV			CBM + 20 eV			CBM + 30 eV		
		X-O	E-O	E-X	X-O	E-O	E-X	X-O	E-O	E-X	X-O	E-O	E-X
mp-390	RMSD	16	18	2	17	17	2	13	13	5	14	13	7
anatase	Max D	37	40	5	32	33	5	32	33	12	36	44	32
mp-2657	RMSD	16	18	3	17	17	2	14	14	5	16	14	7
rutile	Max D	38	40	5	31	33	6	31	38	16	40	57	59
mp-1840	RMSD	16	19	4	16	10	15	12	15	18	14	12	18
brookite	Max D	37	42	8	29	28	26	29	37	27	31	37	27
mp-430	RMSD	17	19	2	13	12	3	11	10	5	14	11	6
TiO ₂	Max D	40	43	4	22	22	6	23	22	11	34	37	16
mvc-11115	RMSD	15	16	2	10	8	3	8	8	5	13	11	5
TiO ₂	Max D	38	37	5	18	15	6	18	29	12	43	60	19
mp-10734	RMSD	13	13	2	5	8	4	11	9	6	22	16	9
Ti ₄ O ₅	Max D	34	35	5	15	15	7	36	30	20	77	99	48
mp-1203	RMSD	13	13	2	4	6	4	8	7	5	19	15	6
TiO	Max D	36	36	5	13	14	7	35	28	17	69	90	44
mp-1215	RMSD	10	10	2	5	9	4	9	8	9	19	11	17
Ti ₂ O	Max D	37	35	7	13	14	8	35	29	34	70	89	77
mp-2664	RMSD	13	12	2	4	6	4	10	8	10	23	15	25
TiO	Max D	36	35	6	15	17	9	42	37	39	82	124	130
mp-458	RMSD	14	15	2	5	7	3	7	7	4	16	11	9
Ti ₂ O ₃	Max D	34	35	4	11	16	6	26	23	20	53	54	51

performing the comparison, the energy levels from different codes are aligned by minimizing the root mean squared deviation (RMSD) of the valence bands (see Appendix A). The differences of the energy eigenvalues are characterized by RMSD and the maximum deviation (Max D). We collect the statistics for each phase and each code in four energy regions: the entire valence band and the conduction band spanning from the CBM to 10, 20, and 30 eV above the CBM. For the entire valence band, comparing OCEAN and XSPECTRA, across the phases the largest RMSD and Max D are 17 meV and 40 meV, respectively. The differences between OCEAN / XSPECTRA and **exciting** are similar, with the largest RMSD of 19 meV and Max D of 43 meV. For most systems, XSPECTRA agrees better with **exciting** (with the largest Max D of 8 meV) than OCEAN, which is likely due to the optimized pseudopotentials in SSSP. The good agreement among the three codes extends to more than 20 eV above CBM. Within CBM+10 eV, the agreement between OCEAN / XSPECTRA and **exciting** is comparable to that for the valence band, with the largest RMSD of 17 meV and max D of 33 meV. Even at CBM+20 eV, the same good agreement still holds with the largest RMSD of 18 meV and Max D of 42 meV. At even higher energies, the differences increase more rapidly because it becomes more difficult to obtain accurate band structure for scattering states at very high energies. In practice, it requires additional projectors in the construction of the pseudopotential or additional local basis functions in the LAPW+LO formalism optimized for high energy scattering states, respectively. Within CBM to CBM+30 eV, the RMSD increases only slightly to 25 meV, as it is averaged over the 30 eV energy range. However, the largest Max D

as a more sensitive quantity grows up to 130 meV in the metallic system, mp-2664. To provide a direct visualization of the energy deviations between different codes, we plot the band structure of mp-2664 as obtained from `exciting` and highlight the difference to that obtained using QUANTUM ESPRESSO with the PseudoDojo pseudopotentials as used in OCEAN in Fig. 3. All other materials have better agreement in the band structure. We found that the large differences mainly come from the energy range between CBM+25 and CBM+30 eV along $\Gamma - X$, $\Gamma - L$ and $\Gamma - K$ lines. The band structure comparison shown in Table IV and Fig. 3 gives us the baseline error due to different treatment of the core electrons, i.e., choice of different pseudopotentials and pseudopotential versus all-electron, when comparing XANES spectra from the three codes. Further optimizing pseudopotentials or the LAPW+LO basis functions might improve the agreement at around CBM+30 eV, but this is non-trivial and beyond the scope of the current work.

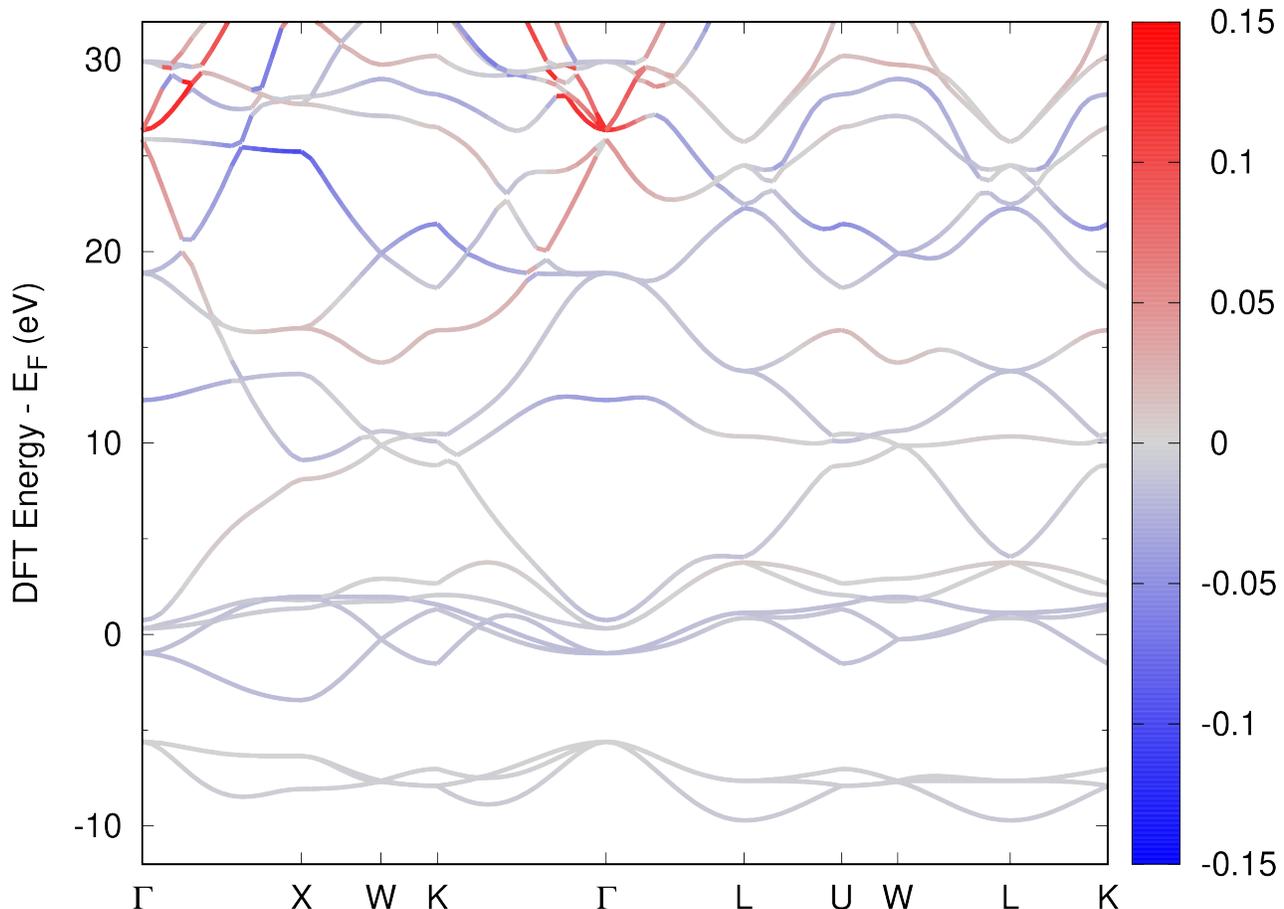


FIG. 3. Band structure of mp-2664 calculated with the `exciting` code. The colors signify the energy difference (in eV) between `exciting` and QUANTUM ESPRESSO (with the PseudoDojo pseudopotentials). The energy zero of `exciting` is set to the Fermi level, the QUANTUM ESPRESSO alignment used in OCEAN was determined by minimizing the RMSD over the occupied states down to -20 eV (a cutoff that neglects the O $2s$ states and the semi-core Ti $3s$ and $3p$ states).

B. $|\mathbf{G} + \mathbf{q}|_{\max}$ Convergence

The `exciting` calculations require convergence of $|\mathbf{G} + \mathbf{q}|_{\max}$ for the summation over reciprocal lattice vectors in Eqs. 9 and 10, which is controlled by the input parameter, GQMAX, in units of Bohr⁻¹. To study its convergence behavior, $|\mathbf{G} + \mathbf{q}|_{\max}$ is varied from 2 to 6 in steps of 1, except for mp-1840, where the highest $|\mathbf{G} + \mathbf{q}|_{\max}$ value is

set to 5 due to the computational cost. For each calculation, we compare the spectrum with the reference obtained with the highest $|\mathbf{G} + \mathbf{q}|_{\max}$ and determine the threshold for convergence when the similarity with the reference drops below -3 . For systems with multiple sites (mp-1203 and mvc-11115), only the convergence for the first site is shown in Fig. 4 as the convergence behavior is similar for different sites. For most materials, converged spectra can be obtained with $|\mathbf{G} + \mathbf{q}|_{\max} = 4$, except for mvc-11115 where a higher value of $|\mathbf{G} + \mathbf{q}|_{\max} = 4.2$ is required. In the following calculations, $|\mathbf{G} + \mathbf{q}|_{\max}$ is set to 4.2 for mvc-11115, while for the rest of the materials a value of 4 is adopted.

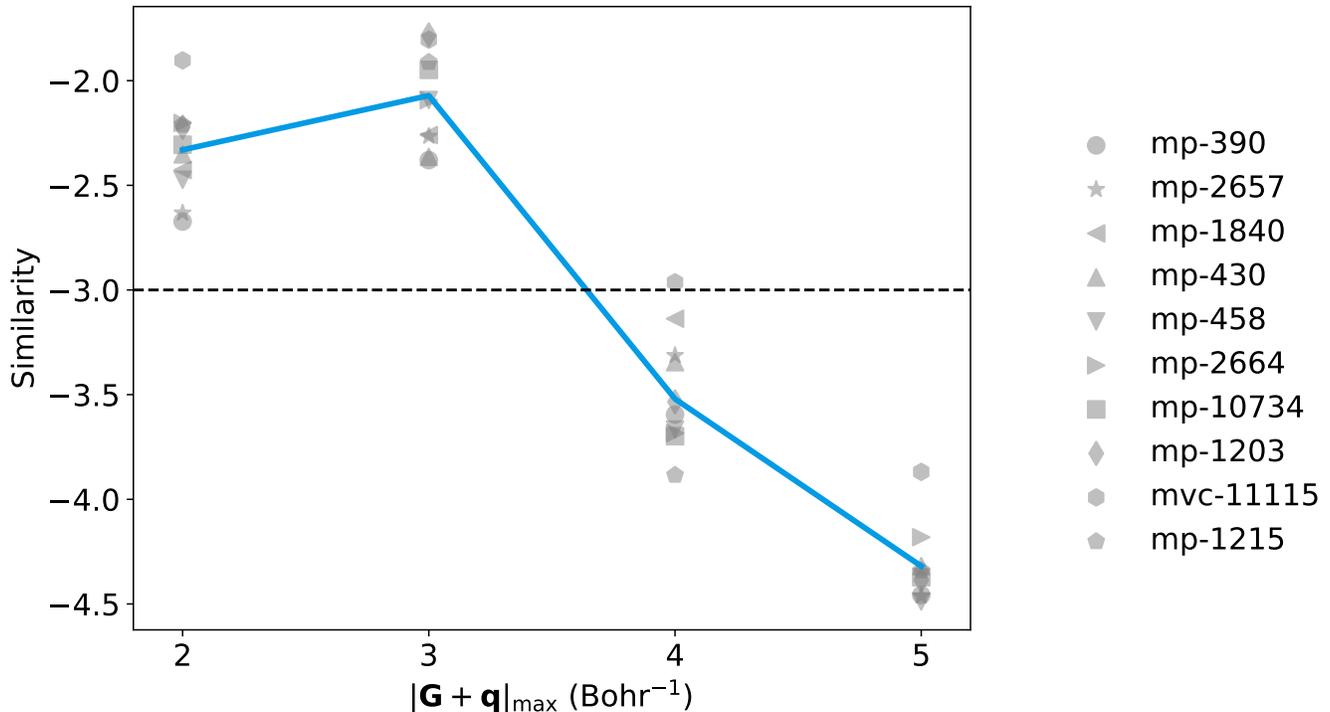


FIG. 4. Convergence of the **exciting** spectra with respect to $|\mathbf{G} + \mathbf{q}|_{\max}$. The blue solid line indicates the average similarity at each $|\mathbf{G} + \mathbf{q}|_{\max}$ value.

C. K -grid Convergence

We perform the k -grid convergence study on the Ti-O-10 dataset against R_{eff} in a wide range up to about 60 Bohr. For each material, we choose an over-converged dense k -mesh as the reference, corresponding to R_{eff} values between 55 to 65 Bohr. For a given R_{eff} , there exist multiple choices of the k -grid $\{n_i\}$, and we choose the smallest one ensuring similar spacings between reciprocal lattice vectors. The calculated spectra on the coarse k -grids are compared to the reference spectra, and we choose $s \leq -3$ as the threshold for convergence (see Fig. S1). As shown in Fig. 5, overall the slopes for the three codes are similar. For each code, we perform a linear regression for data from each site and determine the intersection between the linear fit with the horizontal line of $s = -3$. The largest R_{eff} value at the intersection among all the absorption sites gives the convergence threshold of R_{eff} , as indicated by green (OCEAN), orange (XSPECTRA) and blue (**exciting**) dashed lines in Fig. 5. Out of the three, OCEAN and **exciting** converge at a similar pace ($R_{\text{eff}} > 32.8$ Bohr), while XSPECTRA converges more slowly ($R_{\text{eff}} > 42.7$ Bohr). The different k -grid convergence behavior is likely due to the major difference in the methodology — OCEAN/**exciting** perform BSE calculations on the unit cell and XSPECTRA performs CHP calculations on the supercell containing a core hole.

D. Cross-code Spectral Comparison

For the cross-code comparison, one option is to use spectra computed at the converged k -grid based on R_{eff} (see above section). However, any similarity measure better than -3 is not guaranteed to be meaningful. For this reason, we choose to compare the reference spectra obtained from the densest k -grids used in the three codes. In a pairwise

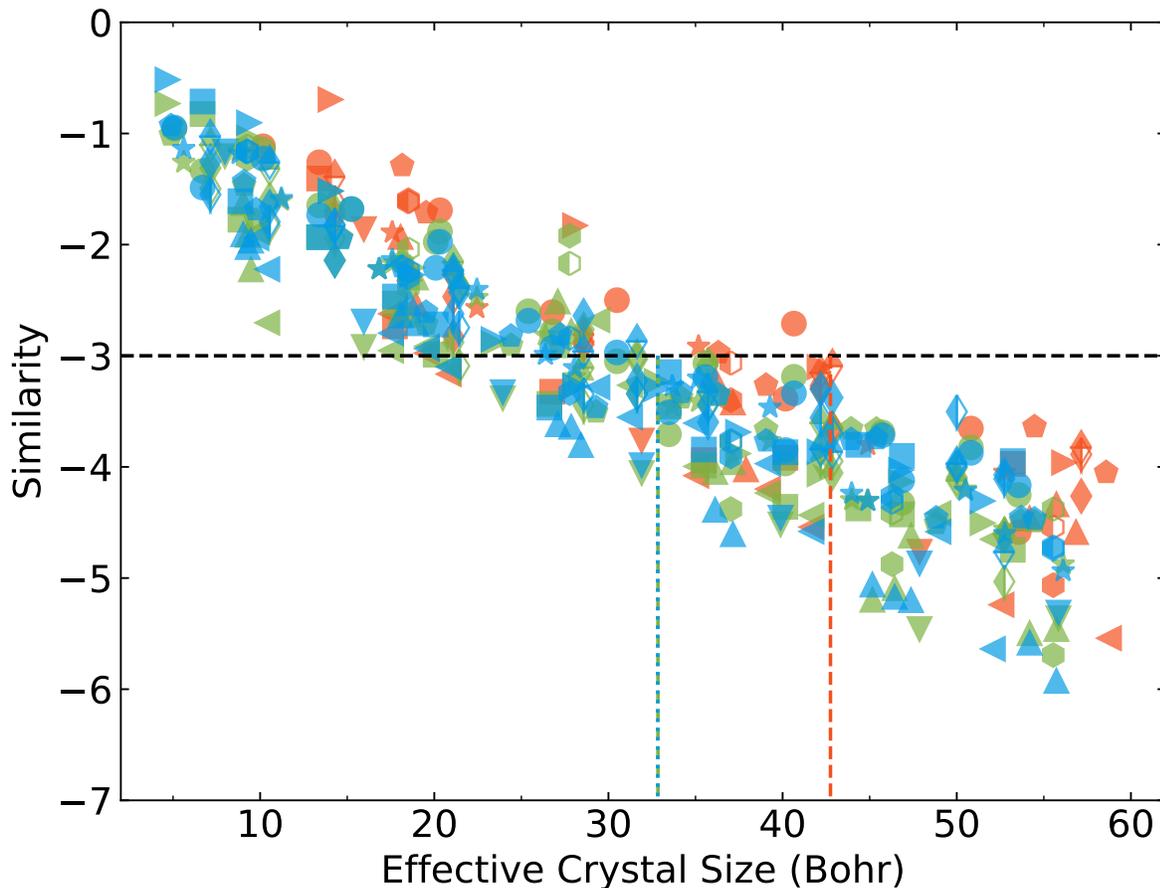


FIG. 5. Multi-code k -grid convergence as a function of the effective crystal size (see text). The spectra obtained from different codes are represented by different colors: green (OCEAN), orange (XSPECTRA) and blue (**exciting**). The symbols designating the materials are the same as in Fig. 4. Partially filled symbols with the same shape represent different absorption sites of the material. Vertical dashed lines denote the effective crystal size beyond which the similarity criterion is fully satisfied for each code.

comparison, we first align two spectra truncated at 35 eV above the onset (see Section II C) and then calculate the similarity of the aligned truncated spectra. The explicit number of bands needs to be provided in OCEAN and **exciting** calculations to yield a large enough energy range of about 35 eV. In this work, we estimate the required number of conduction bands based on the homogeneous electron gas model (see Section III A). The model turns out to work quite well for most of the systems, except for mp-2664, which has the smallest unit cell out of the 10 materials. For mp-2664, we manually increase the number of bands to ensure a sufficiently large energy range (see Appendix B). In Fig. 6, we compare the spectra from the three codes for two representative systems: mp-2657 (rutile TiO_2 , insulator) and mp-458 (Ti_2O_3 , metal). We present the results on both the initial state rule (independent-particle approximation) and final state rule (including electron-hole interactions). Under the initial state rule, the core-hole effects are not considered, and the spectra reflect only the band structure effects. In practice, we switch off the exchange and direct coupling terms in BSE in OCEAN and **exciting**, and use the neutral Ti pseudopotential instead of the core-hole pseudopotential in XSPECTRA. Under the final state rule, the spectra correspond to the BSE or CHP Hamiltonian described in Section II A. The full spectral similarity measure of the Ti-O-10 dataset (including all 3 code pairs) is summarized in Table V.

The calculations for all three codes were carried out to converge the spectra up to 35 eV above the edge. As seen in Fig. 6, deviations start to appear above 35 eV as the impact of the finite number of empty states starts to emerge for OCEAN and **exciting**. The features in the high energy tail where the spectral intensity vanishes are very close

in OCEAN and `exciting`, which is consistent with the good agreement of the band structure shown in Table IV. On the other hand, XSPECTRA spectra extend into a much larger energy range because of the use of the Lanczos formalism that does not require empty states explicitly. At the independent particle level, the spectral shapes show nearly perfect agreement as can be seen in Figs. 6a and 6c, respectively. The XSPECTRA *vs.* OCEAN comparison reveals high similarities of $s = -3.89$ for mp-2657 and -3.98 for mp-458, as expected within the pseudopotential codes. The XSPECTRA results overall show slightly stronger intensity than OCEAN and `exciting`. This is likely due to the use of different pseudopotentials, which can affect the shape of the conduction band wavefunctions as well as the transition matrix elements that depend on choice of the projectors. The convergence criteria based on Spearman's rank correlation is more sensitive to spectral shape and less so to small differences in the absolute spectral intensity. The similarity of `exciting` *vs.* OCEAN and XSPECTRA *vs.* `exciting` are slightly worse with $s = -3.17$ and -3.08 for mp-2657 and -2.71 and -2.59 for mp-458, respectively. Across the Ti-O-10 dataset, the average similarity at the independent particle level is -3.74 between OCEAN and XSPECTRA and becomes slightly worse between OCEAN / XSPECTRA and `exciting` (better than -2.76) as shown in Table V.

Overall, there is good agreement between the three codes at the interacting particle level, as shown in Figs. 6b and 6d. We notice that better agreement is obtained from the two BSE codes and the treatment of core electrons (pseudopotential *vs.* all electron) has a very small impact. As a result, noticeable differences emerge between OCEAN / `exciting` and XSPECTRA, especially regarding peak intensities. For example, near 20 eV in mp-2657, the first peak is higher than the second peak in the OCEAN / `exciting` spectra, while it is the opposite in the XSPECTRA spectrum. In addition, the shoulder peak in OCEAN / `exciting` near 10 eV is higher by about 22% than that in XSPECTRA. As a result, OCEAN and `exciting` have high similarity values of -2.79 in mp-2657 and -2.95 in mp-458, while r_{sp} between XSPECTRA / OCEAN and `exciting` drops by an order of magnitude as compared to the independent particle level, with similarity scores of -2.00 and -2.17 in mp-2657 and -2.79 and -2.71 in mp-458, respectively. From Table V, one can see that at the interacting particle level the average similarity is -2.65 between OCEAN and `exciting`, which is slightly worse than the value of -2.93 at the independent particle level. This small additional deviation most likely results from the different numerical implementation of the BSE Hamiltonian, especially the dielectric screening (see Section II B). The average similarity is about -2.02 between XSPECTRA and OCEAN / `exciting`. We do not find any substantial difference in the overall agreement between insulators and metals.

TABLE V. Spectral similarity of the Ti-O-10 dataset among the three codes at both the independent and interacting particle level. Values in the different columns refer to the comparison between XSPECTRA and OCEAN, `exciting` and OCEAN, `exciting` and XSPECTRA, respectively.

mpid	exciting vs. OCEAN		XSPECTRA vs. OCEAN		XSPECTRA vs. exciting	
	independent	interacting	independent	interacting	independent	interacting
mp-390	-2.80	-2.33	-3.72	-1.49	-2.69	-1.63
mp-2657	-3.17	-2.79	-3.89	-2.00	-3.08	-2.17
mp-1840	-3.25	-2.73	-3.93	-2.02	-3.05	-2.26
mp-430	-3.53	-2.56	-3.82	-1.91	-3.30	-2.09
mvc-11115:0	-3.49	-1.73	-4.08	-1.49	-3.23	-1.16
mvc-11115:1	-2.99	-2.29	-3.78	-1.55	-2.75	-1.43
mp-1203:0	-2.82	-2.87	-3.63	-2.08	-2.67	-2.16
mp-1203:2	-2.32	-2.85	-3.40	-2.30	-2.17	-2.32
mp-1203:4	-2.78	-2.83	-3.20	-2.01	-2.48	-1.92
mp-10734	-2.85	-2.49	-3.75	-1.91	-2.71	-2.10
mp-1215	-2.72	-2.96	-3.66	-2.40	-2.62	-2.16
mp-2664	-2.64	-3.01	-3.83	-2.31	-2.53	-2.19
mp-458	-2.71	-2.95	-3.98	-2.79	-2.59	-2.71
average	-2.93	-2.65	-3.74	-2.02	-2.76	-2.02

E. Comparison of the final-state effects between the BSE and CHP methods

We observe that for the Ti-O-10 dataset, the results from the two BSE codes, OCEAN and `exciting`, exhibit more spectral weight near the edge than those from the CHP code, XSPECTRA (See Fig. 6 and Fig. S2⁹⁴). This trend is more pronounced in the insulating systems than in the metallic systems. This is indicative, but not proof, of a stronger

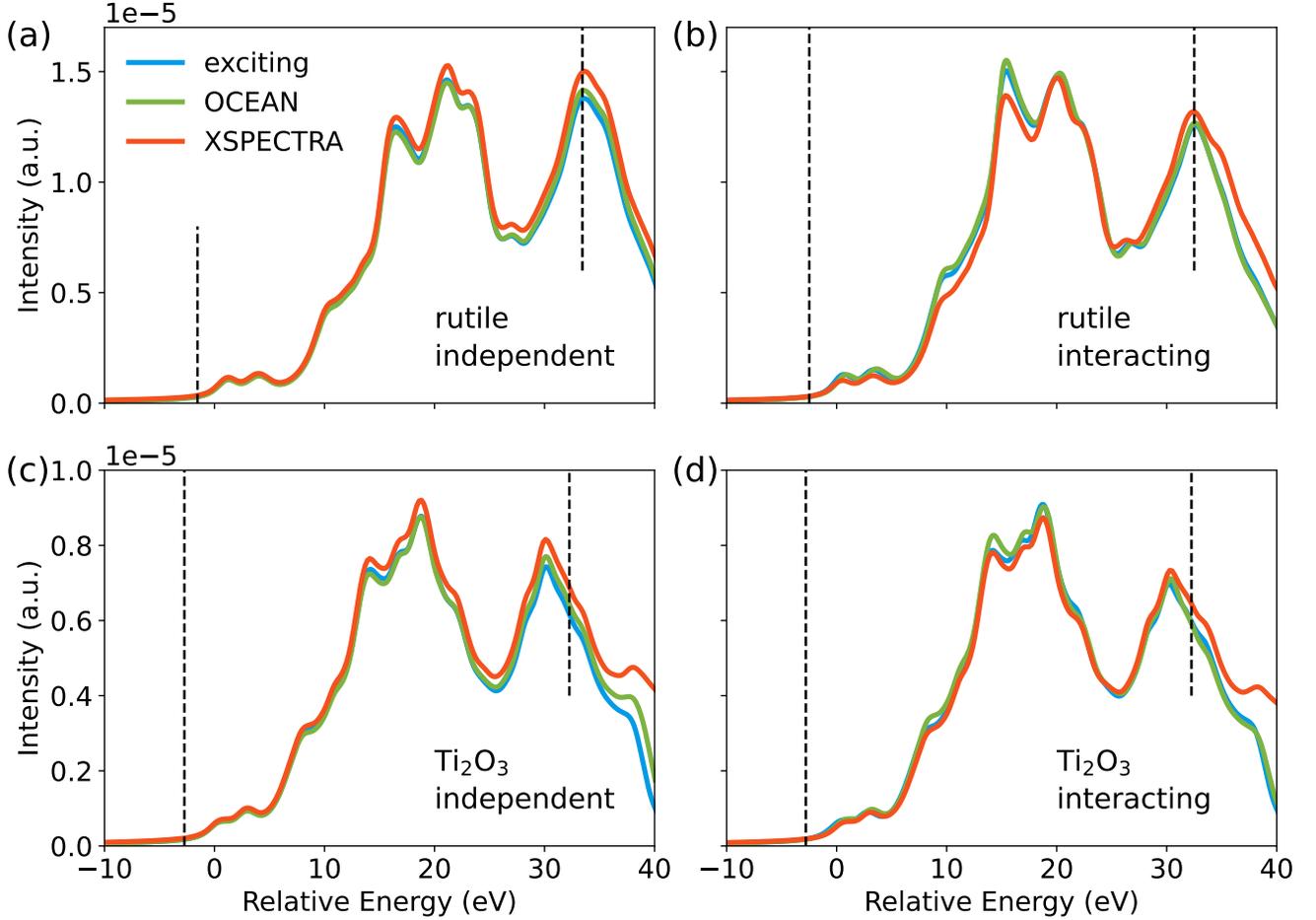


FIG. 6. Comparison of the spectra of rutile (mp-2657) and Ti_2O_3 (mp-458) from all three codes at both the independent and interacting particle level. The first and second vertical dashed lines indicate the absorption onset and 35 eV above it, where a quantitative comparison was performed. The energy is relative to the excitation to the CBM.

core-hole potential, or, equivalently, weaker screening of the core hole in the BSE approach. To better understand the discrepancies between the CHP and BSE calculations, we examine the differences in the two theoretical approaches. Since OCEAN and EXCITING results are nearly identical to each other, we will only use OCEAN for this analysis.

The differences between BSE and the final-state rule at the presence of a core hole have been discussed previously by Rehr, Soininen, and Shirley⁹⁵, and three main differences can be identified. Both the occupied and unoccupied electron states change when a core hole is created⁹⁶. In CHP calculations, the valence band manifold is relaxed in presence of the core hole, while the BSE calculation uses the eigenstates of ground-state DFT. Although this difference can modify the weights of the transition matrix elements, as shown by Roychoudhury and Prendergast⁹⁷, this effect is expected to be small for the particular materials and edge studied here. The empty or mostly empty Ti d -bands sit several eV below the main-edge transitions into the Ti $4p$ states that also hybridize with neighboring oxygen orbitals. Additionally, Liang *et al.* investigated the role of secondary electron-hole excitations in a ΔSCF context, and found that they played only a small role in the O K-edge XAS of TiO_2 (in contrast to later transition metal oxides)^{10,98}. Second, the BSE Hamiltonian includes an exchange interaction between the excited electron – core-hole pairs that is absent in CHP. The importance of this term can be investigated by selectively turning it off. Third, the screened core-hole potential is different between the CHP and BSE methods, owing to differences in the calculation of the induced density response to the core hole. Using OCEAN we investigate several different approximations for calculating the induced density response.

1. Induced density in response to the core-hole potential

In response to the creation of a core hole, the electrons relax, screening the core-hole potential. In the CHP approach the core-hole potential and valence screening are all included in the self-consistent DFT calculation. In the BSE approach, electrons interact with the core hole through the screened potential W , which can be written using the dielectric response ϵ or in terms of an induced density ρ_{ind} and induced potential V_{ind} ⁹⁹,

$$W(\mathbf{r}) = \epsilon^{-1}(\mathbf{r}, \mathbf{r}')V_{\text{ext}}(\mathbf{r}') = V_{\text{ext}}(\mathbf{r}) + V_{\text{ind}}(\mathbf{r}) \quad (15)$$

$$\rho_{\text{ind}}(\mathbf{r}) = \chi(\mathbf{r}, \mathbf{r}')V_{\text{ext}}(\mathbf{r}') \quad (16)$$

$$V_{\text{ind}}(\mathbf{r}) = v(\mathbf{r}, \mathbf{r}')\rho_{\text{ind}}(\mathbf{r}') \quad (17)$$

where v is the Coulomb operator and χ is the reducible polarizability, and implicit summation is assumed. Note that both the CHP and BSE approaches use a static approximation for the screening of the core-hole potential^{100,101}.

Below we consider three methods to construct V_{ind} . In the first method, (as is typical for a BSE calculation), the reducible polarizability is calculated within the RPA,

$$\chi^{\text{RPA}} = (1 - \chi^0 v)^{-1} \chi^0 \quad (18)$$

where χ^0 is the irreducible polarizability. In the second method, the many-body effects can be included through the exchange-correlation kernel, f_{xc} ,

$$\chi = [1 - \chi^0(f_{xc} + v)]^{-1} \chi^0 \quad (19)$$

where V_{xc} is the exchange-correlation potential and $f_{xc} = \delta V_{xc} / \delta \rho$ ¹⁰². Under the linear response framework, normally the static approximation is also used, i.e., an adiabatic f_{xc} . Within OCEAN, f_{xc} can be included within the adiabatic local-density approximation (ALDA)⁷⁵.

A third method for determining the screened potential is based on the CHP method, where a self-consistent DFT calculation is carried out with a core hole and an excited electron at the bottom of the conduction band. The induced density is calculated as the density difference between two self-consistent DFT calculations, with (CH) and without the core hole,

$$\rho_{\text{ind}}^{\text{CH}} = \rho^{\text{CH}} - \rho^{\text{GS}}, \quad (20)$$

where GS stands for ground state. This approach has been used previously for linear-response calculations by applying sufficiently small perturbing potentials^{103,104}, but here we use the screening potential from a full core-hole.

Both the linear response and self-consistent approaches entail approximations. In the case of linear response, the bare core-hole potential is not weak, and non-linearity in the screening response may be important. In the case of the self-consistent approach, the resulting screened potential is due to the electrons relaxing into the lowest energy configuration (in the presence of the core hole), a process that is not instantaneous. However, the relevant time scale for the screening is the inverse of the plasmon frequency (≈ 10 eV to 20 eV) as compared to the core hole lifetime (inverse of the 1 eV broadening). These competing time scales apply equally to the linear response case. In some special cases, such as those with localized orbitals or defects, there may also be very slow screening processes that are incorrectly included in the self-consistent core-hole potential. However, this is not the case in the systems we are investigating here. The issue of screening time can be side-stepped entirely within a real-time formalism, such as real-time time-dependent DFT (RT-TDDFT)¹⁰⁵, though x-ray implementations of RT-TDDFT tend to use non-periodic boundary conditions making comparisons with periodic boundary condition code calculations of extended systems difficult.

2. Comparison of XANES spectra using different approximations

We first investigate the effect of the exchange interaction using a metal (Ti_2O_3 , mp-458) and an insulator (rutile, mp-2657). OCEAN calculations using the CHP-derived screened core-hole potential (W-CHP) are shown in Fig. 7 both with and without the exchange (W-CHP w/o EX). As the exchange term is repulsive, it slightly shifts spectral weight away from the edge, e.g., the small intensity decrease at the first peak in W-CHP as compared to W-CHP w/o EX, but it has little effect on the pre-edge features. While these are dipole transitions enabled due to hybridization between the titanium and neighboring oxygen atoms, the states are primarily $3d$ in nature, and the exchange terms between the $1s$ and $3d$ are small. This comparison highlights the limitation of a local or semi-local exchange potential in capturing the exchange interaction between a core hole and photoelectron. While small in the case of the K edge,

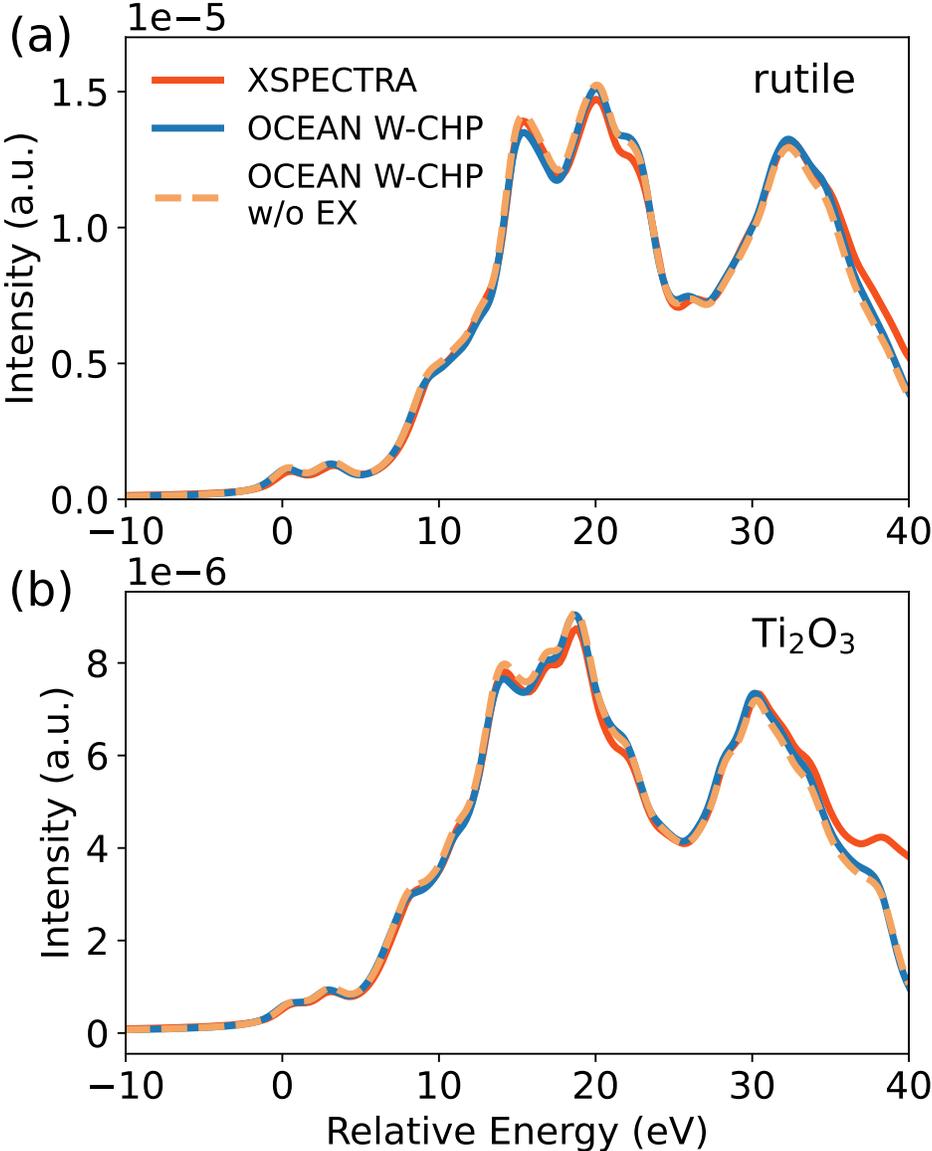


FIG. 7. Comparison of XAS spectra of TiO_2 (rutile, mp-2657) (a) and Ti_2O_3 (mp-458) (b) between XSPECTRA and OCEAN using a screened core-hole based on the CHP method (W-CHP) with and without the exchange interaction (W-CHP w/o EX).

the exchange term is vital for reproducing even qualitatively correct spectra for $3d$ transition metal L edges¹⁰⁶. The contribution of the exchange is associated with the local field effect of the dielectric response, which could play an important role in low-dimensional systems.

XSPECTRA and W-CHP w/o EX are the same level of theory, and, as we see in Fig. 7, they produce nearly perfect agreement in rutile (mp-2657) and Ti_2O_3 (mp-458), except the small intensity differences at the second peak and the shoulder a couple of eV higher in energy. This indicates that, for this system, the unoccupied $4p$ orbitals that are probed in XAS are relatively unaffected by the self-consistent relaxation of the valence bands in the presence of the core hole. In the pre-edge region there is some evidence of these relaxation effects, where the intensities calculated by XSPECTRA might be suppressed by increased occupation of the on-site $3d$ orbitals due to the excited core electron.

Next we compare the three different methods for generating the screened core-hole potential W introduced above: W-RPA, W-ALDA, and W-CHP. In Fig. 8a and 8b, we show the three different induced potentials for rutile (mp-2657: insulator) and Ti_2O_3 (mp-458: metal). Both the CHP and ALDA induced potentials are noticeably stronger than the RPA one, leading to a reduction in the strength of the screened core-hole potential. This causes a reduction in strength of the resulting spectra near threshold. In the case of rutile TiO_2 , visible discrepancies can be seen in

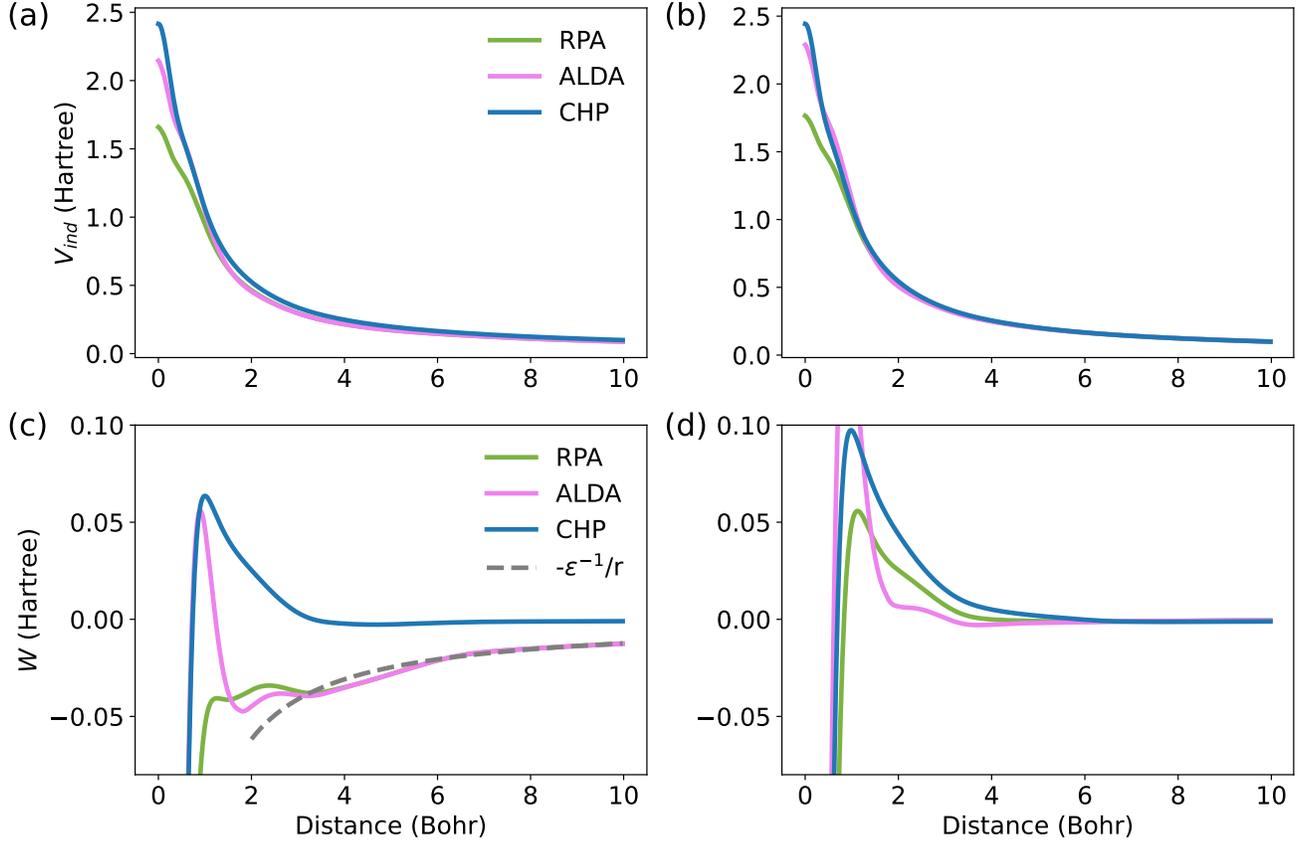


FIG. 8. Comparison of the induced potentials in response to a Ti $1s$ core hole in (a) rutile TiO_2 (mp-2657) and (b) Ti_2O_3 (mp-458) using three different methods for determining the induced density response. The RPA and CHP calculations of the induced potentials for the metallic Ti_2O_3 are in closer agreement than for TiO_2 . The total screened core-hole potential is shown for (c) TiO_2 and (d) Ti_2O_3 . The plots are zoomed in to show detail, with the full core-hole potential approaching 18 Hartree at the origin. For the metallic Ti_2O_3 , all three approaches efficiently screen the core-hole potential, resulting in a total potential that is nearly zero at intermediate ranges. In the case of the insulating TiO_2 , the two linear-response approaches both give the expected $-\epsilon_\infty^{-1}/r$ behavior at large distance. The grey dashed guide line uses the isotropic $\epsilon_\infty = 8.1$ taken from the Materials Project⁷⁶.

the intermediate range up to around 6 Bohr, while in the metallic Ti_2O_3 (mp-458), the three approximations for the induced potential are all in agreement starting relatively close to the core hole, at approximately 1.5 Bohr. We further examine the trend in the total potential W in Fig. 8c and 8d. For the insulating TiO_2 case, the CHP approximation has a finite range, as the excited electron is retained in this approximation leading to local neutrality near the excited Ti site. In contrast, the BSE calculations consider the core hole charge explicitly, leading to the expected long range limit of $-\epsilon_\infty^{-1}/r$. In the metallic Ti_2O_3 (mp-458), as expected, all three total potentials are short-ranged.

In Fig. 9a we show a comparison of the polarization-averaged spectra for rutile TiO_2 under the dipole approximation, calculated with XSPECTRA as well as the three different approximations of W within OCEAN. Both of the modified OCEAN calculations (W-ALDA and W-CHP) show better agreement with XSPECTRA, where the second peak has higher intensity as opposed to the previous W-RPA OCEAN calculations (see Fig 6). Compared to W-RPA, the ALDA kernel has a non-negligible effect most noticeably at the first peak and the shoulder at ≈ 10 eV, which shifts the spectral intensity to higher energy. The W-CHP spectrum exhibits an even lower intensity at the first peak and the shoulder than the W-ALDA spectrum, which can be attributed to the stronger screening potential under the self-consistent relaxation treatment at the presence of a full core hole than that from the linear response treatment as shown in Fig. 8a. The same comparison is shown for the metallic Ti_2O_3 (mp-458) in Fig. 9b, and similar trends are observed but to a lesser extent, due to the smaller discrepancy in the screening potential in the metallic system as shown in Fig. 8b. Previous work showed that the exciton peak position differed between BSE and CHP calculations for the Li K edge of lithium halides¹⁰⁷. We do not see the same behavior here – in Fig. 9 peaks from different methods at the edge (≈ 15 eV) and above the edge (30 eV) are very well aligned. This is due to the much weaker excitonic effects seen in the Ti K edge of titanium oxides, as evidenced by the similarity between the interacting and non-interacting

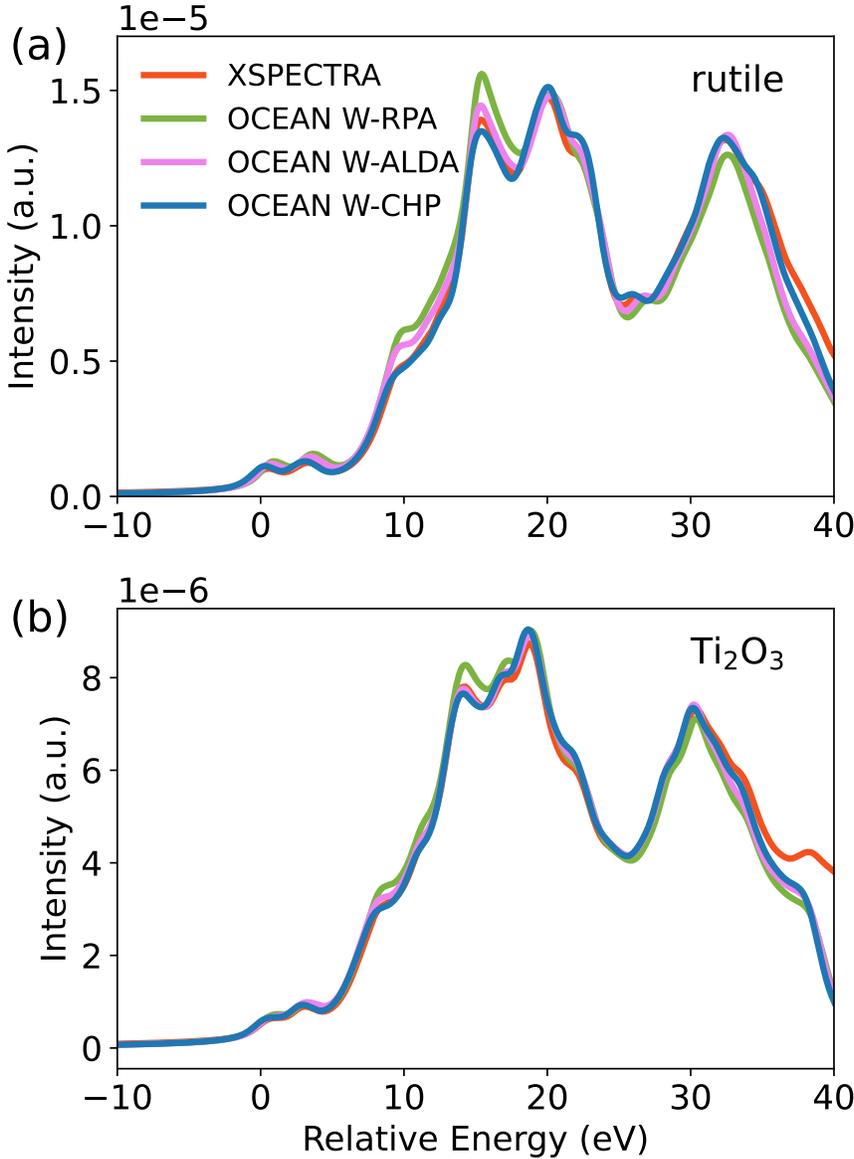


FIG. 9. Comparison of the XAS of (a) rutile (mp-2657) and (b) Ti_2O_3 (mp-458) calculated using XSPECTRA and three different OCEAN calculations under different approximations for the screened Coulomb interaction. The metallic Ti_2O_3 shows the same trends as the insulating rutile TiO_2 , though the differences between the W-ALDA and W-CHP appear smaller.

spectra (Fig. 6) in contrast to the lithium halides (Fig. 2 of Ref. 107).

The W-CHP approximation yields a near-perfect screening of the core hole for rutile TiO_2 (Fig. 8(c)), due to the self-consistent density response as expected for an ionic system, where compensating charge resides closely around the absorbing titanium atom¹⁰⁸. However, as seen in Fig. 9(a), the calculations that use this potential (XSPECTRA and OCEAN W-CHP) are largely in agreement with the calculations whose screened potential has the correct long-range form (W-RPA and W-ALDA). In addition, practical CHP calculations are also performed with a core hole whose occupancy (strength) can be varied between 1 (full core hole) and 1/2 (half core hole). The applicability of these approximations is system dependent, but the CHP method we use here enjoys success for both metallic and semi-conducting systems³⁹.

F. Comparison with Experiment

In this section, we compare the simulated spectra of rutile (mp-2657) and anatase (mp-390) to measurements from Carta *et al.*¹⁰⁹. While both the CHP and BSE methods attempt to simulate the full x-ray spectra, we note that there are several important effects that are not considered. These include the physical effects of the quasiparticle lifetime broadening, the effect of vibrational disorder, self-energy corrections, and satellite features. Below we summarize several limitations of this comparison. First, our calculations are carried out on perfectly ordered structures with the atoms fixed during the x-ray excitation. Including vibrational disorder has been shown to reveal symmetry-forbidden, but experimentally observed, “dark states,” suppress expected states^{46–48}, or change the intensities of pre-edge features^{48,49}, while excited-state vibrational coupling can change the shape and energies of peaks⁵⁰. A second large approximation is the use of orbitals from semi-local DFT as quasiparticle excitations of the x-ray excited state, neglecting many-body effects that could be incorporated with more accurate functionals⁵¹, *GW* self-energy corrections^{52,53}, the cumulant approximation⁵⁴, and multi-configuration methods¹¹⁰. Lastly, the approaches we use here are limited to a single electron or single electron–core-hole pair approximation. It is well known that this approximation can fail for $L_{2,3}$ edges ($2p \rightarrow 3d$ transitions) where local dd^* or ligand-to-metal charge transfer excitations play an important role⁵⁵. This is less of an issue at the K edge due to the small quadrupole intensity and the reduced exchange interaction between the $3d$ electron and the $1s$ core hole as compared to a $2p$ hole⁵⁶.

Nevertheless, the comparison conducted here can provide the reader with a basic calibration for the degree to which the present methods capture the main features in experiment over the full XANES photon energy range. With this foundation, future investigations can incorporate further physical effects, building on prior work as well^{46–56,86}. For example, the effects of vibrational disorder can be studied by statistically averaging atomic structures under a given temperature using the workflow developed in this study for each snapshot⁴⁶. A systematic benchmark study against experiment for a broader range of Titanium oxide polymorphs is the subject of future work.

In this comparison, we present results including both dipole and quadrupole contributions from only OCEAN and XSPECTRA because the results from the two BSE codes have been shown to be nearly identical. Since both rutile and anatase belong to the $4/mmm$ point group, the powder average of the quadrupole term is performed by a weighted average of three specially chosen photon orientations ($\{\mathbf{e}, \mathbf{q}\}$)¹¹¹. The choice of photon orientations is given in Appendix C.

The comparisons between theory and experiment for rutile and anatase TiO_2 are shown in Figs. 10 and 11. The simulated spectra are aligned to the experiment at the first peak of the main edge near 4986 eV. In rutile, the experimental spectrum shows a shoulder peak around 4981 eV and three peaks in the main edge region at 4986.6 eV, 4991.8 eV, and 5004.0 eV. These main spectral features are well reproduced in both OCEAN and XSPECTRA. The energy separation between the first two peaks are underestimated in the simulation by about 0.3 eV, likely due to the missing the self-energy correction at the PBE level of theory. For anatase, both simulation methods reproduce the main features, including the shoulder peak around 4980 eV, the pronounced main edge at 4986.8 eV and the three-peak feature between 4993 eV and 5005 eV.

In the rutile experimental spectrum, the second peak has higher intensity than the first one. While XSPECTRA shows a similar pattern, it is the opposite in OCEAN. This discrepancy between XSPECTRA and OCEAN can be attributed to the different approximations for the screening potential. As discussed in detail in the previous section, the screening potential in RPA is weaker than in the core-hole potential method, which results in stronger electron–core-hole coupling that redistributes oscillator strength towards lower energy in OCEAN as compared to XSPECTRA. There is a similar trend in anatase, where the OCEAN spectrum produces stronger low energy features than XSPECTRA, but to a lesser degree than rutile. However, it has been shown that when the shake-up satellites are considered as a post treatment of BSE by the convolution with the cumulant spectral function¹¹², the overstated BSE intensity of the excitonically enhanced $\text{Ti } 1s \rightarrow 4p$ transition in rutile is reduced and the resulting spectrum is in much better agreement with experiment⁵⁴. Therefore one needs to be cautious about drawing quantitative conclusions when comparing theory to experiment on Ti K-edge XANES without considering the effects of satellites. Nevertheless, a systematic study of the satellite effects is beyond the scope of the current work.

The Ti K-edge XANES pre-edge features of rutile and anatase have been studied extensively^{2,54,60,113–124}, and the three pre-edge peaks have been assigned to 1) the dipole forbidden $\text{Ti } 1s \rightarrow 3d t_{2g}$ and 2) $1s \rightarrow 3d e_g$ transitions under the crystal field splitting and 3) the $\text{Ti } 1s$ to nearest neighbor $\text{Ti } 3d$ transitions mediated by the hybridization with ligand O $2p$ orbitals. The comparisons of the pre-edge features in rutile and anatase between theory and experiment are shown at the inset of Figs. 10 and 11. Overall the pre-edge peaks are more pronounced in OCEAN than XSPECTRA, as the stronger electron - core hole coupling in BSE sharpens the pre-edge features that have significant excitonic character. OCEAN also shows better agreement with experiment in the pre-edge peak positions than XSPECTRA. A major discrepancy between theory and experiment is that the first pre-edge peak in rutile at 4968.6 eV is almost missing in both OCEAN and XSPECTRA. The first pre-edge peak has a quadrupole nature, which is an order of magnitude smaller than the dipole contribution, as shown in Figs. S4 and S5. The Ti octahedron is nearly ideal in

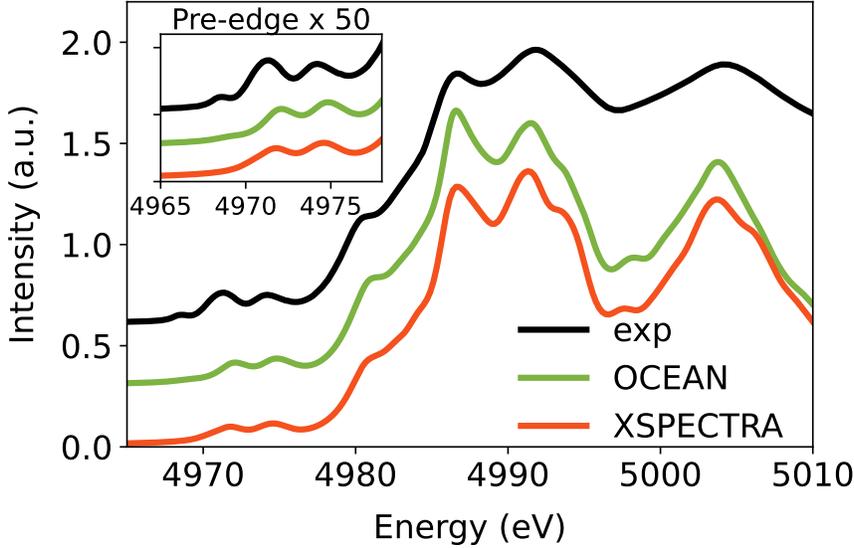


FIG. 10. Comparison of Ti K-edge XANES of rutile (mp-2657) between theory (OCEAN and XSPECTRA) and experiment¹⁰⁹. Inset shows the pre-edge magnified by fifty fold.

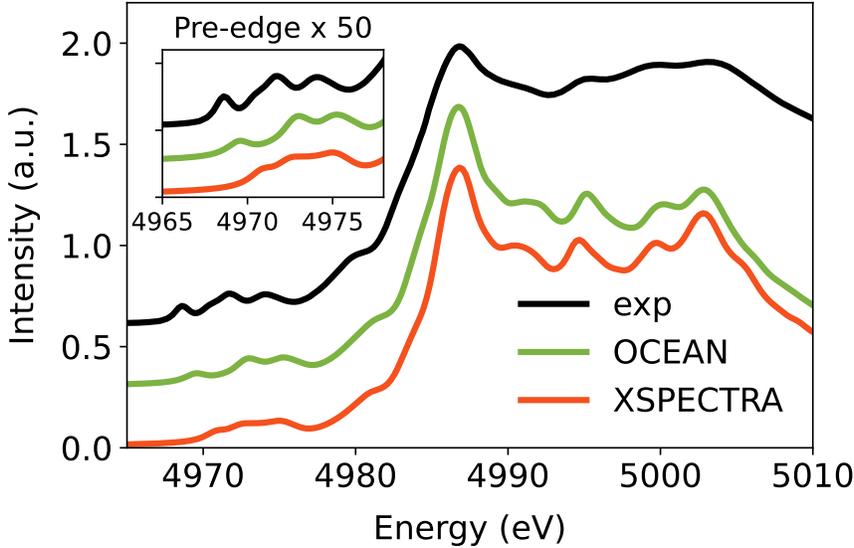


FIG. 11. Comparison of Ti K-edge XANES of anatase (mp-390) between theory (OCEAN and XSPECTRA) and experiment¹⁰⁹. Inset shows the pre-edge magnified by fifty fold.

rutile, but has a large bond angle distortion in anatase with two O-Ti-O bond angles at 154.67 degree in the mp-390 structure. As a result, the intensity of the Ti $1s \rightarrow 3d t_{2g}$ excitation in rutile is negligible in simulated spectra based on the zero temperature structure. The temperature effects on XANES pre-edge features have been discussed in the literature^{49,125-127}. In particular, Cockayne *et al.* showed that the $t_{1u}^{(1)}$ phonon mode, which moves the Ti^{4+} relative to its axial O neighbors, has a strong spectroscopic signature⁴⁹. This centrosymmetry-breaking distortion can cause $p - d$ mixing, which results in a finite pre-edge peak comparable to experiment⁴⁹.

While important physical effects are visible in certain spectral features such as pre-edge peaks and relative peak intensities, the present methods represent the measured XANES spectra quite well over the 35 eV energy range. The predictive power of first-principles XANES simulations in the whole spectral range is crucial to spectral analysis that employs fingerprinting and modern ML-based methods.

G. Edge Alignment

As discussed in Sec. II C all three codes make different approximations in calculating the core-level shifts, and none of the codes provides an absolute energy scale that can be compared to experiment without correction. All three attempt to provide correct relative energy shifts, using different levels of theory. Correct relative energy shifts means that a single parameter can be used to align the calculated Ti K-edge XAS from one code with experimental measurements, or, in our case, a single parameter can be used to align the calculations from two different codes. However, we would not expect this single alignment to work equally well across all systems due to differences in theory and implementation between codes. In this section, we examine the differences to assess the key approximations that the three codes make.

We compare both the alignment of the XAS spectra and the $1s$ removal energies as explained in Sec. II C. For each pair of codes, system, and unique Ti site, we calculate the optimal relative alignment of the x-ray spectra. The simulated spectra are truncated using an energy range of 35 eV from the onset (defined as the 2% of the maximum intensity) to make sure that they cover similar energy ranges. We treat each spectrum as a vector. The similarity between two spectra (μ_1 and μ_2) are characterized by their Cosine similarity, $\cos(\theta) = \frac{\mu_1 \cdot \mu_2}{\|\mu_1\| \|\mu_2\|}$. We align spectra from two codes using the optimal relative shift, which yields the highest Cosine similarity. We subtract out the average relative alignment – the best fit single alignment parameter – and report the system-by-system deviation. For the $1s$ removal energies, no fitting is necessary because we are directly comparing energies. However, as for the spectra, the use of pseudopotentials means that only the relative shifts are meaningful.

First we compare the two BSE code OCEAN and `exciting` as shown in Table VI for (final state) spectra and $1s$ removal energies. For both, the two codes are using the same level of theory, except for different approximations in implementation. The $1s$ removal energies are calculated at the level of DFT, and they are given by the energy difference between the conduction band minimum (Fermi level for metals) and the energy of the $1s$ orbital. The two codes agree to ± 0.13 eV. We attribute this to the use of the frozen-core approximation in OCEAN. As a check, we found that the agreement in shifts of the non-interacting XAS is similar at ± 0.14 eV (see Table S1). Given the previously shown agreement in band structure (Table IV and Fig. 3), the non-interacting XAS comparison is primarily a measure of the $1s$ energies. In the case of the interacting XAS, the agreement in the relative shifts drops to ± 0.20 eV. Treating these errors as uncorrelated, the differences in the BSE implementation between OCEAN and `exciting` is responsible for approximately a ± 0.15 eV discrepancy in the core-level shifts. Like the spectral comparisons in Sec. IV D, this shows that the two codes are in close agreement despite differences in basis sets and the treatment of core-level electrons.

TABLE VI. The difference in the relative core-level shifts in eV between OCEAN and `exciting` for both $1s$ removal and XAS spectra. Both the frozen-core approximation in OCEAN as well as the differences in the BSE implementation between the two codes are found to lead to small, but significant differences between the two codes. A negative number indicates that the OCEAN spectrum is blue-shifted, relative to the `exciting` one. σ denotes the standard deviation of the differences.

mpid:site	XAS	$1s$ removal
mp-390	-0.17	-0.12
mp-2657	-0.23	-0.15
mp-1840	0.14	-0.14
mp-430	0.24	-0.12
mvc-11115:0	-0.33	-0.09
mvc-11115:1	-0.34	-0.16
mp-1203:0	0.13	0.10
mp-1203:2	0.11	0.10
mp-1203:4	-0.04	-0.02
mp-10734	0.14	0.19
mp-1215	0.10	0.11
mp-2664	0.20	0.12
mp-458	0.05	0.18
σ	0.20	0.13

Next, we turn to the comparison between CHP (XSPECTRA) and BSE (OCEAN) codes. We start with the calculation of the $1s$ removal energy. This is a component of the shift seen in XAS (along with the energy cost of adding the core level electron to the conduction bands and the interaction energy of the exciton). In the case of XSPECTRA, the $1s$ removal energy is obtained from the Δ SCF procedure explained in Sec. II C. The core-excited final state has a single Ti $1s$ core hole using the same core-hole pseudopotential as the spectral calculations. The core-excited state nominally

has a net +1 charge, which, with periodic DFT, is neutralized by a uniform background charge. The effects of both the uniform background charge and the periodic images of the core hole fall off with supercell size. We extrapolate the calculated core-level removal energies (at the minimum supercell size of 9 Å and 13 Å) to the infinite supercell limit. Note that the neutral excitation of the CHP method for x-ray spectra converges at smaller cell sizes, and as mentioned in Sec. II B, a 9 Å minimum lattice vector was found to be sufficient. In OCEAN, the core-level removal energy is estimated as outlined in Sec. II C as the Hartree potential minus $1/2W_c$ to account for the relaxation of the system due to the removal of the core electron.

From Table VII, we find that the two methods give reasonably similar 1s removal energies with a standard deviation of only ± 0.19 eV. There is a slight bimodal grouping of the results by metal (first six rows) and non-metal (last seven rows). For non-metals, the estimate of the 1s removal energy from XSPECTRA tends to be smaller than that of OCEAN by 0.29 eV. This could indicate a sensitivity to the model dielectric that is used in OCEAN to model the long-range response, different convergence behavior with respect to cell size of the $N - 1$ core-hole calculation in XSPECTRA, or it could reflect differences between linear-response and self-consistent calculations of the density response.

Finally, we compare the spectral shifts in XANES between OCEAN and XSPECTRA, and we again find reasonable agreement between the codes and methods with a standard deviation of ± 0.22 eV. This would suggest that the contribution due to the different treatments of the final state effects is about ± 0.11 eV, but we can not rule out some cancellation of errors.

TABLE VII. The difference in the relative core-level shifts in eV between OCEAN and XSPECTRA for both XAS and 1s removal. Unlike the comparison between BSE codes in Table VI, here the OCEAN results include the self-energy correction to the core-hole energy (see text). The sign follows the same convention as in Table VI. σ denotes the standard deviation of the differences.

mpid:site	XAS	1s removal
mp-390	-0.06	-0.25
mp-2657	0.14	-0.17
mp-1840	-0.16	-0.30
mp-430	-0.36	-0.30
mvc-11115:0	-0.33	0.08
mvc-11115:1	-0.23	0.00
mp-1203:0	0.11	0.04
mp-1203:2	-0.04	0.18
mp-1203:4	0.17	0.25
mp-10734	0.31	0.23
mp-1215	0.09	0.09
mp-2664	0.02	0.13
mp-458	0.34	0.01
σ	0.22	0.19

V. CONCLUSIONS

We have developed an automated workflow to generate and standardize input files for simulating x-ray absorption spectra using three widely used codes: OCEAN, `exciting`, and XSPECTRA. By carefully converging the disparate settings for each of the three codes, the workflow ensures that the resulting spectra are free from numerical artifacts unrelated to the underlying theory or implementation details. We carried out quantitative comparisons between Ti K-edge XANES spectra from the three codes using ten representative titanium oxide compounds, referred to as the Ti-O-10 dataset.

We found that the two BSE-based codes, `exciting` and OCEAN, produce spectra measurably closer to each other than to the core-hole potential code XSPECTRA. We quantified the similarity between spectra using Spearman’s rank correlation score. We found that the average score between the two BSE codes is 0.998, while between either BSE code and XSPECTRA it is 0.990. While this difference appears small, it corresponds to visible differences in the spectra (see Fig. 6). Nonetheless, the overall spectra are quite similar. From these results, we draw several conclusions. First, despite significant implementation differences between the all-electron `exciting` and pseudopotential-based OCEAN, they produce nearly identical spectra, suggesting that both implementations are robust. Second, the BSE and CHP methods produce quite close agreement in main features and peak positions, though differences are noticeable,

especially in the relative spectral intensities. Such discrepancies in spectral shape can be primarily attributed to the difference in the strength of the screened core-hole potential. Carrying out BSE calculations using the ALDA instead of RPA to screen the core hole greatly reduces the difference between the CHP and BSE results.

We also conducted a theory (OCEAN and XSPECTRA) to experiment comparison for rutile and anatase TiO_2 , where the quadrupole contributions are included to give a more complete description of pre-edge features. Other effects, such as thermal disorder and satellite effects, although very important for a thorough comparison, are subject of future work. With these limitations, our main findings are the following. First, both OCEAN and XSPECTRA reproduce quite well the main features of the experimental spectra in a wide energy range of about 35 eV, including the number of peaks, the overall spectral shape, and the positions of main and shoulder peaks. Second, XSPECTRA exhibits a seemingly better match to the relative intensity of the double-peak feature in rutile compared to OCEAN. We found that the electron-core hole coupling is stronger in BSE under a weaker dielectric screening than CHP. As a result, the oscillator strength is redistributed more towards low energy in OCEAN than XSPECTRA. This should be considered in the context of studies showing the impact of other physical effects. As shown by Woicik *et al.*⁵⁴, including the satellite effects using the cumulant method can modify the relative intensity of the two peaks and yield a much better agreement between OCEAN and experiment. On the other hand, the cumulant correction is expected to worsen the agreement between XSPECTRA and experiment. Finally, overall OCEAN pre-edge features are more pronounced and agree better with experiment than XSPECTRA. However, the $\text{Ti } 1s \rightarrow 3d t_{2g}$ peak in rutile is severely underestimated in both codes, due to the lack of the thermal disorder correction. This discussion underscores the importance of thoroughly accounting for additional effects such as vibrational disorder and satellites. On a material by material basis these have been shown to change positions and weights of spectral features to a degree comparable to the code- and method-dependent differences we have shown here.

Finally, we compared the relative edge alignment among the three codes. Between OCEAN and `exciting`, the standard deviation is 0.20 eV in the XAS spectra. This difference is partially due to the frozen core hole approximation used in OCEAN, which gives rise a the standard deviation of 0.13 eV at the Kohn-Sham DFT level. Between OCEAN and XSPECTRA, the standard deviation in the $1s$ removal energy is 0.19 eV. The standard deviation in the XAS spectra is 0.22 eV, though the two codes have quite different treatments for the self-energy correction and final state effects.

This study supports broad application of first-principles simulation for x-ray spectral analysis, in particular, modern data-driven methods that take into account the full spectral range of XANES. The automated workflow and the heuristics to achieve the numerical convergence provide a good standard for first-principles x-ray spectral simulations. Furthermore, the workflow can play a big role in developing simulated XAS spectral databases using high-throughput computing.

VI. DATA AVAILABILITY

Data of the benchmark study, including input files, output files, and metadata¹²⁸, can be downloaded from the Novel Materials Discovery (NOMAD) Laboratory¹²⁹ under the CC-BY-4.0 license.

VII. ACKNOWLEDGMENT

We thank Dr. Daniela Carta and Dr. Themistoklis Prodromakis for providing us their measured Ti K-edge XANES spectra of rutile and anatase TiO_2 . F.M., X.Q. and D.L. thank Eli Stavitski and Bruce Ravel for helpful discussion. This work is partially supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Numbers FWP PS-030. The research used the theory and computation resources of the Center for Functional Nanomaterials, which is a U.S. DOE Office of Science Facility, and the Scientific Data and Computing Center, a component of the Computational Science Initiative, at Brookhaven National Laboratory under Contract No. DE-SC0012704. This research also used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAP-20690, 18006 and 14811. This work received partial funding by the German Research Foundation (DFG) through the CRC 1404 (FONDA), Projektnummer 414984028, and the NFDI consortium FAIRmat – project 46019701. Certain software packages are identified to facilitate understanding. Such identification does not imply recommendations or endorsement by NIST.

* vorwerk@uchicago.edu; Also at Institut für Physik and IRIS Adlershof, Humboldt-Universität zu Berlin, Berlin Germany

- [†] claudia.draxl@physik.hu-berlin.de; Also at European Theoretical Spectroscopy Facility (ETSF)
- [‡] john.vinson@nist.gov
- [§] dlu@bnl.gov
- ¹ John J Rehr and Robert C Albers, “Theoretical approaches to x-ray absorption fine structure,” *Reviews of Modern Physics* **72**, 621 (2000).
 - ² Takashi Yamamoto, “Assignment of pre-edge peaks in k-edge x-ray absorption spectra of 3d transition metal compounds: Electric dipole or quadrupole?” *X-Ray Spectrometry: An International Journal* **37**, 572–584 (2008).
 - ³ T Saitoh, AE Bocquet, T Mizokawa, H Namatame, A Fujimori, M Abbate, Y Takeda, and M Takano, “Electronic structure of La 1-x sr x mno 3 studied by photoemission and x-ray-absorption spectroscopy,” *Physical Review B* **51**, 13942 (1995).
 - ⁴ David Prendergast and Giulia Galli, “X-ray absorption spectra of water from first principles calculations,” *Physical Review Letters* **96**, 215502 (2006).
 - ⁵ Wei Zhang, Mehmet Topsakal, Christina Cama, Christopher J Pelliccione, Hu Zhao, Steven Ehrlich, Lijun Wu, Yimei Zhu, Anatoly I Frenkel, Kenneth J Takeuchi, *et al.*, “Multi-stage structural transformations in zero-strain lithium titanate unveiled by in situ x-ray absorption fingerprints,” *Journal of the American Chemical Society* **139**, 16591–16603 (2017).
 - ⁶ Y Li, D Zakharov, S Zhao, R Tappero, U Jung, A Elsen, Ph Baumann, Ralph G Nuzzo, EA Stach, and AI Frenkel, “Complex structural dynamics of nanocatalysts revealed in operando conditions by correlated imaging and spectroscopy probes,” *Nature Communications* **6**, 1–6 (2015).
 - ⁷ Junko Yano and Vittal K Yachandra, “X-ray absorption spectroscopy,” *Photosynthesis Research* **102**, 241–254 (2009).
 - ⁸ Frank De Groot and Akio Kotani, *Core level spectroscopy of solids* (CRC press, 2008).
 - ⁹ M. W. Haverkort, M. Zwierzycki, and O. K. Andersen, “Multiplet ligand-field theory using wannier orbitals,” *Phys. Rev. B* **85**, 165113 (2012).
 - ¹⁰ Yufeng Liang, John Vinson, Sri Pemmaraju, Walter S. Drisdell, Eric L. Shirley, and David Prendergast, “Accurate x-ray spectral predictions: An advanced self-consistent-field approach inspired by many-body perturbation theory,” *Phys. Rev. Lett.* **118**, 096402 (2017).
 - ¹¹ Fujie Tang, Zhenglu Li, Chunyi Zhang, Steven G Louie, Roberto Car, Diana Y Qiu, and Xifan Wu, “Many-body effects in the x-ray absorption spectra of liquid water,” *Proceedings of the National Academy of Sciences* **119**, e2201258119 (2022).
 - ¹² Zhantao Chen, Nina Andrejevic, Nathan C. Drucker, Thanh Nguyen, R. Patrick Xian, Tess Smidt, Yao Wang, Ralph Ernstorfer, D. Alan Tennant, Maria Chan, and Mingda Li, “Machine learning on neutron and x-ray scattering and spectroscopies,” *Chemical Physics Reviews* **2**, 031301 (2021).
 - ¹³ Janis Timoshenko and Anatoly I Frenkel, ““inverting” x-ray absorption spectra of catalysts by machine learning in search for activity descriptors,” *Acs Catalysis* **9**, 10192–10211 (2019).
 - ¹⁴ Janis Timoshenko, Deyu Lu, Yuewei Lin, and Anatoly I. Frenkel, “Supervised machine-learning-based determination of three-dimensional structure of metallic nanoparticles,” *The Journal of Physical Chemistry Letters* **8**, 5091–5098 (2017), PMID: 28960990.
 - ¹⁵ Nicholas Marcella, Yang Liu, Janis Timoshenko, Erjia Guan, Mathilde Luneau, Tanya Shirman, Anna M Plonka, Jessi ES van der Hoeven, Joanna Aizenberg, Cynthia M Friend, *et al.*, “Neural network assisted analysis of bimetallic nanocatalysts using x-ray absorption near edge structure spectroscopy,” *Physical Chemistry Chemical Physics* **22**, 18902–18910 (2020).
 - ¹⁶ Yang Liu, Avik Halder, Soenke Seifert, Nicholas Marcella, Stefan Vajda, and Anatoly I Frenkel, “Probing active sites in Cu x Pd y cluster catalysts by machine-learning-assisted x-ray absorption spectroscopy,” *ACS Applied Materials & Interfaces* **13**, 53363–53374 (2021).
 - ¹⁷ Yang Liu, Nicholas Marcella, Janis Timoshenko, Avik Halder, Bing Yang, Lakshmi Kolipaka, Michael J Pellin, Soenke Seifert, Stefan Vajda, Ping Liu, *et al.*, “Mapping xanes spectra on structural descriptors of copper oxide clusters using supervised machine learning,” *The Journal of Chemical Physics* **151**, 164201 (2019).
 - ¹⁸ Alexander A Guda, Sergey A Guda, Andrea Martini, AN Kravtsova, Alexander Algasov, Aram Bugaev, SP Kubrin, LV Guda, P Šot, JA van Bokhoven, *et al.*, “Understanding x-ray absorption spectra by means of descriptors and machine learning algorithms,” *npj Computational Materials* **7**, 1–13 (2021).
 - ¹⁹ Liang Li, Mindren Lu, and Maria KY Chan, “A deep learning model for atomic structures prediction using x-ray absorption spectroscopic data,” *arXiv preprint arXiv:1905.03928* (2019), 10.48550/arXiv.1905.03928.
 - ²⁰ Anja Aarva, Volker L Deringer, Sami Sainio, Tomi Laurila, and Miguel A Caro, “Understanding x-ray spectroscopy of carbonaceous materials by combining experiments, density functional theory, and machine learning. part i: Fingerprint spectra,” *Chemistry of Materials* **31**, 9243–9255 (2019).
 - ²¹ Anja Aarva, Volker L. Deringer, Sami Sainio, Tomi Laurila, and Miguel A. Caro, “Understanding x-ray spectroscopy of carbonaceous materials by combining experiments, density functional theory, and machine learning. part ii: Quantitative fitting of spectra,” *Chemistry of Materials* **31**, 9256–9267 (2019).
 - ²² Prahlad K Routh, Yang Liu, Nicholas Marcella, Boris Kozinsky, and Anatoly I Frenkel, “Latent representation learning for structural characterization of catalysts,” *The Journal of Physical Chemistry Letters* **12**, 2086–2094 (2021).
 - ²³ Shuting Xiang, Peipei Huang, Junying Li, Yang Liu, Nicholas Marcella, Prahlad K Routh, Gonghu Li, and Anatoly I Frenkel, “Solving the structure of “single-atom” catalysts using machine learning–assisted xanes analysis,” *Physical Chemistry Chemical Physics* **24**, 5116–5124 (2022).
 - ²⁴ Orlando Trejo, Anup L Dadlani, Francisco De La Paz, Shinjita Acharya, Rob Kravec, Dennis Nordlund, Ritimukta Sarangi, Fritz B Prinz, Jan Torgersen, and Neil P Dasgupta, “Elucidating the evolving atomic structure in atomic layer deposition reactions with in situ xanes and machine learning,” *Chemistry of Materials* **31**, 8937–8947 (2019).
 - ²⁵ Matthew R Carbone, Mehmet Topsakal, Deyu Lu, and Shinjae Yoo, “Machine-learning x-ray absorption spectra to quantitative accuracy,” *Physical Review Letters* **124**, 156401 (2020).

- ²⁶ Samantha Tetef, Niranjana Govind, and Gerald T Seidler, “Unsupervised machine learning for unbiased chemical classification in x-ray absorption spectroscopy and x-ray emission spectroscopy,” *Physical Chemistry Chemical Physics* **23**, 23586–23601 (2021).
- ²⁷ Matthew R Carbone, Shinjae Yoo, Mehmet Topsakal, and Deyu Lu, “Classification of local chemical environments from x-ray absorption spectra using supervised machine learning,” *Physical Review Materials* **3**, 033604 (2019).
- ²⁸ Steven B. Torrisi, Matthew R. Carbone, Brian A. Rohr, Joseph H. Montoya, Yang Ha, Junko Yano, Santosh K. Suram, and Linda Hung, “Random forest machine learning models for interpretable x-ray absorption near-edge structure spectrum-property relationships,” *npj Computational Materials* **6**, 109 (2020).
- ²⁹ Chen Zheng, Chi Chen, Yiming Chen, and Shyue Ping Ong, “Random forest models for accurate identification of coordination environments from x-ray absorption near-edge structure,” *Patterns* **1**, 100013 (2020).
- ³⁰ Conor D Rankine, Marwah MM Madkhali, and Thomas J Penfold, “A deep neural network for the rapid prediction of x-ray absorption spectra,” *The Journal of Physical Chemistry A* **124**, 4263–4270 (2020).
- ³¹ Danhua Yan, Mehmet Topsakal, Sencer Selcuk, John L Lyons, Wenrui Zhang, Qiyuan Wu, Iradwikanari Waluyo, Eli Stavitski, Klaus Attenkofer, Shinjae Yoo, *et al.*, “Ultrathin amorphous titania on nanowires: Optimization of conformal growth and elucidation of atomic-scale motifs,” *Nano letters* **19**, 3457–3463 (2019).
- ³² Martin Kuban, Santiago Rigamonti, Markus Scheidgen, and Claudia Draxl, “Density-of-states similarity descriptor for unsupervised learning from materials data,” *Sci. Data* **9**, 646 (2022).
- ³³ Martin Kuban, Šimon Gabaj, Wahib Aggoune, Cecilia Vona, Santiago Rigamonti, and Claudia Draxl, “Similarity of materials and data-quality assessment by fingerprinting,” *MRS Bulletin* **47**, 1 (2022).
- ³⁴ Conor Douglas Rankine and TJ Penfold, “Accurate, affordable, and generalizable machine learning simulations of transition metal x-ray absorption spectra using the xanesnet deep neural network,” *The Journal of Chemical Physics* **156**, 164102 (2022).
- ³⁵ Animesh Ghose, Mikhail Segal, Fanchen Meng, Zhu Liang, Mark S. Hybertsen, Xiaohui Qu, Eli Stavitski, Shinjae Yoo, Deyu Lu, and Matthew R. Carbone, “Uncertainty-aware predictions of molecular x-ray absorption spectra using neural network ensembles,” *Phys. Rev. Res.* **5**, 013180 (2023).
- ³⁶ Kiran Mathew, Chen Zheng, Donald Winston, Chi Chen, Alan Dozier, John J. Rehr, Shyue Ping Ong, and Kristin A. Persson, “High-throughput computational x-ray absorption spectroscopy,” *Scientific Data* **5**, 180151 (2018).
- ³⁷ Yiming Chen, Chi Chen, Chen Zheng, Shyam Dwaraknath, Matthew K. Horton, Jordi Cabana, John Rehr, John Vinson, Alan Dozier, Joshua J. Kas, Kristin A. Persson, and Shyue Ping Ong, “Database of ab initio l-edge x-ray absorption near edge structure,” *Scientific Data* **8**, 153 (2021).
- ³⁸ Kiyohito Shibata, Kakeru Kikumasa, Shin Kiyohara, and Teruyasu Mizoguchi, “Simulated carbon k edge spectral database of organic molecules,” *Scientific Data* **9**, 214 (2022).
- ³⁹ Eric L Shirley, LGM Pettersson, and D Prendergast, “Core-hole potentials and related effects,” in *International Tables for Crystallography Volume I: X-ray absorption spectroscopy and related techniques*, edited by C. T. Chantler, F. Boscherini, and B. Bunker (Wiley Online Library, 2021).
- ⁴⁰ Claudia Draxl and Caterina Cocchi, “exciting core-level spectroscopy,” in *International Tables for Crystallography Volume I: X-ray absorption spectroscopy and related techniques*, edited by C. T. Chantler, F. Boscherini, and B. Bunker (Wiley Online Library, 2021).
- ⁴¹ JJ Rehr, JA Soininen, and Eric L Shirley, “Final-state rule vs the bethe-salpeter equation for deep-core x-ray absorption spectra,” *Physica Scripta* **2005**, 207 (2005).
- ⁴² F. Fossard, G. Hug, K. Gilmore, J. J. Kas, J. J. Rehr, F. D. Vila, and E. L. Shirley, “Quantitative first-principles calculations of valence and core excitation spectra of solid c_{60} ,” *Phys. Rev. B* **95**, 115112 (2017).
- ⁴³ Kurt Lejaeghere, Gustav Bihlmayer, Torbjörn Björkman, Peter Blaha, Stefan Blügel, Volker Blum, Damien Caliste, Ivano E Castelli, Stewart J Clark, Andrea Dal Corso, *et al.*, “Reproducibility in density functional theory calculations of solids,” *Science* **351**, aad3000 (2016).
- ⁴⁴ Michiel J van Setten, Fabio Caruso, Sahar Sharifzadeh, Xinguo Ren, Matthias Scheffler, Fang Liu, Johannes Lischner, Lin Lin, Jack R Deslippe, Steven G Louie, *et al.*, “Gw 100: Benchmarking g0w0 for molecular systems,” *Journal of Chemical Theory and Computation* **11**, 5665–5687 (2015).
- ⁴⁵ Fabio Caruso, Matthias Dauth, Michiel J Van Setten, and Patrick Rinke, “Benchmark of gw approaches for the gw 100 test set,” *Journal of Chemical Theory and Computation* **12**, 5076–5087 (2016).
- ⁴⁶ John Vinson, Terrence Jach, W. T. Elam, and J. D. Denlinger, “Origins of extreme broadening mechanisms in near-edge x-ray spectra of nitrogen compounds,” *Phys. Rev. B* **90**, 205207 (2014).
- ⁴⁷ Tod A. Pascal, Ulrike Boesenberg, Robert Kostecki, Thomas J. Richardson, Tsu-Chien Weng, Dimosthenis Sokaras, Dennis Nordlund, Eamon McDermott, Alexander Moewes, Jordi Cabana, and David Prendergast, “Finite temperature effects on the X-ray absorption spectra of lithium compounds: First-principles interpretation of X-ray Raman measurements,” *The Journal of Chemical Physics* **140** (2014), 10.1063/1.4856835, 034107.
- ⁴⁸ Christian Brouder, Delphine Cabaret, Amélie Juhin, and Philippe Sainctavit, “Effect of atomic vibrations on the x-ray absorption spectra at the k edge of al in α - al_2o_3 and of ti in tio_2 rutile,” *Phys. Rev. B* **81**, 115125 (2010).
- ⁴⁹ Eric Cockayne, Eric L. Shirley, Bruce Ravel, and Joseph C. Woicik, “Local atomic geometry and ti 1s near-edge spectra in $pbtio_3$ and $srtio_3$,” *Phys. Rev. B* **98**, 014111 (2018).
- ⁵⁰ Andrey Geondzhian and Keith Gilmore, “Demonstration of resonant inelastic x-ray scattering as a probe of exciton-phonon coupling,” *Phys. Rev. B* **98**, 214305 (2018).
- ⁵¹ John Vinson, “Faster exact exchange in periodic systems using single-precision arithmetic,” *The Journal of Chemical Physics* **153**, 204106 (2020).

- ⁵² John Vinson, Terrence Jach, Matthias Müller, Rainer Unterumsberger, and Burkhard Beckhoff, “Quasiparticle lifetime broadening in resonant x-ray scattering of nh_4no_3 ,” *Phys. Rev. B* **94**, 035163 (2016).
- ⁵³ Fujie Tang, Zhenglu Li, Chunyi Zhang, Steven G. Louie, Roberto Car, Diana Y. Qiu, and Xifan Wu, “Many-body effects in the x-ray absorption spectra of liquid water,” *Proceedings of the National Academy of Sciences* **119**, e2201258119 (2022).
- ⁵⁴ J. C. Woicik, C. Weiland, C. Jaye, D. A. Fischer, A. K. Rumaiz, E. L. Shirley, J. J. Kas, and J. J. Rehr, “Charge-transfer satellites and chemical bonding in photoemission and x-ray absorption of SrTiO_3 and rutile TiO_2 : Experiment and first-principles theory with general application to spectroscopic analysis,” *Phys. Rev. B* **101**, 245119 (2020).
- ⁵⁵ Frank M.F. de Groot, Hebatalla Elnaggar, Federica Frati, Ru pan Wang, Mario U. Delgado-Jaime, Michel van Veenendaal, Javier Fernandez-Rodriguez, Maurits W. Haverkort, Robert J. Green, Gerrit van der Laan, Yaroslav Kvashnin, Atsushi Hariki, Hidekazu Ikeno, Harry Ramanantoanina, Claude Daul, Bernard Delley, Michael Odelius, Marcus Lundberg, Oliver Kuhn, Sergey I. Bokarev, Eric Shirley, John Vinson, Keith Gilmore, Mauro Stener, Giovanna Fronzoni, Piero Decleva, Peter Kruger, Marius Retegan, Yves Joly, Christian Vorwerk, Claudia Draxl, John Rehr, and Arata Tanaka, “2p x-ray absorption spectroscopy of 3d transition metal systems,” *Journal of Electron Spectroscopy and Related Phenomena* **249**, 147061 (2021).
- ⁵⁶ Frank de Groot, György Vankó, and Pieter Glatzel, “The 1s x-ray absorption pre-edge structures in transition metal oxides,” *Journal of Physics: Condensed Matter* **21**, 104207 (2009).
- ⁵⁷ J Vinson, JJ Rehr, JJ Kas, and EL Shirley, “Bethe-salpeter equation calculations of core excitation spectra,” *Physical Review B* **83**, 115106 (2011).
- ⁵⁸ John Vinson, “Advances in the OCEAN-3 spectroscopy package,” *Phys. Chem. Chem. Phys.* **24**, 12787 (2022).
- ⁵⁹ Andris Gulans, Stefan Kontur, Christian Meisenbichler, Dmitrii Nabok, Pasquale Pavone, Santiago Rigamonti, Stephan Sagmeister, Ute Werner, and Claudia Draxl, “Exciting: a full-potential all-electron package implementing density-functional theory and many-body perturbation theory,” *Journal of Physics: Condensed Matter* **26**, 363202 (2014).
- ⁶⁰ Christian Vorwerk, Caterina Cocchi, and Claudia Draxl, “Addressing electron-hole correlation in core excitations of solids: An all-electron many-body approach from first principles,” *Physical Review B* **95**, 155121 (2017).
- ⁶¹ Christian Vorwerk, Benjamin Aurich, Caterina Cocchi, and Claudia Draxl, “Bethe-salpeter equation for absorption and scattering spectroscopy: Implementation in the exciting code,” *Electronic Structure* **1**, 037001 (2019).
- ⁶² Mathieu Taillefumier, Delphine Cabaret, Anne-Marie Flank, and Francesco Mauri, “X-ray absorption near-edge structure calculations with the pseudopotentials: Application to the k edge in diamond and α -quartz,” *Physical Review B* **66**, 195107 (2002).
- ⁶³ Christos Gougoussis, Matteo Calandra, Ari P Seitsonen, and Francesco Mauri, “First-principles calculations of x-ray absorption in a scheme based on ultrasoft pseudopotentials: From α -quartz to high- t_c compounds,” *Physical Review B* **80**, 075102 (2009).
- ⁶⁴ Peter Puschnig and Claudia Ambrosch-Draxl, “Optical absorption spectra of semiconductors and insulators including electron-hole correlations: An ab initio study within the lapw method,” *Physical Review B* **66**, 165105 (2002).
- ⁶⁵ J. Vinson and J. J. Rehr, “Ab initio bethe-salpeter calculations of the x-ray absorption spectra of transition metals at the l -shell edges,” *Phys. Rev. B* **86**, 195135 (2012).
- ⁶⁶ Dorothea Golze, Levi Keller, and Patrick Rinke, “Accurate absolute and relative core-level binding energies from gw,” *The Journal of Physical Chemistry Letters* **11**, 1840–1847 (2020).
- ⁶⁷ Yi Yao, Dorothea Golze, Patrick Rinke, Volker Blum, and Yosuke Kanai, “All-electron bse@gw method for k-edge core electron excitation energies,” *Journal of Chemical Theory and Computation* **18**, 1569–1583 (2022), pMID: 35138865.
- ⁶⁸ Michael Rohlfing and Steven G Louie, “Electron-hole excitations and optical spectra from first principles,” *Physical Review B* **62**, 4927 (2000).
- ⁶⁹ Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, *et al.*, “Quantum espresso: a modular and open-source software project for quantum simulations of materials,” *Journal of Physics: Condensed Matter* **21**, 395502 (2009).
- ⁷⁰ Paolo Giannozzi, Oliviero Andreussi, Thomas Brumme, Oana Bunau, M Buongiorno Nardelli, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Matteo Cococcioni, *et al.*, “Advanced capabilities for materials modelling with quantum espresso,” *Journal of Physics: Condensed Matter* **29**, 465901 (2017).
- ⁷¹ Shang-Di Mo and W. Y. Ching, “Ab initio calculation of the core-hole effect in the electron energy-loss near-edge structure,” *Phys. Rev. B* **62**, 7901–7907 (2000).
- ⁷² Koblar A. Jackson and Mark R. Pederson, “New theoretical model for the diamond 1s core exciton,” *Phys. Rev. Lett.* **67**, 2521–2524 (1991).
- ⁷³ Cornelius Lanczos, “An iteration method for the solution of the eigenvalue problem of linear differential and integral operators,” *Journal of Research of the National Bureau of Standards* **45**, 255 (1950).
- ⁷⁴ Cornelius Lanczos, “Solution of systems of linear equations by,” *Journal of Research of the National Bureau of Standards* **49**, 33 (1952).
- ⁷⁵ John Vinson and Eric L. Shirley, “Fast, efficient, and accurate dielectric screening using a local real-space approach,” *Phys. Rev. B* **103**, 245143 (2021).
- ⁷⁶ Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, *et al.*, “Commentary: The materials project: A materials genome approach to accelerating materials innovation,” *APL materials* **1**, 011002 (2013).
- ⁷⁷ Ioannis Petousis, David Mrdjenovich, Eric Ballouz, Miao Liu, Donald Winston, Wei Chen, Tanja Graf, Thomas D Schladt, Kristin A Persson, and Fritz B Prinz, “High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials,” *Scientific Data* **4**, 1–12 (2017).

- ⁷⁸ J Bergengren, “Über die röntgenabsorption des phosphors,” *Zeitschrift für Physik* **3**, 247–249 (1920).
- ⁷⁹ Axel E Lindh, “Zur kenntnis des röntgenabsorptionsspektrums von chlor,” *Zeitschrift für Physik* **6**, 303–310 (1921).
- ⁸⁰ O Stelling, *Z. Phys. Chem., B* **7**, 210 (1930).
- ⁸¹ V Kunzl, “A linear dependence of energy levels on the valency of elements,” *Collection of Czechoslovak Chemical Communications* **4**, 213–224 (1932).
- ⁸² F. Farges, “Chromium speciation in oxide-type compounds: application to minerals, gems, aqueous solutions and silicate glasses,” *Phys Chem Minerals* **36**, 463–481 (2009).
- ⁸³ Moniek Tromp, Jerome Moulin, Gillian Reid, and John Evans, “Cr k-edge xanes spectroscopy: Ligand and oxidation state dependence — what is oxidation state?” *AIP Conference Proceedings* **882**, 699–701 (2007).
- ⁸⁴ Jacques Paul Suchet, *Chemical physics of semiconductors* (Van Nostrand, 1965).
- ⁸⁵ Michiel J. van Setten, Ramon Costa, Francesc Viñes, and Francesc Illas, “Assessing gw approaches for predicting core level binding energies,” *Journal of Chemical Theory and Computation* **14**, 877–883 (2018), pMID: 29320628.
- ⁸⁶ Dorothea Golze, Jan Wilhelm, Michiel J. van Setten, and Patrick Rinke, “Core-level binding energies from gw: An efficient full-frequency approach within a localized basis,” *Journal of Chemical Theory and Computation* **14**, 4856–4869 (2018).
- ⁸⁷ Prior to version 3.1, OCEAN incorrectly did not remove the V_{xc} contribution to the core-hole energy. Removing V_{xc} gives an almost system-independent shift of 42.58 eV \pm 0.03 eV, suggesting that the errors in relative alignment from this bug are minor.
- ⁸⁸ Matthew R Carbone, Fanchen Meng, Christian Vorwerk, Benedikt Maurer, Fabian Peschel, Xiaohui Qu, Eli Stavitski, Claudia Draxl, John Vinson, and Deyu Lu, “Lightshow: a python package for generating computational x-ray absorption spectroscopy input files,” *Journal of Open Source Software* **8**, 5182 (2023).
- ⁸⁹ Michiel J van Setten, Matteo Giantomassi, Eric Bousquet, Matthieu J Verstraete, Don R Hamann, Xavier Gonze, and G-M Rignanese, “The pseudodojo: Training and grading a 85 element optimized norm-conserving pseudopotential table,” *Computer Physics Communications* **226**, 39–54 (2018).
- ⁹⁰ Gianluca Prandini, Antimo Marrazzo, Ivano E Castelli, Nicolas Mounet, and Nicola Marzari, “Precision and efficiency in solid-state pseudopotential calculations,” *npj Computational Materials* **4**, 1–13 (2018).
- ⁹¹ John P Perdew, Kieron Burke, and Matthias Ernzerhof, “Generalized gradient approximation made simple,” *Physical Review Letters* **77**, 3865 (1996).
- ⁹² C. Hébert, “Practical aspects of running the wien2k code for electron spectroscopy,” *Micron* **38**, 12–28 (2007).
- ⁹³ Charles Spearman, “The proof and measurement of association between two things.” *Am. J. Psychol.* **15**, 72–101 (1904).
- ⁹⁴ See Supplemental Material at [URL will be inserted by publisher].
- ⁹⁵ J J Rehr, J A Soininen, and E L Shirley, “Final-state rule vs the bethe-salpeter equation for deep-core x-ray absorption spectra,” *Physica Scripta* **2005**, 207 (2005).
- ⁹⁶ K. Ohtaka and Y. Tanabe, “Theory of the soft-x-ray edge problem in simple metals: historical survey and recent developments,” *Rev. Mod. Phys.* **62**, 929–991 (1990).
- ⁹⁷ Subhayan Roychoudhury and David Prendergast, “Efficient core-excited state orbital perspective on calculating x-ray absorption transitions in determinant framework,” *Phys. Rev. B* **107**, 035146 (2023).
- ⁹⁸ Yufeng Liang and David Prendergast, “Quantum many-body effects in x-ray spectra efficiently computed using a basic graph algorithm,” *Phys. Rev. B* **97**, 205127 (2018).
- ⁹⁹ Giovanni Onida, Lucia Reining, and Angel Rubio, “Electronic excitations: density-functional versus many-body green’s-function approaches,” *Rev. Mod. Phys.* **74**, 601–659 (2002).
- ¹⁰⁰ G. Strinati, “Effects of dynamical screening on resonances at inner-shell thresholds in semiconductors,” *Phys. Rev. B* **29**, 5718–5726 (1984).
- ¹⁰¹ A. L. Ankudinov, A. I. Nesvizhskii, and J. J. Rehr, “Dynamic screening effects in x-ray absorption spectra,” *Phys. Rev. B* **67**, 115120 (2003).
- ¹⁰² Mark S. Hybertsen and Steven G. Louie, “Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies,” *Phys. Rev. B* **34**, 5390–5413 (1986).
- ¹⁰³ Andrzej Fleszar and Raffaele Resta, “Dielectric matrices in semiconductors: A direct approach,” *Phys. Rev. B* **31**, 5305–5310 (1985).
- ¹⁰⁴ K. Kunc and E. Tosatti, “direct” evaluation of the inverse dielectric matrix in semiconductors,” *Phys. Rev. B* **29**, 7045–7047 (1984).
- ¹⁰⁵ Xiaosong Li, Niranjan Govind, Christine Isborn, A. Eugene III DePrince, and Kenneth Lopata, “Real-time time-dependent electronic structure theory,” *Chemical Reviews* **120**, 9951–9993 (2020), pMID: 32813506.
- ¹⁰⁶ Eric L. Shirley, “Bethe–salpeter treatment of x-ray absorption including core-hole multiplet effects,” *Journal of Electron Spectroscopy and Related Phenomena* **144-147**, 1187–1190 (2005), Proceeding of the Fourteenth International Conference on Vacuum Ultraviolet Radiation Physics.
- ¹⁰⁷ W Olovsson, I Tanaka, T Mizoguchi, P Puschnig, and C Ambrosch-Draxl, “All-electron bethe-salpeter calculations for shallow-core x-ray absorption near-edge structures,” *Physical Review B* **79**, 041102 (2009).
- ¹⁰⁸ Vincent Mauchamp, Michel Jaouen, and Peter Schattschneider, “Core-hole effect in the one-particle approximation revisited from density functional theory,” *Phys. Rev. B* **79**, 235106 (2009).
- ¹⁰⁹ Daniela Carta, Gavin Mountjoy, Anna Regoutz, Ali Khiat, Alexantrou Serb, and Themistoklis Prodromakis, “X-ray absorption spectroscopy study of tio2-x thin films for memory applications,” *The Journal of Physical Chemistry C* **119**, 4362–4370 (2015).
- ¹¹⁰ Atsushi Hariki, Keisuke Higashi, Tatsuya Yamaguchi, Jiebin Li, Curran Kalha, Manfred Mascheck, Susanna K. Eriksson, Tomas Wiell, Frank M. F. de Groot, and Anna Regoutz, “Satellites in the Ti 1s core level spectra of SrTiO₃ and TiO₂,”

- Phys. Rev. B **106**, 205138 (2022).
- ¹¹¹ C Brouder, “Angular dependence of x-ray absorption spectra,” *J. Phys.: Condens. Matter* **2**, 701 (1990).
- ¹¹² J. J. Kas, F. D. Vila, J. J. Rehr, and S. A. Chambers, “Real-time cumulant approach for charge-transfer satellites in x-ray photoemission spectra,” *Phys. Rev. B* **91**, 121112 (2015).
- ¹¹³ T Uozumi, K Okada, A Kotani, O Durmeyer, JP Kappler, E Beaurepaire, and JC Parlebas, “Experimental and theoretical investigation of the pre-peaks at the ti k-edge absorption spectra in tio2,” *Europhysics Letters* **18**, 85 (1992).
- ¹¹⁴ RV Vedrinskii, VL Kraizman, AA Novakovich, Ph V Demekhin, and SV Urazhdin, “Pre-edge fine structure of the 3d atom k x-ray absorption spectra and quantitative atomic structure determinations for ferroelectric perovskite structure crystals,” *Journal of Physics: condensed matter* **10**, 9561 (1998).
- ¹¹⁵ D Cabaret, Y Joly, H Renevier, *et al.*, “Pre-edge structure analysis of ti k-edge polarized x-ray absorption spectra in tio2 by full-potential xanes calculations,” *Journal of Synchrotron Radiation* **6**, 258–260 (1999).
- ¹¹⁶ Yves Joly, Delphine Cabaret, Hubert Renevier, and Calogero R. Natoli, “Electron population analysis by full-potential x-ray absorption simulations,” *Phys. Rev. Lett.* **82**, 2398–2401 (1999).
- ¹¹⁷ J Danger, P Le Fèvre, H Magnan, D Chandesaris, S Bourgeois, J Jupille, T Eickhoff, and W Drube, “Quadrupolar transitions evidenced by resonant auger spectroscopy,” *Physical review letters* **88**, 243001 (2002).
- ¹¹⁸ Tomoyuki Yamamoto, Teruyasu Mizoguchi, and Isao Tanaka, “Core-hole effect on dipolar and quadrupolar transitions of SrTiO₃ and BaTiO₃ at Ti K edge,” *Physical Review B* **71**, 245113 (2005).
- ¹¹⁹ JC Woicik, Eric L Shirley, CS Hellberg, KE Andersen, S Sambasivan, DA Fischer, BD Chapman, EA Stern, P Ryan, DL Ederer, *et al.*, “Ferroelectric distortion in sr tio 3 thin films on si (001) by x-ray absorption fine structure spectroscopy: Experiment and first-principles calculations,” *Physical Review B* **75**, 140103 (2007).
- ¹²⁰ ZY Wu, DC Xian, TD Hu, YN Xie, Y Tao, CR Natoli, E Paris, and A Marcelli, “Quadrupolar transitions and medium-range-order effects in metal k-edge x-ray absorption spectra of 3 d transition-metal compounds,” *Physical Review B* **70**, 033104 (2004).
- ¹²¹ FMF De Groot, JC Fuggle, BT Thole, and GA Sawatzky, “ $L_{2,3}$ x-ray-absorption edges of d^0 compounds: K^+ , Ca^{2+} , Sc^{3+} , and Ti^{4+} in O_h (octahedral) symmetry,” *Physical Review B* **41**, 928 (1990).
- ¹²² Delphine Cabaret, Amélie Bordage, Amélie Juhin, Mounir Arfaoui, and Emilie Gaudry, “First-principles calculations of x-ray absorption spectra at the k-edge of 3d transition metals: an electronic structure analysis of the pre-edge,” *Phys. Chem. Chem. Phys.* **12**, 5619–5633 (2010).
- ¹²³ T. C. Rossi, D. Grolimund, M. Nachttegaal, O. Cannelli, G. F. Mancini, C. Bacellar, D. Kinschel, J. R. Rouxel, N. Ohannessian, D. Pergolesi, T. Lippert, and M. Chergui, “X-ray absorption linear dichroism at the ti k edge of anatase tio₂ single crystals,” *Phys. Rev. B* **100**, 245207 (2019).
- ¹²⁴ T. C. Rossi, D. Grolimund, O. Cannelli, G. F. Mancini, C. Bacellar, D. Kinschel, J. R. Rouxel, N. Ohannessian, D. Pergolesi, and M. Chergui, “X-ray absorption linear dichroism at the Ti K -edge of rutile (001) TiO₂ single crystal,” *Journal of Synchrotron Radiation* **27**, 425–435 (2020).
- ¹²⁵ O Durmeyer, E Beaurepaire, J-P Kappler, Ch Brouder, and F Baudelet, “Temperature dependence of the pre-edge structure in the ti k-edge x-ray absorption spectrum of rutile,” *Journal of Physics: Condensed Matter* **22**, 125504 (2010).
- ¹²⁶ Silvia Tinte and Eric L Shirley, “Vibrational effects on sr tio3 ti 1s absorption spectra studied using first-principles methods,” *Journal of Physics: Condensed Matter* **20**, 365221 (2008).
- ¹²⁷ N. Jiang, D. Su, and J. C. H. Spence, “Determination of ti coordination from pre-edge peaks in ti k -edge xanes,” *Phys. Rev. B* **76**, 214117 (2007).
- ¹²⁸ Data can be downloaded at the following link: <https://doi.org/10.17172/NOMAD/2023.03.24-1>.
- ¹²⁹ Claudia Draxl and Matthias Scheffler, “The nomad laboratory: from data sharing to artificial intelligence,” *Journal of Physics: Materials* **2**, 036001 (2019).

Appendix A: Band structure RMSD comparison

The comparisons for the ground-state band structure energies, given in Table IV, are calculated using the regular k -point grids from the Materials Project. The RMSD is calculated in the standard way, but with pseudo-occupation numbers f to limit the range of states, e.g., valence band, or limited energy ranges of the conduction band. The RMSD between the energies ϵ from two codes i and j is given by,

$$\text{RMSD}[i, j] = \left[\frac{\sum_{\mathbf{k}} w_{\mathbf{k}} \sum_n f_{n\mathbf{k}}^{ij} (\epsilon_{n\mathbf{k}}^i - \epsilon_{n\mathbf{k}}^j)^2}{\sum_{\mathbf{k}} w_{\mathbf{k}} \sum_n f_{n\mathbf{k}}^{ij}} \right]^{1/2} \quad (\text{A1})$$

where the sum is over symmetry reduced k -points \mathbf{k} with weights $w_{\mathbf{k}}$ and bands n . The occupation numbers are given by

$$f_{n\mathbf{k}}^{ij} = \sqrt{f_{n\mathbf{k}}^i f_{n\mathbf{k}}^j} \quad (\text{A2})$$

$$f_{n\mathbf{k}}^i = \left[e^{(\epsilon_{n\mathbf{k}}^i - \epsilon_L)/\gamma} + 1 \right]^{-1} \left[e^{(\epsilon_H - \epsilon_{n\mathbf{k}}^i)/\gamma} + 1 \right]^{-1} \quad (\text{A3})$$

where the lower and upper bounds are given by ϵ_L and ϵ_H , respectively, and a broadening γ of 0.02 Rydberg is used for both.

Appendix B: Effect of number of bands

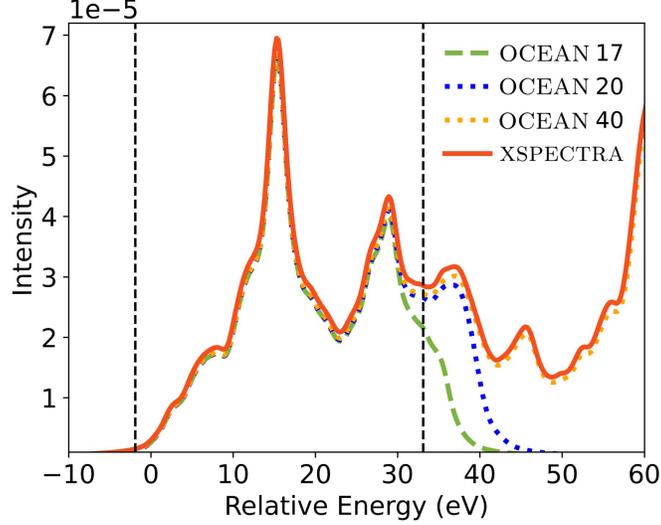


FIG. B1. The effect of number of empty bands on the XANES spectra of mp-2664 calculated at the independent particle level. OCEAN 17 (20 or 40) corresponds to OCEAN calculations with 17 (20 or 40) empty bands. The first dashed vertical line corresponds to our definition of the onset, and the second is 35 eV above onset.

We carefully examined the energy range (35 eV from onset) used to compare the spectra. The estimation for the number of empty bands described in the main text based on the homogeneous electron gas model is not accurate enough for mp-2664, which is likely due to the small unit cell volume of mp-2664. In this case, comparing the effect of different number of bands, we found that by using 20 empty bands, the spectra converged within the energy range where the comparison of spectra are performed. Therefore, in this particular case, we use 20 empty bands instead of 17 from our homogeneous electron gas model.

Appendix C: Isotropic quadrupole transitions

In Section IV F we compare calculations to measurements on powdered samples, requiring the calculation of the isotropic spectra. For all three of the codes shown here, spectra are calculated for an explicit, cartesian polarization direction and, for quadrupole transitions, momentum vector. Both OCEAN and XSPECTRA can be set to calculate only the quadrupole transition without the dipole operator. In the case of non-magnetic systems and linear polarization the dipole and quadrupole contributions do not interfere and can be summed separately. The dipole case is straightforward. Regardless of the symmetry of the system, an average of spectra calculated with three orthogonal polarization directions is equal to the isotropic average. The number of required calculations can be reduced to 1 or 2, depending on the symmetry of the crystal, but for simplicity we average over $\mathbf{e} = \{\hat{x}, \hat{y}, \hat{z}\}$.

The case of the quadrupole term $(\mathbf{e} \cdot \mathbf{r})(\mathbf{q} \cdot \mathbf{r})$ is significantly more complicated. Brouder detailed the angular dependence by writing the x-ray transitions in terms of spherical tensors.¹¹¹ Following Brouder, we note that rutile and anatase both have a crystallographic point group 4/mmm. Their quadrupole spectra therefore is given by Eq. 5.7 in Ref. 111

$$\begin{aligned}
 \sigma(\mathbf{e}, \mathbf{k}) = & \sigma^Q(0, 0) + \sqrt{5/14} (3 \sin^2 \theta \sin^2 \psi - 1) \sigma^Q(2, 0) \\
 & + 1/\sqrt{14} (35 \sin^2 \theta \cos^2 \theta \cos^2 \psi + 5 \sin^2 \theta \sin^2 \psi - 4) \sigma^Q(4, 0) \\
 & + \sqrt{5} \sin^2 \theta [(\cos^2 \theta \cos^2 \psi - \sin^2 \psi) \cos 4\phi - 2 \cos \theta \sin \psi \cos \psi \sin 4\phi] \sigma^{Qr}(4, 4)
 \end{aligned} \tag{C1}$$

where $\sigma^Q(a, b)$ are the elements of the tensor decomposition that are not zero by symmetry and $\sigma^Q(0, 0)$ is the desired isotropic spectrum. The directions of the polarization and momentum are, in terms of the three angles,

$$\begin{aligned}\mathbf{e} &= (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta) \\ \mathbf{q} &= (\cos \theta \cos \phi \cos \psi - \sin \phi \sin \psi, \\ &\quad \cos \theta \sin \phi \cos \psi + \cos \phi \sin \psi, \\ &\quad -\sin \theta \cos \psi)\end{aligned}$$

By a careful choice of polarization and momentum direction, three spectra can be combined to give only the isotropic response. We choose:

$$\begin{aligned}(\theta_1, \phi_1, \psi_1) &= (\arccos[1/\sqrt{7}], -1/4 \operatorname{atan}[19/7\sqrt{11}], 1/2 \arccos[2/9]) \\ (\theta_2, \phi_2, \psi_2) &= (-1/4 \arccos[1/49], \pi/8, 0) \\ (\theta_3, \phi_3, \psi_3) &= (0, 0, 0)\end{aligned}$$

giving

$$\sigma^Q(0, 0) = \sigma^Q(\mathbf{e}_1, \mathbf{q}_1) - 1/15\sigma^Q(\mathbf{e}_2, \mathbf{q}_2) + 1/15\sigma^Q(\mathbf{e}_3, \mathbf{q}_3) \quad (\text{C2})$$

which is added to the isotropic dipole spectra to give the total isotropic spectra including both dipole and quardupole terms.