

Supplemental material

I. SPECTRA DECONVOLUTION

The fitting is done using the LMFIT (Non-Linear Least-Square Minimization and Curve-Fitting) Python package [1]. Voigt profiles for the resonances and a step function for the edge step were used. The Voigt profile is given by a convolution of a Lorentzian distribution to account for the core-hole lifetime broadening and a Gaussian distribution to account for instrumental broadening and is defined as

$$f(x; A, \mu, \sigma, \gamma) = A \frac{\text{Re}[w(z)]}{\sigma \sqrt{2\pi}} \quad (1)$$

with $\text{Re}[w(z)]$ as the real part of the Faddeeva function and $z = \frac{x+i\gamma}{\sigma}$, where γ is half the FWHM of the Lorentzian and σ is defined as $\sigma = \frac{f_G}{2 \ln 4}$ with f_G as the FWHM of the Gaussian. Once the parameters and their boundaries are defined, a residual between the model and the data is defined, with which a fit is performed using Levenberg-Marquardt minimization. In the case of many parameters and complex model functions, one might run the risk of local minima. However, in our case, we use as many parameters as necessary but as few as possible, and the parameters have physically sensible defined boundaries to avoid local minima. The parameter and their boundaries are explained in more detail below on the Ti K-edge XAS spectra example.

Peak Positions

The initial guess for the peak energy positions of the OCEAN calculation fit is done visually since the peaks are clearly pronounced in the spectra.

Once the OCEAN fit was done, the peak positions were scaled using the energy scaling as explained in the paper, to match the experimental energy scale. Since a slight discrepancy in relative peak positions is expected between OCEAN and experimental spectra, the boundaries for the experimental data fit were set slightly higher around ± 1.0 eV in most cases (except at the XES K-edge, where the boundaries are set to ± 2.0 eV).

Peak widths

The only ad hoc broadening included in the OCEAN calculation is the user-provided constant approximated core-hole lifetime which is used as a constant across the whole energy scale. For the Ti-K edge this was set to 0.94 eV and for the Ti L-edges it was set to 0.24 eV. In a simplified single atom (or molecule) model, one would expect the spectral features to be well fitted with a Lorentzian with the width of the lifetime broadening, and later in the experimental data an additional Gaussian with the instrumental broadening. However, it is important to note that although we are conducting calculations on a unit cell, the results of the calculations are done in terms of band structure in a condensed phase, i.e., treating the system as a crystalline solid. This means that the shape of the spectral features is more complex and reflects the density of states, which can have a somewhat arbitrary shape. Furthermore, the measurements of the systems were done without the preference for any crystal direction. Thus, the OCEAN calculation results were averaged over orthogonal photon polarizations (momentum transfer) directions. To account for this complexity, OCEAN spectra were fitted using Voigt profiles, where the minimum boundaries for the Lorentzian were set to the user provided approximated core-hole lifetime.

The width of the peaks in the experimental spectra is influenced by several factors. The core-hole lifetime, which is energy-dependent and broadens for higher energies. The higher the energy of the final state, the shorter its lifetime and hence the broader the peak. This is especially the case for resonances in the continuum (i.e. σ^* resonance). To describe the energy-dependency, an empirical energy-dependent Lorentzian function is assumed for the γ parameters [2]:

$$\gamma(E) = f_L + \alpha(E - E_0) \quad (2)$$

with f_L the core-hole lifetime, E_0 the Fermi level, and α as a varying parameter based on literature value ranges (from 0.07 to 0.15 for Ti K-edge). The instrumental broadening is dominated by the beamline resolution. The resolution at the FCM beamline at Bessy II is estimated to be around 0.5 eV for the Ti K edge and for the PGM beamline around 0.23 eV for Ti L edges. Vibrational disorder: the OCEAN code treats ions as fixed during the excitation. In reality, they are slightly displaced from the perfect crystal positions. In consequence, the spectral shape is broadened and slightly asymmetric. We expect this effect to be rather small for the Ti-K edge.

Amplitude

The spectra were normalized to the peak with the lowest probability to be dampened by self-absorption. In the case of Ti K edge, this was the pre-edge. The pre-edge structure is complex, and it is possible that not all pre-edge aspects are covered by a theoretical calculation, or intensities are over- or underestimated. However, looking at Fig. 3 in the manuscript, we see that the pre-edge peak shapes are quite well reproduced by OCEAN. Still keeping uncertainties in mind, the amplitudes of the OCEAN calculation were taken as upper boundaries for the experimental fits. For the experimental data fits, we were conscious of the self-absorption effect. This means, that the peaks should not be unreasonably different between the OCEAN calculations and the experimental spectra, but self-absorption might affect the peak intensities disproportionately, meaning that the higher the intensity of a peak in the calculation, the more suppressed it might be in the experimental data in relation to less intense peaks. This effect was not quantitatively accounted for, however, qualitatively checked for every fit. Indeed, developing a method of using OCEAN to calculate the self-absorption correction for a measured spectrum is a meaningful and interesting future work possibility.

Step edge

The step edge refers to the abrupt increase in absorption that occurs when a core electron is excited to a continuum or quasi-continuum of final states. Due to the overlap of other spectral features in this region, it is difficult to discern. The model function of the step edge depends on the lifetime broadening of the core hole and the instrumental broadening. The influence of the lifetime broadening can be described as an arctan() function, whereas the instrumental broadening can be described as an erf():

$$s(E) = \frac{h_s}{2} + h_s \frac{1}{2} (1 - A_{arc}) \operatorname{erf} \left(\frac{E - E_0}{1.2\sigma} \right) + \frac{A_{arc}}{\pi} \arctan \left(\frac{E - E_0}{\gamma} \right) \quad (3)$$

with h_s as the height of the step edge, A_{arc} the portion of Arctan, E_0 the step edge position, σ the instrumental broadening, and γ the according core-hole lifetime broadening. For the fit of the OCEAN calculations γ was allowed to vary from 0.47 eV (core-hole lifetime broadening for the Ti K spectra, and 0.12 eV for the Ti L spectra) and 1 eV. For the experimental data fit, was allowed to vary around 0.43 eV-1 eV and σ was allowed to vary between the minimal instrumental broadening estimation (0.5 eV) and 2 eV to account for the additional self-absorption effect, which might broaden the step edge. There are several methods to determine the edge position including the inflection point or the first or most pronounced peak of the first derivative of the spectrum. Another method is defining the edge step as the energy position where a spectrum, that was normalized such that a region above the edge has an intensity equal to 1, first reaches an intensity of 0.5. Due to the pronounced pre-edge structure, the latter method was not chosen as applicable in the case of the titanium systems. Thus, the former method of the inflection point was chosen as guidance. The step edge position parameter for the OCEAN calculation fits was thus chosen to vary between the two first most pronounced peaks after the inflection point of the first pre-edge peak. For the experimental step edge position parameter, the OCEAN fit result was used as a starting point and allowed to vary in the range of ± 1 eV.

Goodness of fit

Based on the additional reported data from the Imfit minimizer, the goodness of the fits are analyzed using the chi-square value, weighted chi-square value, and the reduced chi-square value. For the

experimental data fits, the residual to be minimized is calculated as a weighted residual:

$$Res = \frac{(model - y)^2}{weights^2} \quad (4)$$

with $weights = \frac{1}{\sigma_y}$ where $u_y = \sqrt{\frac{y}{N}}$ are the statistical errors of the measurement (N – events detected).

Uncertainties

The result of the minimizer is the final parameters with the confidence interval being calculated as an estimation of the standard errors from the estimated covariance matrix. In the case of simple models with parameters that are not strongly correlated, these estimates are sufficiently good. For the OCEAN calculations, the resulting uncertainties are good and can be used to provide an estimate of the uncertainties (confidence interval!). For the experimental data, the algorithm fails to estimate sensible uncertainties. This has two reasons: 1. the experimental data has fewer data points than the OCEAN calculation (ocean calc. 1500 points, exp. 120 points). The second reason is the less distinct features, which appear broad and not as evident as the peaks in theoretical calculations. And the result of the fit also provides a correlation matrix, which shows that many of the parameters are highly correlated. The uncertainties calculated by Imfit are a measure of how the fit changes when parameters are changed. Since in our case, the model function is ‘overfitting’ the data, the uncertainties calculated in this manner will not be correct, and highly overestimated.

Thus, further analysis of the result using bootstrap sampling was done. Given the original dataset of size N, a bootstrap sample is generated by randomly sampling N data points with replacement from the original data. For each sample, the model function with the resulting parameters as starting estimates, is applied and fitted again. These parameters are saved. Then, this procedure is repeated 10,000 times, so that a set of 10,000 samples and estimated parameters for each sample is generated. The variability among the estimated parameters from the bootstrap samples is used to estimate the uncertainty. This is done by calculating the standard deviation, confidence intervals, and percentiles of the parameter estimates. By generating multiple bootstrap samples and estimating the parameter for each sample, the bootstrap method provides an empirical distribution of the parameter, allowing us to estimate its uncertainty. In a way, we are not assuming how well the model fits the data, but how well the data is representative of the model.

II. COMPUTATIONAL DETAILS

The OCEAN code version 2.9.6.3 was used for theoretical calculations throughout this work [3, 4]. For the calculation of the ground-state using DFT the QUANTUM ESPRESSO code version [5, 6] was used.

Atomic Parameter Source

The atomic parameters are taken from the Crystallography Open Database [7–10] as shown in Tab. I.

System	Source
Ti	COD: 9016190
TiO	COD: 1536851
TiO ₂	COD: 9009083

Table I. Sources and IDs of the atomic parameters for all three materials.

The pseudopotentials are taken from the PseudoDojo collection [11] and generated using the ON-CVPSP code [12] (ONCVPSP v0.4, XC: PW). In the following the full input files for the OCEAN code are shown.

OCEAN Input

Ti

```
#####
#Job Info
#####
dft qe
para-prefix{ mpirun -n 48}

calc xas
opf.program hamann

#####
#Convergence Parameters
#####
# Plane-wave basis cut-offs
ecut 200

# Number of k-points for calculation of final state
nkpt { 12 12 10 }

# Number of k-points for the ground state calculation in the DFT
ngkpt{ 12 12 10 }

# n bands (over 100 eV should set CNBSE.niter to 200-1000)
nbands 400
CNBSE.niter 200

# CNBSE.xmesh sets a grid to be converted into the NIST BSE from wave-functions
CNBSE.xmesh { 12 12 12 }

# Number of bands for screening calculation
screen.nbands 700

# Only to be changed for some systems , e.g. very small or very large unit cells
screen.nkpt{ 2 2 2 }

metal .true .

#####
# Edges to be calculated
#####
#edge information
nedges 1
edges{ -22 1 0 }
core-offset true

# #####
# ABINIT style crystallographic info
# #####
acell { 5.574691995 5.574691995 5.574691995 }

rprim {
  0.866025403784439   -0.5000000000000000   0.0000000000000000
  0.0000000000000000   1.0000000000000000   0.0000000000000000
  0.0000000000000000   0.0000000000000000   1.586440677966102
}

```

```

znucl { 22 }

typat { 1 1 }

xred {
0.3333333333333333 0.6666666666666667 0.2500000000000000
      0.6666666666666667 0.3333333333333333 0.7500000000000000
}

# Static dielectric const
diemac 100

#####
#Pseudo potentials
#####

pp_list {
Ti-sp . oncvpsp
}

#####
#Atomic and Screening Info
#####
screen.final.dr 0.02

screen.grid.rmax 10

screen.grid.rmode { legendre uniform }

screen.grid.ang { 5 11 11 9 7 }

screen.grid.deltar { 0.10 0.15 0.25 0.25 0.25 }

screen.grid.shells { -1 4 6 8 10 }

screen.lmax 2

cnbse.rad 5.5

screen.shells { 3.5 4.0 4.5 5.0 5.5 6.0 }

# spectral broadening in eV
cnbse.broaden { 0.94 }

                                TIO

#####
#Job Info
#####
para_prefi { mpirun -n 32 }
dft qe
opf.program hamann
calc xas

#####
#Convergence Parameters

```

```

#####
# Plane-wave basis cut-offs
ecut 150

# Number of k-points for calculation of final state
nkpt { 8 8 8 }

# Number of k-points for the ground state calculation in the DFT
ngkpt { 8 8 8 }

# n bands (over 100 eV should set CNBSE.niter to 200-1000)
nbands 400
CNBSE.niter 200

# CNBSE.xmesh sets a grid to be converted into the NIST BSE from wave-functions
CNBSE.xmesh { 8 8 8 }

# Number of bands for screening calculation
screen.nbands 397

# Only to be changed for some systems, e.g. very small or very large unit cells
screen.nkpt { 2 2 2 }

# Core offset (set to true or take from a calculation from same pseudopotential < SCREEN
core_offset 213.227510936361
#core_offset true

metal .true .

#####
# Edges to be calculated
#####
#edge information
#nedges 1
edges { -22 1 0 }

#####
# ABINIT style crystallographic info
# #####
acell { 8.1125941473 8.1125941473 8.1125941473 }

rprim {
    0.0000000000000000    0.5000000000000000    0.5000000000000000
    0.5000000000000000    0.0000000000000000    0.5000000000000000
    0.5000000000000000    0.5000000000000000    0.0000000000000000 }

znucl { 22 8 }

typat { 1 2 }

xred {
    0.0000000000000000    0.0000000000000000    0.0000000000000000
    0.5000000000000000    0.5000000000000000    0.5000000000000000
}

# Static dielectric const

```

diemac 83

```
#####
#Pseudo potentials
#####
pp_list { Ti-sp.oncvpsp
          O-high.oncvpsp }

#####
#Atomic and Screening Info
#####
screen.final.dr 0.02

screen.grid.rmax 10

screen.grid.rmode { legendre uniform }

screen.grid.ang { 5 11 11 9 7 }

screen.grid.deltar { 0.10 0.15 0.25 0.25 0.25 }

screen.grid.shells { -1 4 6 8 10 }

screen.lmax 2

cnbse.rad 5.5

screen.shells { 3.5 4.0 4.5 5.0 5.5 6.0 }

# Energy tol
toldfe 1.1d-6

# wftol
tolwfr 1.1d-16

# iterations
nstep 50

#Scaling Factor
scfac 0.80

cnbse.broaden 0.94
```

TiO₂

```
#####
#Job Info
#####
para_prefix { mpirun -n 48 }
dft qe
opf.program hamann
calc xas

#####
#Convergence Parameters
#####
# Plane-wave basis cut-offs
```

```

ecut 150

# Number of k-points for calculation of final state
nkpt { 12 12 8 }

# Number of k-points for the ground state calculation in the DFT
ngkpt { 8 8 8 }

# n bands (over 100 eV should set CNBSE.niter to 200-1000)
nbands 500
CNBSE.niter 200

# CNBSE.xmesh sets a grid to be converted into the NIST BSE from wave-functions
CNBSE.xmesh { 12 12 12 }

# Number of bands for screening calculation
screen.nbands 1235

# Only to be changed for some systems, e.g. very small or very large unit cells
screen.nkpt { 2 2 2 }

# chemical shift on (section 2.2.2 manual)
# core_offset 218.708060117373
# core_offset 210.95301141751
core_offset 213.227510936361

#####
# Edges to be calculated
#####
#edge information
nedges 1
edges { -22 1 0 }

#####
# ABINIT style crystallographic info
# #####
acell { 8.68 8.68 5.6 }

rprim {
  1 0 0
  0 1 0
  0 0 1 }

znucl { 22 8 }

typat { 1 1 2 2 2 2 }

xred {
  0      0      0
  .5     .5     .5
  .3053  .3053  0
  -.3053 -.3053  0
  .8053  .1947  .5
  .1947  .8053  .5 }

# Static dielectric const
diemac 100

#####

```



```

#Pseudo potentials
#####
pp_list { Ti-sp.oncvpsp
          O-high.oncvpsp }

#####
#Atomic and Screening Info
#####
screen.final.dr 0.02

screen.grid.rmax 10

screen.grid.rmode { legendre uniform }
screen.grid.ang { 5 11 11 9 7 }

screen.grid.deltar { 0.10 0.15 0.25 0.25 0.25 }

screen.grid.shells { -1 4 6 8 10 }

screen.lmax 2

cnbse.rad 5.5

screen.shells { 3.5 4.0 4.5 5.0 5.5 6.0 }

# iterations
nstep 50

# Energy tol
toldfe 1.1d-6

# wftol
tolwfr 1.1d-16

#Scaling Factor
scfac 0.80
cnbse.broaden 0.94

```

-
- [1] M. Newville, T. Stensitzki, D. B. Allen, and A. Ingargiola, LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python (2014).
- [2] P. J. W. Weijs, M. T. Czyżyk, J. F. van Acker, W. Speier, J. B. Goedkoop, H. van Leuken, H. J. M. Hendrix, R. A. de Groot, G. van der Laan, K. H. J. Buschow, G. Wiech, and J. C. Fuggle, Core-hole effects in the x-ray-absorption spectra of transition-metal silicides, *Physical Review B* **41**, 11899 (1990).
- [3] J. Vinson, J. J. Rehr, J. J. Kas, and E. L. Shirley, Bethe-Salpeter equation calculations of core excitation spectra., *Phys. Rev. B* **83**, 115106, 115106 (2011).
- [4] K. Gilmore, J. Vinson, E. L. Shirley, D. Prendergast, C. D. Pemmaraju, J. J. Kas, F. D. Vila, and J. J. Rehr, Efficient implementation of core-excitation Bethe-Salpeter equation calculation., *Comp. Phys. Comm.* **197**, 109, 109 (2015).
- [5] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, *Journal of Physics: Condensed Matter* **21**, 395502 (2009).
- [6] P. Giannozzi, O. Baseggio, P. Bonfa, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. de Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru, and S. Baroni, Quantum ESPRESSO toward the exascale, *The Journal of Chemical Physics* **152**, 154105 (2020).

- [7] S. B. Hendricks and L. Pauling, The crystal structures of sodium and potassium trinitrides and potassium cyanate and the nature of the trinitride group, *Journal of the American Chemical Society* **47**, 2904 (1925).
- [8] S. Glavatskikh and A. I. Gorshkov, Natural analog of alpha-titanium in the exhalation products of the Great Tolbachik Fissure Eruption (Kamchatka) (1992).
- [9] T.-H. Yu, S.-J. Lin, P. Chao, C.-S. Fang, and C.-S. Huang, A preliminary study of some new minerals of the platinum-group and another associated new one in platinum-bearing intrusions in a region of China, *Acta Geologica Sinica* **2**, 202 (1974).
- [10] R. W. G. Wyckoff, *Crystal structures*, 2nd edition. Vol. 1, *Acta Crystallographica* **18**, 139 (1965).
- [11] M. J. van Setten, M. Giantomassi, E. Bousquet, M. J. Verstraete, D. R. Hamann, X. Gonze, and G. M. Rignanese, The PseudoDojo: Training and grading a 85 element optimized norm-conserving pseudopotential table, *Computer Physics Communications* **226**, 39 (2018).
- [12] D. R. Hamann, Optimized norm-conserving Vanderbilt pseudopotentials, *Physical Review B* **88**, 085117 (2013).