

Multichannel active-space embedding of atomic multiplets in plane-wave density functional theory/projector augmented-wave for core-level spectroscopies

Alessandro Mirone^{1,*}, Mauro Rovezzi², Christoph J. Sahle¹ and Alessandro Longo^{2,3}

¹*European Synchrotron Radiation Facility, 71, Avenue des Martyrs, F-38000 Grenoble, France*

²*Univ. Grenoble Alpes, CNRS, IRD, Irstea, Météo France, OSUG, FAME, F-38000 Grenoble, France*

³*Istituto per lo Studio dei Materiali Nanostrutturati, Consiglio Nazionale delle Ricerche, Via Ugo La Malfa 153, 90146 Palermo, Italy*



(Received 6 February 2025; accepted 6 March 2026; published 16 April 2026)

We introduce an active-space embedding framework for core-level spectroscopies that connects localized atomic multiplets to continuum resonances within a plane-wave density functional theory/projector augmented-wave description. The approach is complementary to widely used core-level Bethe-Salpeter implementations based on a two-particle (core-exciton) picture with typically static screening: Here, a correlated multiplet manifold of the absorber (including the core hole and open-shell configurations) is coherently coupled to a plane-wave photoelectron, enabling a unified treatment of localized multiplet structure and continuum lineshapes. Spectra are computed in a general time-domain formulation equivalent to Fermi's golden rule: A transition operator tailored to the specific spectroscopy technique is applied to the correlated ground state to generate an excited wave packet, and the corresponding wave packet autocorrelation function is evaluated without explicit real-time propagation, using Lanczos tridiagonalization or the kernel polynomial method; the spectral intensity follows from its Fourier representation. We validate the method at the Ce $N_{4,5}$ edges, reproducing in quantitative agreement with experiment both high- Q multiplet features and the low- Q giant dipole resonance continuum, including a characteristic low-energy shoulder relevant for robust Ce valence assignments. The implementation is available open source within the QUANTUM ESPRESSO xspectra package (xspectruple mode), together with reproducible inputs and scripts.

DOI: [10.1103/gnd5-sxl2](https://doi.org/10.1103/gnd5-sxl2)

I. INTRODUCTION

Since its original formulation, density functional theory (DFT) [1] has evolved into a versatile tool that accurately describes ground-state properties of complex, many-electron systems. By employing advanced exchange-correlation functionals—such as hybrid and meta-generalized gradient approximation (meta-GGA) forms—DFT now often yields results that closely track experimental findings, even for systems with pronounced electron-electron correlations [2,3]. In parallel, many-body approaches, including the GW approximation [4] and dynamical mean-field theory [5], capture correlation effects and electron dynamics with remarkable fidelity.

However, when the goal shifts from ground states to spectroscopy, the situation changes dramatically. X-ray spectroscopies probe states far from the ground state; core excitations produce a core hole and a photoelectron whose entanglement with open-shell electrons yields intricate multiplet structures and resonances with strong continuum character.

For *localized* excitations, multiplet approaches remain remarkably predictive. Early implementations based on small active spaces and symmetry exploitation [6] evolved into charge-transfer multiplet theory [7], arbitrary symmetry [8], and *ab initio* parameter extraction via Wannier Hamiltonians [9–11]. Yet, the Hilbert space of the active spin-orbitals grows exponentially, limiting the explicit treatment of continuum states.

Alternatively, one may start from *ab initio* single-particle frameworks and add correlations: self-energy corrections to the single-particle propagator (GW) or corrections in the particle-hole sector via the Bethe-Salpeter equation (BSE) [12–14]. While the BSE provides, in principle, an exact representation of linear response, practical core-level implementations typically rely on approximate (often static) kernels and on a restricted description of the absorber degrees of freedom. Implementations such as Obtaining Core Excitations from *ab initio* electronic structure and NBSE (OCEAN) [12,15] model a photoexcited electron coupled to a localized core hole, yielding a practical core-exciton treatment. For open-shell ions (e.g., Ce^{3+} with $4f^1$), however, the relevant final states span a correlated manifold of ionic configurations coupled to the photoelectron, which is inefficient to represent within the standard two-particle core-exciton picture used in most practical BSE-based calculations. From a formal many-body perspective, the linear density response can be obtained from the four-point polarizability (two-particle Green's function) via the BSE, which is formally exact. [16]

*Contact author: mirone@esrf.fr

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

In practice, however, most core-level implementations rely on a two-particle core-exciton reduction with statically screened interactions and additional approximations, which can hinder the description of dynamical correlation effects (e.g., satellite structures and higher-order excitations) and of strongly correlated open-shell manifolds coupled to the continuum [17]. Frequency-dependent (dynamical) kernels have been explored to recover such effects (including double excitations), at the price of increased complexity and, in approximate schemes, the possible appearance of spurious solutions [18].

Here, we take a complementary route tailored to open-shell absorbers, going beyond the standard static-kernel two-particle core-exciton practice by explicitly embedding a correlated atomic active space and coherently coupling it to the plane-wave (PW)/projector augmented-wave (PAW) photoelectron framework of *xspectra* [19–21] within QUANTUM ESPRESSO [22,23]. PAW pseudopotentials [24,25] are pivotal: Besides accurate pseudopotentials, PAW projectors provide localized channels to map excited atomic states onto the plane-wave pseudobasis.

Unlike the original *XSPECTRA*—which solves a single-particle problem—we represent the photoelectron as a superposition of states, each given by the tensor product of a single-particle plane-wave function and an ionic basis vector $|a_i\rangle$. Here, we discuss how this approach, named *xspectrumlet-mode* within the *XSPECTRA* code, captures both localized multiplet peaks and the continuum component of the giant dipole resonance (GDR), and we illustrate it on the Ce $N_{4,5}$ edges. The term giant dipole resonance originates from photonuclear absorption studies [26,27]; in rare-earth core spectroscopies, it is used by analogy to denote a strong, broad dipolar resonance with substantial continuum character.

A. Intuitive picture

Before developing the full formalism, it is useful to outline, as an illustration of the importance of entanglement, the exchange-driven mechanism underlying the GDR (Fig. 1). At the Ce $N_{4,5}$ edges, a photon with multipolarity (l_{ph}, m_{ph}) promotes a $4d$ electron (l_d, m) to $4f$ $(l_f, m+m_{ph})$, creating a coherent superposition of core-photoelectron pairs (Fig. 1, left). The first-order exchange contribution to the resonance energy can be cast as the trace of the product between the excited-state density matrix and an exchange vertex of rank K (Fig. 1, right). By orthogonality and completeness of Wigner $3j$ symbols [28,29], the dominant contribution arises for $K = l_{ph}$; other K are orthogonal and do not contribute. Consequently, dipole excitations ($l_{ph} = 1$) experience a large shift (of order $2G_1$), whereas higher multipoles are much weaker (e.g., $\sim 0.38 G_3$ and $\sim 0.28 G_5$), with G_K the $4d$ - $4f$ exchange Slater integral of rank K . The hierarchy stems from the structure of the superposition: For $l_{ph} = 1$, many core-electron pairs interfere constructively, whereas for $l_{ph} = 5$ the allowed channels are severely constrained (e.g., for $m_{ph} = 5$, only $m = -2$ is promoted). Crucially, the exchange vertex is a short-range, near-core interaction: It acts only where the photoelectron has finite amplitude inside the absorber and overlaps the highly localized $4d$ hole. In the dipole channel, this contribution is repulsive and large—of order $2G_1 \approx 30$ eV for Ce—so it lifts the would-be localized $4d^{-1}4f^1$ configuration above

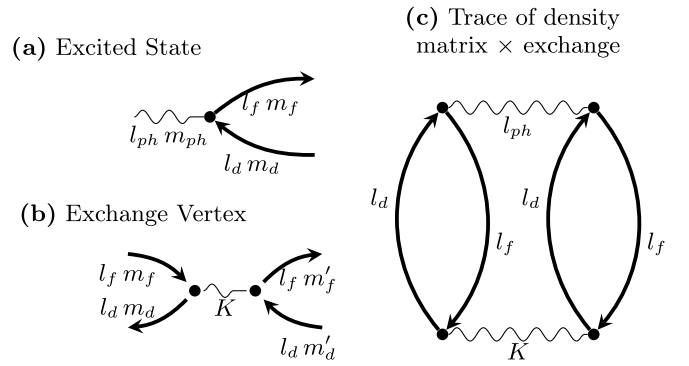


FIG. 1. Bird's-eye view of the leading exchange contribution to the dipolar resonance energy at the Ce $N_{4,5}$ edges. (a) Schematic excitation step: A photon of multipolarity (l_{ph}, m_{ph}) promotes a $4d$ electron into the $4f$ manifold, creating a coherent superposition of core-photoelectron pairs. (b) The exchange vertex. (c) The corresponding first-order exchange shift can be written as the trace of the excited-state density matrix times an exchange vertex of rank K (snake line), where solid arrows denote electronic angular-momentum channels. In spherical symmetry (and after averaging over m_{ph}), orthogonality of Wigner $3j$ symbols selects $K = l_{ph}$, strongly enhancing the dipole ($K = 1$) contribution.

the ionization threshold, endowing it with strong continuum character and producing a GDR. In other words, the exchange term can “jump” the barrier between localized states and the continuum. The effect is also *resonance selective*: It is strong for the dipole ($K = 1$) channel and much weaker for higher multipoles ($K = 3, 5$), so the amount of spectral-weight transfer into the continuum depends on the specific resonance. In our formulation, this near-core exchange arises naturally as an interference effect between the entangled components; the rigorous development follows in Sec. II.

B. Practical relevance: Ce $N_{4,5}$ thresholds and valence assignment

In recent *in situ* studies on ceria-based systems, the Ce $N_{4,5}$ edge has been widely used as a valence fingerprint to quantify Ce^{4+} vs Ce^{3+} fractions [30,31]. While sharp high- Q features are well captured by multiplet models and provide robust constraints on Slater reductions, the low- Q response is dominated by the GDR, whose broad lineshape often presents a low-energy shoulder. If interpreted within a purely localized picture, this shoulder can be mistakenly assigned to a chemically shifted line from reduced species, biasing the inferred $\text{Ce}^{3+}/\text{Ce}^{4+}$ ratio. Our entangled propagation clarifies that this shoulder is a continuum effect—driven by the $K = 1$ exchange rather than a separate atomic peak.

II. THEORY

We consider the scattering atom after excitation of the photoelectron and enumerate by M the number of ionic configurations relevant to the multiplet structure. Each ionic state is a vector $|a_i\rangle$ in a chosen ionic Hilbert basis. The photoelectron is represented by $2M$ spinless PW channels $|\psi^i\rangle$ (spin is encoded in the index i , first M for spin-down, next M for

spin-up), and the total *entangled* state is

$$|\widehat{\psi}\rangle = \sum_{i=1}^{2M} |a_i\rangle \otimes |\psi^i\rangle. \quad (1)$$

In the following, we refer to a generic vector $|\widehat{\psi}\rangle$ as an *entangled state*, and to the specific initial condition $|\psi_0\rangle$ created by the probe as the (*entangled*) *excited wave packet*.

Throughout this work, the term “ionic” refers to localized *electronic* Fock-space configurations of the absorber (core hole and valence/open-shell electrons), not to nuclear/vibrational degrees of freedom. The meaning of the symbols and the construction of the ionic states are detailed below, together with the time evolution operator:

$$\partial_t |\widehat{\psi}\rangle = \widehat{D}_t |\widehat{\psi}\rangle, \quad (2)$$

which defines the formal generator of the evolution operator; its action $e^{t\widehat{D}_t}$ on the (entangled) excited wave packet defined below is evaluated by Lanczos/kernel polynomial method (KPM) recursions, without explicit real-time stepping. Once these two key elements—the wavefunction and the time evolution—are defined, the spectrum is obtained from the formal time evolution of an (*entangled*) *excited wave packet* $|\widehat{\psi}_0\rangle$ generated by applying a probe-specific transition operator to the correlated ground state:

$$|\widehat{\psi}_0(\mathbf{Q})\rangle = \widehat{O}(\mathbf{Q})|g\rangle. \quad (3)$$

Here, $|g\rangle$ is the correlated ground state and $\widehat{O}(\mathbf{Q})$ is adapted to the specific spectroscopy technique (and carries the dependence on momentum transfer and polarization when relevant). For example, in nonresonant inelastic x-ray scattering one may take $\widehat{O}(\mathbf{Q}) = \sum_j e^{i\mathbf{Q}\cdot\mathbf{r}_j}$, whereas in x-ray absorption $\widehat{O}$ reduces to the appropriate dipole/quadrupole operator. All selection rules and Q -dependence therefore enter through $\widehat{O}$ and the construction of $|\widehat{\psi}_0(\mathbf{Q})\rangle$; for notational simplicity, we will often omit the explicit $\mathbf{Q}$ label. In a correlation-function formulation equivalent to Fermi’s golden rule, the spectrum can then be written as

$$f(\omega) = \int_{-\infty}^{\infty} \langle \widehat{\psi}_0 | \widehat{S} e^{t\widehat{D}_t} | \widehat{\psi}_0 \rangle e^{i\omega t} dt, \quad (4)$$

which can be evaluated either by the KPM [32] or via a continued fraction obtained from Lanczos tridiagonalization [33]. Here, $\widehat{S}$ denotes the overlap (metric) operator in the entangled space; it reduces to the PAW overlap for closed shells and is generalized as in Eq. (9) to remove open-shell redundancy. The advantage of KPM is that it is numerically stable, being defined by a recursive relation with eigenvalues <1 , but it may be slower to reach a given resolution because it methodically adds information over the entire spectrum. In fact, in analogy with the Fourier transform, the Chebyshev components span the whole spectral range, so that the number of iterations is proportional to the ratio between the spectral range and the required resolution, even when the final spectrum is sharply peaked in a narrow subrange. In contrast, the Lanczos tridiagonalization may converge rapidly when the spectral weight of the excited wavefunction is confined to a narrow region, but the numerical orthogonality of the generated Krylov space is more fragile. Both methods are implemented.

A. Transition operator and Q dependence

For completeness, we recall the standard Fermi-golden-rule expression for $S(\mathbf{Q}, \omega)$ to make explicit the meaning of $\widehat{O}(\mathbf{Q})$ and the role of $\mathbf{Q}$; this also provides a direct route to the autocorrelation formulation in Eq. (4). For nonresonant inelastic x-ray scattering (XRS/NIXS), the measured intensity is proportional to the dynamic structure factor,

$$S(\mathbf{Q}, \omega) = \sum_f |\langle f | \widehat{O}(\mathbf{Q}) | g \rangle|^2 \delta[\omega - (E_f - E_g)], \quad (5)$$

with $\widehat{O}(\mathbf{Q}) = \sum_j e^{i\mathbf{Q}\cdot\mathbf{r}_j}$ the Fourier component of the electron density (restricted in practice to the core-level excitation subspace on the absorber in our PAW embedding). Since we focus on normalized spectral shapes, the polarization-dependent kinematical prefactor is omitted here. Therefore, all selection rules and the full $\mathbf{Q}$ dependence enter through the construction of $|\widehat{\psi}_0(\mathbf{Q})\rangle$, while the propagator itself is $\mathbf{Q}$ -independent. Expanding $e^{i\mathbf{Q}\cdot\mathbf{r}}$ into spherical Bessel functions and spherical harmonics makes explicit that low Q emphasizes dipolelike ($L = 1$) channels, while increasing Q progressively enhances higher multipoles.

B. Equivalence of golden rule and autocorrelation formulation

Starting from Eq. (5), inserting $\sum_f |f\rangle\langle f| = \mathbb{I}$, using $\delta(x) = \frac{1}{2\pi} \int dt e^{ixt}$, and defining the excited wave packet $|\widehat{\psi}_0(\mathbf{Q})\rangle = \widehat{O}(\mathbf{Q})|g\rangle$ leads to Eq. (4).

C. Meaning of the symbols and numerical organization

In Eq. (1), the ket on the *left* of each tensor product, $|a_i\rangle$, is a *Fock-space* basis vector for the absorber’s localized subspace. Concretely, it is specified in second quantization by the binary occupations (0/1) of all relevant on-site spin-orbitals (n, ℓ, m, σ)—and, when present, of additional radial channels associated with the same ℓ (e.g., $4f, 5f, 6f$, or $6p$ arising from multiple PAW projectors). In this sense, $|a_i\rangle$ compactly encodes an *ionic configuration* (including the core hole) without any real-space expansion: It is a discrete label that fixes which localized spin-orbitals are occupied in that channel [34]. The ket on the *right*, $|\psi^i\rangle$, is instead a PW pseudowavefunction living in the one-electron space of the photoelectron. Unlike the ionic factor, it is a freely evolving object: Its components change in time under the entangled dynamics introduced below, and it can acquire any spatial form allowed by the PW representation. The index i therefore labels *channels* formed by an ionic basis state tensored with an associated PW wavefunction; by construction, the sum in Eq. (1) runs over all relevant ionic labels (and both spin channels).

From a computational standpoint, all PW wavefunctions $\{|\psi^i\rangle\}_{i=1}^{2M}$ are handled *as a block*. We reuse the existing QUANTUM ESPRESSO infrastructure that applies linear operations to sets of PW states (as commonly done for multiple bands); here, the same batched machinery acts on *channels* rather than on eigenbands. In practice, the “band index” of the PW engine is repurposed to enumerate the $2M$ entangled channels $|\psi^i\rangle$ attached to the different ionic labels $|a_i\rangle$.

The coupling between channels—responsible for mixing amplitudes associated with different ionic configurations—is provided by precomputed scattering matrices imported from a multiplet code. These matrices encode the *local atomic physics*, namely, Coulomb multipole interactions and spin-orbit coupling.

D. Local charge fluctuations in the open shell

The entangled form in Eq. (1) also implies that the method naturally explores *local charge fluctuations* on the absorber. Concretely, consider an ionic label whose open shell contains n electrons. Whenever the associated PW factor carries nonzero amplitude on the absorber's open-shell orbitals, the *local* occupancy of that shell is effectively $n+1$: One electron comes from the ionic label and one from the PW component that momentarily resides in the localized channel. Conversely, when the PW amplitude has no overlap with those localized orbitals, the shell remains at occupancy n . As the PW wavefunctions evolve, amplitude is continuously transferred between these two situations.

E. Active orbitals of the Hilbert space

The Hilbert space is constructed by the multiplet code, together with the scattering factors. It is determined by the user-choice concerning the core shell ($4d$ in our examples), the open shell (here, $4f$) and its base occupancy n (here, $n = 0, 1$), and the complementing active orbitals which in our examples are $5f$, $6f$, and $6p$. They must correspond to existing PAW orbitals on whose basis the multiplet code computes matrix elements, using the all-electron wavefunctions. This procedure is further illustrated by the worked-out examples, which come with the updated version of xspectra integrated in the QUANTUM ESPRESSO (QE) suite [35].

In constructing the Hilbert basis, we impose the following constraints: the occupancy in the open shell fluctuates between n and $n+1$; the core shell contains one hole; the remaining active shells contain zero or one electrons; and finally, the total number of electrons fluctuates between n and $n+1$, this without counting the core shell which has one hole.

F. PW plane-wave single-particle base model and spin treatment

In the present implementation, the plane-wave single-particle basis is obtained from a *spinless* self-consistent (SCF) calculation. The underlying potential and the PAW projectors define a single-particle framework that does not carry an explicit spinor structure. The spin degree of freedom is instead introduced at the level of the entangled Hamiltonian, through the scattering terms imported from the multiplet code. These matrices include both Coulomb multipole interactions and spin-orbit coupling, so that the relevant spin physics—particularly the core and open-shell splittings that shape the experimental spectra—is incorporated via the atomic part of the problem. This design reflects the fact that, in core-level x-ray spectroscopies, the experimentally visible spin dynamics originates almost entirely from near-core, localized effects, whereas the spin behavior of the extended photoelectron plays only a secondary role.

In future developments, nothing prevents extending the framework to use a fully spinor plane-wave single-particle base model. A noncollinear (or fully relativistic) SCF calculation would naturally provide two-component spinor wavefunctions, allowing one to describe spin-orbit effects also in the delocalized region, while retaining the same entangled construction presented here.

G. Why a Pauli projector is needed

In each entangled channel $|a_i\rangle \otimes |\psi^i\rangle$, the PW factor $|\psi^i\rangle$ is free to explore the full plane-wave space. When we look at its *local* content on the absorbing atom via PAW, nothing *a priori* prevents $|\psi^i\rangle$ from acquiring a finite projection onto an orbital that is *already occupied* in the ionic labeled configuration $|a_i\rangle$. That would amount to placing two indistinguishable electrons (with the same spin as channel i) on the same localized orbital—i.e., a violation of Pauli. We therefore enforce, for each channel, a Pauli projector that removes from $|\psi^i\rangle$ any component parallel to the subspace spanned by the occupied local orbitals of $|a_i\rangle$.

H. PAW notation

Within PAW, we denote by $|\beta_\eta\rangle$ the projectors centered on the absorber (distinguished by angular momentum and, if needed, radial index), and by $|\phi_\eta\rangle$ the corresponding pseudo partial waves that are *dual* to the projectors, i.e., $\langle\beta_\eta|\phi_{\eta'}\rangle = \delta_{\eta\eta'}$. These provide localized channels to test and manipulate the amplitude of $|\psi^i\rangle$ on specific atomiclike orbitals (e.g., the $4f$ manifold). Because we use non-norm-conserving pseudopotentials, scalar products in PW space are taken with the overlap operator S^{PW} .

I. Definition and properties

Let $\text{Occ}(a_i)$ be the set of PAW channels corresponding to orbitals that are occupied in the ionic state $|a_i\rangle$ with the same spin as channel i . The Pauli projector acting on $|\psi^i\rangle$ is

$$P_i^p = I - \sum_{\eta \in \text{Occ}(a_i)} \frac{|\phi_\eta\rangle\langle\phi_\eta| S^{\text{PW}}}{\langle\phi_\eta| S^{\text{PW}} |\phi_\eta\rangle}. \quad (6)$$

In words, P_i^p subtracts from $|\psi^i\rangle$ its component along each occupied local orbital, where “component” and “orthogonality” are meant in the S^{PW} metric appropriate for PAW. All other components are left unchanged. In the Cerium case considered here, the sum typically runs over the occupied $4f$ orbitals (and may be extended to any additional local shells explicitly included).

J. Application during time evolution

We apply P_i^p whenever the formal evolution operator is applied to the (entangled) excited wave packet (e.g., within the Lanczos or KPM recursion). This guarantees that, even if the dynamics mixes channels, no PW factor $|\psi^i\rangle$ can ever develop a nonzero projection onto an orbital that is already occupied in its ionic label $|a_i\rangle$. For the full entangled state, we

use the block-diagonal operator

$$\hat{P}^p = \sum_{i=1}^{2M} |a_i\rangle \langle a_i| P_i^p, \quad (7)$$

which enforces Pauli independently in each channel.

K. Redundancy and scalar product

The entangled representation in Eq. (1) is deliberately redundant whenever the photoelectron has nonzero amplitude on the *open-shell* atomic subspace. Intuitively, an ionic configuration in which $(n+1)$ electrons sit on an open shell can be realized in $n+1$ equivalent ways by deciding which of the indistinguishable electrons we call “the photoelectron.” Our tensor-product construction materializes these $n+1$ choices as $n+1$ different channels $|a_i\rangle \otimes |\psi^i\rangle$. As a result, the *atomic-projected part* of the state would be counted multiple times in naive norms and overlaps, unless we correct for this redundancy.

L. A concrete example ($n = 1$)

Consider an ion with a single $4f$ electron ($n = 1$). Let one ionic label be $|a_{(z^3)}\rangle$, meaning that the $4f$ electron occupies (schematically) the z^3 orbital. If $|\phi_{xyz}\rangle$ is the PAW pseudo partial wave corresponding to a $4f$ orbital, the product $|a_{(z^3)}\rangle \otimes |\phi_{xyz}\rangle$ represents one way of forming a two-electron $4f$ configuration with orbitals $\{z^3, xyz\}$. However, the *same physical configuration* also appears in a different channel whose ionic label is $|a_{(xyz)}\rangle$ and whose PW factor carries the z^3 component. Thus, when we sum over channels, the atomic-projected contribution associated with $\{z^3, xyz\}$ is counted *twice*. For a general open shell with occupancy n , any $(n+1)$ -electron configuration obtained by placing the photoelectron in the open-shell subspace is represented $n+1$ times across the entangled sum.

This is why only the part of $|\psi^i\rangle$ that lives in the open-shell atomic subspace needs a correction: Components orthogonal to that subspace (i.e., truly delocalized or with different l) are unique and are not overcounted.

M. Projector onto the open-shell subspace

To isolate the redundant piece, we introduce the PAW projector

$$P^a = \sum_{\eta \in \text{o.s.}} \frac{|\phi_\eta\rangle \langle \phi_\eta| S^{\text{PW}}}{\langle \phi_\eta| S^{\text{PW}} |\phi_\eta\rangle}, \quad (8)$$

where the sum runs over all PAW channels (η) that span the *open-shell* on the absorber (e.g., the $4f$ manifold in Ce), $|\phi_\eta\rangle$ are the dual pseudo partial waves, and S^{PW} is the PW overlap operator appropriate for PAW. The operator P^a extracts, from any PW state, precisely the localized component that gives rise to redundancy.

N. Redundancy-corrected scalar product

With this notation, a redundancy-safe norm (and, more generally, overlap) for the entangled state is obtained by summing the standard PW metric over channels and then

subtracting the overcounted atomic-projected weight with the correct combinatorial factor.

O. Single open shell with fixed n

A redundancy-safe scalar product is then

$$\langle \hat{\psi} | \hat{S} | \hat{\psi} \rangle = \sum_{i=1}^{2M} \left(\langle \psi^i | S^{\text{PW}} | \psi^i \rangle - \frac{n}{n+1} \langle \psi^i | P^a S^{\text{PW}} P^a | \psi^i \rangle \right). \quad (9)$$

The coefficient $n/(n+1)$ does exactly what the example suggests: For $n = 0$ (closed shell, e.g., Ce^{4+}), there is no redundancy and the correction vanishes; for $n = 1$, each atomic-projected piece is present twice, so we subtract one half of its total weight, leaving a single physical copy; in general, every $(n+1)$ -electron open-shell configuration appears $n+1$ times, and subtracting $n/(n+1)$ of its total restores a single representation.

P. PW Hamiltonian and ionic Hamiltonians

The one-particle PW Hamiltonian H^{PW} (from QE) governs the photoelectron dynamics; for non-norm-conserving pseudopotentials (PPs) [36], the generalized Schrödinger equation, without considering the ionic Hamiltonians but only the QE part, reads

$$\partial_t^{\text{QE}} |\psi\rangle = -i(S^{\text{PW}})^{-1} H^{\text{PW}} |\psi\rangle, \quad (10)$$

which, together with Lanczos tridiagonalization to compute the spectra, is the basis of the former xspectra implementation. Here, we consider, beside this term, also the action of the ionic Hamiltonian, which we split into $H^{A_{n+1}}$, acting on the PAW projections of $|\psi^i\rangle$ over the open shell plus extra projectors, and H^{A_n} for the sector where the photoelectron does not project over the ionic orbitals. Their matrix elements are precomputed with a multiplet code (Hilbert++ [8,10]) considering electron-electron Coulomb interaction and spin-orbit. The relevant sources are distributed with QE.

The Hamiltonian $\hat{H}^{A_{n+1}}$ is

$$\hat{H}^{A_{n+1}} = \sum_{\kappa, i, j, \eta} |\beta_\kappa\rangle \langle a_i| \frac{h_{\kappa, i, j, \eta}}{f_\kappa f_\eta} \langle a_j | \langle \beta_\eta |, \quad (11)$$

where the sum over κ, η spans all PAW projectors. The factors f_κ and f_η compensate the redistribution induced by $\hat{R}$ [defined in Eq. (13) in the open-shell sector:

$$f_\xi = \begin{cases} n+1, & \xi \in \text{o.s.}, \\ 1, & \text{otherwise}, \end{cases} \quad (\xi = \kappa, \eta).$$

Consequently, terms involving two open-shell projectors carry an overall prefactor $1/(n+1)^2$, while terms with exactly one open-shell index (i.e., $\kappa \in \text{o.s.}$ and $\eta \notin \text{o.s.}$ or vice versa) carry $1/(n+1)$. In short, the $h_{\kappa, i, j, \eta}$ coefficients describe the many-body interactions of the ion and the photoelectron considered together.

The Hamiltonian $\hat{H}^{A_n}$ acts on the subspace where the photoelectron does not project on the absorbing site:

$$\hat{H}^{A_n} = \sum_{i, j} |a_i\rangle h'_{i, j} \langle a_j | \left(S^{\text{PW}} - \sum_{\eta} |\beta_\eta\rangle \langle \beta_\eta | \right). \quad (12)$$

This operator satisfies two requirements: (1) Its expectation value is zero for a wavefunction whose PW part is a linear combination of $|\phi_\eta\rangle$ (which are dual to projectors); and (2) it recovers its *n-occupancy* limit when the PW wavefunction has no projection on the ion. In short, the $h'_{i,j}$ coefficients describe the many-body interactions within the ion without the photoelectron.

Q. Redistribution operator

As discussed above, the atomic-projected part of a given $(n+1)$ -electron open-shell configuration appears $n+1$ times across different channels. Our *ansatz* is that all these *equivalent representations* carry the *same* complex amplitude at any time, i.e., the coefficient of a physical configuration must not depend on how we label “which electron is the photoelectron.” To enforce this indistinguishability in the entangled representation at each application of the time-evolution operator, we apply to its result the following operator:

$$\hat{R} = I + \sum_{a_i, \eta \in \text{o.s.}} \sum_{(a_j, \kappa) \equiv' (a_i, \eta)} s_{j\kappa i\eta} |a_j\rangle \frac{|\phi_\eta\rangle \langle \phi_\kappa| S^{\text{PW}} \langle a_i|}{\|\phi_\eta\|^{1/2} \|\phi_\kappa\|^{1/2}}, \quad (13)$$

where the terms appearing in the denominator are defined as $\|\phi_\eta\| \equiv \langle \phi_\eta | S^{\text{PW}} | \phi_\eta \rangle$, and the condition $(a_j, \kappa) \equiv' (a_i, \eta)$ means that the two-channel-projector labels are two different but equivalent representations of the *same* $(n+1)$ -electron open-shell configuration. The sign $s_{j\kappa i\eta}$ enforces the correct Fermi statistics. The operator acts only in the atomic-projected sector spanned by the PAW partial waves $\{|\phi_\eta\rangle\}$ corresponding to the open shell ($\eta \in \text{o.s.}$).

R. Why it is needed (illustrative view)

Consider the $n = 1$ case with two distinct $4f$ orbitals, say z^3 and xyz (as in the example above). The same local two-electron configuration $\{z^3, xyz\}$ appears in two channels: (1) the channel with ionic label $|a_{(z^3)}\rangle$, whose PW factor has a component along $|\phi_{xyz}\rangle$, and (2) the channel with ionic label $|a_{(xyz)}\rangle$, whose PW factor has a component along $|\phi_{z^3}\rangle$. During time evolution, scattering may create (or modify) the amplitude of $\{z^3, xyz\}$ *asymmetrically* in these two channels. If left unchecked, the two “copies” of the *same* physical configuration would drift to different coefficients purely due to labeling. The action of $\hat{R}$ restores equality of those coefficients by redistributing the increment from one copy to all its equivalents, so that the configuration $\{z^3, xyz\}$ is represented with a single, consistent amplitude, independent of channel labels.

Operational use. We apply $\hat{R}$ after each application of the formal evolution operator to the excited wave packet $|\psi_0\rangle$ (e.g., within the Lanczos or KPM recursion): We act on the block of PW channels, enforce Pauli in each channel, then symmetrize with $\hat{R}$. This keeps the redundancy-corrected scalar product consistent and guarantees that numerical results do not depend on arbitrary choices in channel labeling. To compensate for the demultiplexing effect of the redistribution operator on the strength of matrix elements, we use the factors f_κ and f_η in $\hat{H}^{A_{n+1}}$ as defined above.

S. Entangled time evolution

Based on the above-defined operators, the time evolution of the entangled state is

$$\partial_t |\hat{\psi}\rangle = \hat{D}_t |\hat{\psi}\rangle, \quad (14)$$

$$\hat{D}_t = -i \hat{R} \hat{P} (\hat{S}^{\text{PW}})^{-1} (\hat{H}^{\text{PW}} + \hat{H}^{A_n} + \hat{H}^{A_{n+1}}). \quad (15)$$

Here, $\hat{S}^{\text{PW}}$ and $\hat{H}^{\text{PW}}$ are direct sum, one of S^{PW} and the other of H^{PW} , over the ionic channels. This equation, together with Eq. (4), defines the spectra.

III. COMPUTATIONAL SETUP

The calculation begins with the evaluation of the Slater integrals involving the core-hole orbital, the open-shell wavefunctions, and the all-electron wavefunctions associated with those pseudopotential projectors that are included in the definition of the ionic Hamiltonians. When the core-hole orbital is not part of the pseudopotential, it is generated by a self-consistent spherical-atom calculation, which also provides the corresponding spin-orbit coupling strength.

With these ingredients, a straightforward multiplet calculation is performed to determine the ionic ground state. The initial (entangled) excited wave packet is then constructed by applying the excitation operator to this ground state and subsequently projecting the resulting state onto the plane-wave representation through the PAW wavefunctions.

To provide an accurate description of the crystal environment, the ionic Hamiltonians also require the crystal-field contribution. These ingredients—the crystal field and the corresponding density matrix—can optionally be extracted from a preliminary SCF calculation followed by a Wannierization of the open-shell orbitals, together with the rotation of both the one-particle Hamiltonian and the density matrix into the Wannier basis. Using the resulting open-shell density matrix, we then remove the double-counting term associated with the nonspherical mean-field electron-electron interaction within the open shell. This correction is necessary to avoid counting this contribution twice: once through the crystal-field term and again through the explicit many-body interactions.

Alternatively, in the example discussed below, we adopt empirical tight-binding Slater-Koster parameters [10] to model the crystal field in the ground-state multiplet calculation. The resulting density matrix is then used to subtract the double-counting contribution from the $h_{\kappa,i,j,\eta}$ and $h'_{i,j}$ coefficients before performing the *xspectrumlet* calculation.

The crystal field plays an important role not only through its overall strength but, more importantly, through its ability to orient the open-shell orbitals in the ground state. When these orbitals are partially occupied, such orientation may induce an angular dependence in the spectra. In our test case, which concerns the highly localized $4f$ shell, we employed an arbitrary crystal-field strength of less than one tenth of an eV in the Slater-Koster parametrization. This magnitude is sufficient to enforce a well-defined orbital orientation once the crystallographic positions of the neighboring oxygen atoms are provided to the multiplet code.

The spherical component of the ionic Hamiltonians is then made neutral with respect to the plane-wave Hamiltonian.

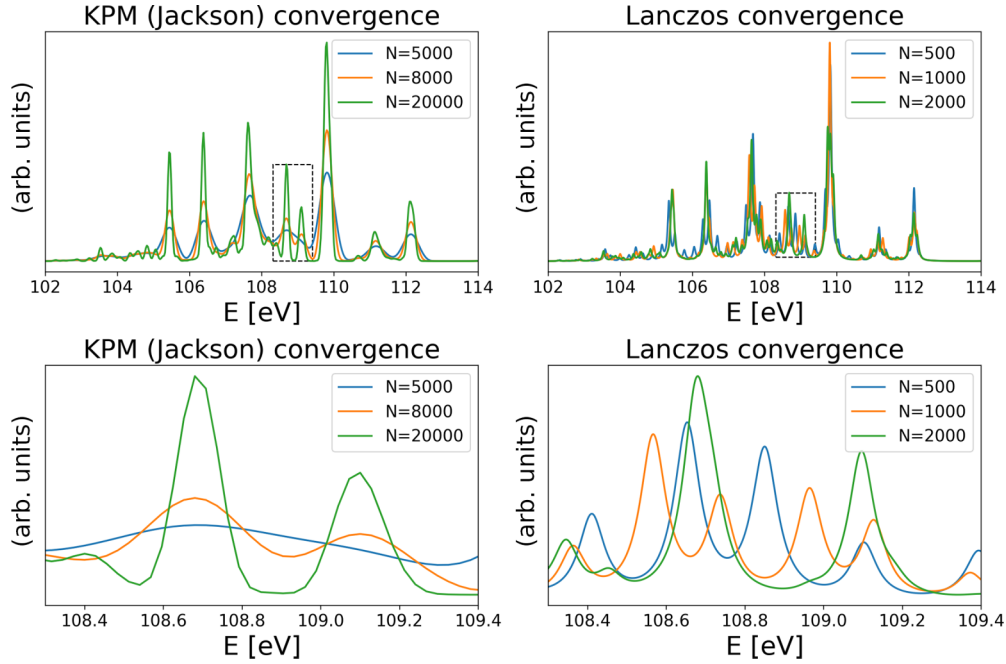


FIG. 2. Convergence of the $\text{Ce}^{4+} N_{4,5}$ XRS spectrum showing only the high- Q part (red curve in Fig. 3) with respect to the number of KPM recursion steps N_C (left) and the number of Lanczos steps N_L (right). Top: full energy range; bottom: close-up of 108.3–109.4 eV (dashed box in the top panels). Curves correspond to $N_C = 5000, 8000, 20\,000$ and $N_L = 500, 1000, 2000$.

This is achieved by setting the trace of $h'_{i,j}$ to zero, while for the coefficients $h_{\eta,i,j,\eta}$, the offset is applied separately for each value of η .

Convergence and resolution control (Lanczos versus KPM)

In practice, evaluating Eq. (4) requires truncating the iterative expansion. In the Lanczos approach, the correlation function is represented by a continued fraction obtained from the Krylov subspace generated by $\hat{D}_t$ acting on $|\psi_0\rangle$; the recursion is therefore stopped after N_L steps. Convergence is assessed empirically by monitoring the change of the spectrum upon increasing N_L (e.g., requiring that $f_{N_L+\Delta N}(\omega)$ differs from $f_{N_L}(\omega)$ by less than a user-defined threshold over the energy window of interest). As is well known, Lanczos can suffer from a progressive loss of orthogonality in finite precision arithmetic, which may lead to numerical instabilities or spurious features if pushed to very large N_L . This can be mitigated by full or selective reorthogonalization, at the price of additional computational and memory overhead, which can become prohibitive in large-scale high-performance computing (HPC) runs.

In the KPM, the effective operator is first rescaled so that its spectrum lies within the domain of the Chebyshev recursion, and the spectral resolution is then controlled *deterministically* by the expansion order N_C (or number of steps of the recursion relation), together with the chosen kernel/broadening. Thus, for a target resolution ΔE over a fixed spectral range, N_C is fixed *a priori*, and the procedure does not rely on an empirical stopping criterion. In practice, the two methods can be used in a complementary way: One may (1) use KPM alone when a guaranteed target resolution is desired, or (2) use KPM as a numerically stable reference to calibrate the minimum

N_L needed by Lanczos to reach the same effective resolution and lineshape, thereby providing an *a posteriori* validation of Lanczos stability.

To make the convergence behavior explicit, Fig. 2 shows only the high- Q part of the $\text{Ce}^{4+} N_{4,5}$ edge (red curve in Fig. 3, i.e., CeO_2) computed with increasing KPM order N_C (left) and Lanczos depth N_L (right). The top panels span the full $N_{4,5}$ window, while the bottom panels show a close-up of 108.3–109.4 eV. The Chebyshev (KPM) spectra are shown without any extra convolution. For Lanczos, the continued fraction is evaluated with $\eta = 0.04$ eV (i.e., the imaginary part added to the spectral energy to shift away from the poles), which corresponds to a Lorentzian FWHM of 0.08 eV. For KPM, the Jackson kernel [32] introduces a Gaussian-like effective lineshape whose width narrows as N_C increases. These effective widths are intentionally much narrower than the physical Lorentzian broadenings used in Fig. 3 (FWHM 1.5 eV at high Q and 3 eV at low Q), but they allow a richer illustration of fine multiplet details.

From the comparison, one can see that Lanczos yields localized peaks rapidly, but intermediate (step-dependent) spurious features and small oscillations in peak positions appear as N_L varies. At the physical linewidths of Fig. 3, these effects have little impact, and in this test case convergence is already reached with about $N_L \sim 200$ iterations. KPM instead converges in a more systematic, Fourier-like fashion (Chebyshev polynomials replacing trigonometric functions), and the Jackson kernel suppresses truncation (Gibbs) oscillations; the trade-off is that a larger N_C is required to reach comparable resolution. Pushing the KPM recursion to very large N_C can eventually reveal isolated, very narrow peaks. In principle, the Chebyshev recursion can also be used as a spectral filter to construct an approximate wavefunction associated with such

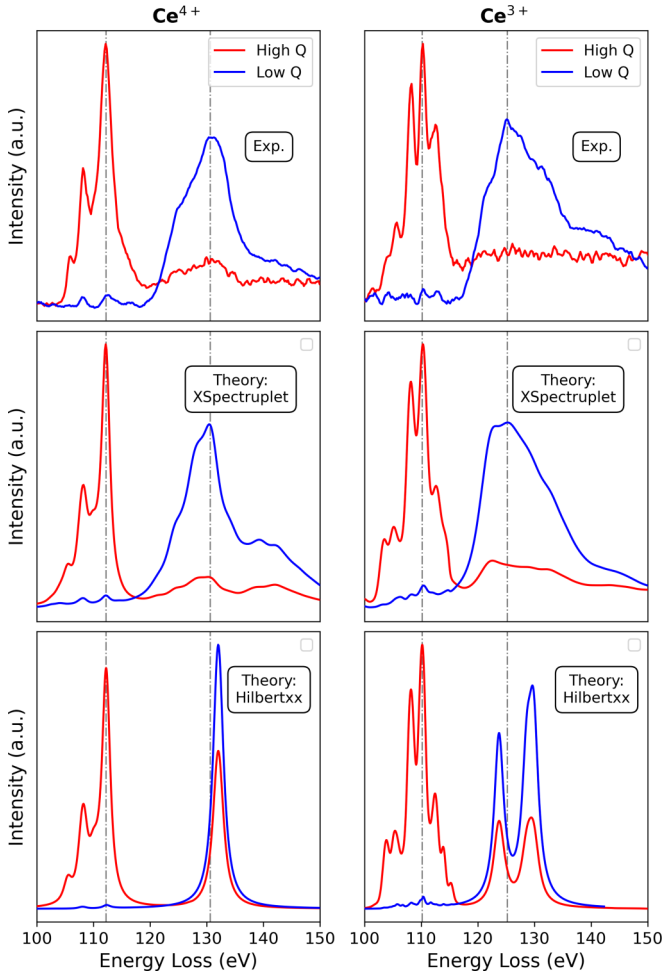


FIG. 3. XRS at the Ce $N_{4,5}$ edge for CeO_2 (left) and $\text{Ce}_2(\text{SO}_4)_3$ [31] (right): experiment (top), xspectrumlet (middle), and multiplet (bottom). Red: high Q ($9.5 \text{ \AA}^{-1}$); blue: low Q ($3.5 \text{ \AA}^{-1}$). The entangled propagation reproduces localized peaks and continuum GDR structure with correct energy and lineshape, including low-energy shoulders otherwise absent in pure multiplet.

an isolated feature, or more generally a coherent superposition of states within a narrow energy window around a target resonance energy E_{res} [32,37]. This provides a practical way to analyze the contribution of final states within a selected energy window (e.g., a given peak) without explicitly constructing or diagonalizing the full final-state manifold, which would be prohibitively expensive in the present plane-wave multichannel setting.

IV. RESULTS AND VALIDATION

We show in Fig. 3, the experimental (top row) and computed (middle row) XRS spectra at the $N_{4,5}$ edge of cerium for CeO_2 and cerium(III) sulfate [$\text{Ce}_2(\text{SO}_4)_3$] [31]. In the bottom line of Fig. 3, the spectra have been calculated switching off the PW Hamiltonian, and considering only the $4f$ projectors and pseudofunctions, thus reproducing a simple multiplet calculation.

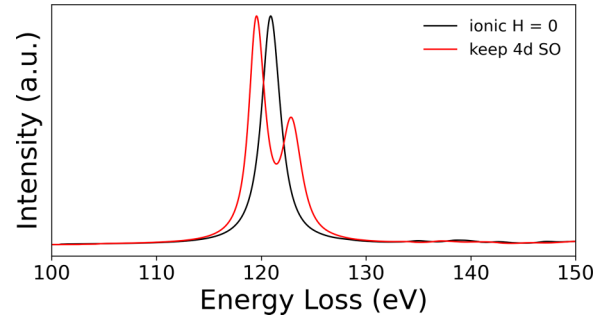


FIG. 4. Ce^{4+} $N_{4,5}$ spectrum obtained by setting all ionic-Hamiltonian coefficients to zero (black, pristine xspectra-like baseline). The curve is broadened with FWHM 1.5 eV below 120 and 3 eV above, so the $4f$ crystal-field splitting is not resolved and only one strong peak remains. If the $4d$ spin-orbit term is retained (red), the expected $\sim 3.3 \text{ eV}$ $N_{4,5}$ splitting appears while the photoelectron remains in a single-particle-like picture.

Independent-particle baseline and why it is not shown

A natural reference would be a one-particle plane-wave/PAW spectrum as provided by the standard xspectra workflow. However, in its current form xspectra is not directly applicable to the Ce $N_{4,5}$ thresholds with the PAW datasets employed here, and extending it to this edge would require additional dedicated setup for the $4d$ core channel. More importantly, even if such an independent-particle calculation were performed, the resulting lineshape would largely reflect (1) the single-particle crystal-field splitting of the $4f$ -derived final states (typically sub-eV) and (2) the $4d$ spin-orbit separation between the N_5 and N_4 components (of order a few eV), while missing the correlated $4d^{-1}4f^{n+1}$ multiplet manifold and the resonance-continuum mixing that redistribute oscillator strength over a $\gtrsim 20 \text{ eV}$ window and produce the giant dipole resonance. For this reason, we use as controlled baselines the purely local multiplet limit obtained by switching off $\hat{H}^{\text{PW}}$ (bottom panels of Fig. 3) and the full entangled propagation including $\hat{H}^{\text{PW}}$ (middle panels). To still provide a comparison with a pristine xspectra-like independent-particle calculation, we also construct a proxy spectrum by setting to zero all coefficients of the ionic Hamiltonian. In this limit, channels do not interfere because they differ in the core-hole quantum numbers of the final state. The resulting Ce^{4+} spectrum is shown in Fig. 4; after applying the same phenomenological broadening used for the Ce^{4+} high- Q curve, it consists of one strong $4f$ peak (the crystal field being not resolved due to the peak broadening).

The first compound contains $\text{Ce } 4f^0$ ions, for which Eq. (15) propagates two particles: the core hole and the photoelectron. The second compound instead contains $\text{Ce } 4f^1$ ions, for which Eq. (15) describes an entangled three-body final state (core hole, open-shell electron, and photoelectron), i.e., a multichannel extension of the standard core-exciton treatment.

Given the close chemical similarity and the nearly identical ionic radii of trivalent Ce and Pr, we used the experimentally established structure of $\text{Pr}_2(\text{SO}_4)_3$ as a proxy for $\text{Ce}_2(\text{SO}_4)_3$. The graph shows the energy loss spectra for an exchanged Q of $9.5 \text{ \AA}^{-1}$ (red line, high Q) and $3.5 \text{ \AA}^{-1}$ (blue line,

low Q). The procedure to reproduce the calculation with the QUANTUM ESPRESSO package is provided in the examples for the *xspectrumlet* mode, which accompany the QE sources [35] (see Appendix A). In all the calculations, the Slater integrals are rescaled by a reduction factor equal to 0.85 for all the Slater integrals in the case of Ce^{4+} , and equal to 0.75 for Ce^{3+} . The spectra in Fig. 3 were computed with KPM (Jackson kernel) using $N_C = 2000$ recursion steps, and are then convoluted with a FWHM = 1.5 eV Lorentzian for the high- Q part and with FWHM = 3 eV for the low- Q part. This choice was motivated by the agreement with the experimental shape.

Concerning Ce^{4+} , the extra shoulder in the low-energy part of the resonance emerges naturally as a result of the formalism. This is an important result, which clarifies how to interpret correctly these features in relation to the oxidation state of the system.

The high- Q part is well reproduced both in shape and in the energy positions of the peaks. These positions strictly determine the Slater reduction factor. With the above so determined Slater reduction factors (one for Ce^{4+} and another one for Ce^{3+}), the energy position of the GDR peak is well reproduced in the *xspectrumlet* calculation but is found at higher energy in the simple multiplet calculation. This is an effect of the GDR weight on components belonging to the continuum, all having a minor overlap with $4d$ than $4f$ has, hence a lesser exchange with $4d$. Similar effects, where the renormalization of the effective Slater integrals depends on the resonance, have been already observed in the literature and ascribed to the hybridization with states in the continuum, reproducing the same behavior with a multiplet calculation where the atomic orbitals of the model Hamiltonian have been complemented by a discrete number of fictitious states [38].

Computational details for the studied cases are reported in Appendix B.

The numerical complexity is dominated by the PW dimension per channel ($N_{\text{PW}} \sim 10^5$) and by the number of channels (Ce^{4+} : $N_{\text{ch}} = 20$; Ce^{3+} : $N_{\text{ch}} = 280$); full details are given in Appendix C.

V. DISCUSSION AND OUTLOOK

We have developed a quantum-mechanical formalism that treats the coherent electronic entanglement between a plane-wave photoelectron and a correlated manifold of localized ionic configurations of the absorber in *ab initio* core-level spectroscopy calculations.

Our method has allowed the correct interpretation of the GDR features in cerium oxide systems, in particular their shape and energy position, and has the potential to shed light on the “gray zone” between transitions to well-localized states and transitions to the continuum.

As shown in this work, a particularly relevant example is the GDR in rare-earth elements, where electronic transitions involve deep-lying orbitals (often $4f$ and $5d$), which show an intriguing interplay between tightly bound excitations and partial escape into the continuum. Similar phenomena also appear in uranium compounds and in $2p \rightarrow 3d$ transition, where threshold effects and state mixing can make it challenging to

distinguish between discrete energy levels and those that “spill over” into higher-energy continua.

All these observations highlight how our approach, focused on exploring phenomena in this “hybrid” energy range, can pave the way for both the development of more refined theoretical models and the design of experiments aimed at exploiting these feature-rich spectroscopic domains.

ACKNOWLEDGMENTS

We thank Matteo Calandra, Joshua Elliot, and Francesco Mauri for the advice and help during the integration of *xspectrumlet* within QUANTUM ESPRESSO, and the QUANTUM ESPRESSO community for providing a well-written and well-documented suite that enables smooth implementation of ideas. We acknowledge the European Synchrotron Radiation Facility (ESRF) for provision of synchrotron radiation facilities under Proposals No. IH-CH-1633 and No. IH-CH-1634, and we would like to thank F. Gerbon for assistance and support in using beamline ID20.

DATA AVAILABILITY

The data that support the findings of this article are openly available [35].

APPENDIX A: INSTALLATION AND RUNNING THE EXAMPLES

1. Repository overlay and build

The *xspectrumlet* source distribution is provided as an overlay on top of the official QE 7.4 sources. The repository (see Ref. [35]) contains only the modified/added files; these must be copied on top of a pristine QE 7.4 tree (or a prepatched QE 7.4 tarball from the project wiki). After overlaying, QE can be built with the standard CMake workflow described in the repository README (Conda-based build recommended), followed by installation of the Hilbertxx tools via `XSpectra/tool_hilbertxx/install.sh`.

2. Running the Ce^{4+} / Ce^{3+} demos

Reproducible examples are provided in the following directories (relative to the QE root):

- (1) `XSpectra/examples_hilbertxx/demo_ce4p`
- (2) `XSpectra/examples_hilbertxx/demo_ce3p`

Each demo contains the SCF input `scf_enes.qe.in`, the *xspectrumlet* input `xspectrumlet_input.qe.in` (default: KPM/Chebyshev), and a Slurm launcher `launch_xspectrumlet.sh`. To switch to Lanczos, the input file comments indicate that one should remove the three Chebyshev-related lines. The Chebyshev coefficients can be post-processed with the provided `jackson.py` script. Full, step-by-step instructions are available in `INSTRUCTIONS_DEMO_Ce4p.md` and `INSTRUCTIONS_DEMO_Ce3p.md`.

APPENDIX B: COMPUTATIONAL DETAILS FOR THE STUDIED CASES

SCF calculations were performed at the Γ point (one k point) using large supercells; we verified that this cell size is sufficient so that additional k points do not change the spectra. A relatively large Gaussian smearing (degauss = 0.1 Ry) was used to facilitate SCF convergence; x-ray spectral features in the energy range of interest are insensitive to this choice.

(1) CeO_2 (Ce^{4+}): $a = 20.7536$ a.u., $V = 8936.73$ (a.u.)³, 96 atoms, 768 electrons, 1400 KS states, $N_{\text{PW}} = 157\,683$ plane waves.

(2) $\text{Ce}_2(\text{SO}_4)_3$ (Ce^{3+}): $a = 18.3005$ a.u., $V = 13486.34$ (a.u.)³, 102 atoms, 684 electrons, 2000 KS states, $N_{\text{PW}} = 238\,273$ plane waves.

Plane-wave cutoffs were $E_{\text{cut}}^{\psi} = 103$ Ry and $E_{\text{cut}}^{\rho} = 1100$ Ry in both cases. The spectra were calculated with KPM (Chebyshev expansion) using $N_{\text{C}} = 2000$ moments and the Jackson kernel.

The PAW setup for the absorber includes nine projectors ($l = 0, 1, 2, 3$); the $l = 3$ (f) and $l = 1$ (p) channels enter the active space of the ionic $N + 1$ Hamiltonian. Three f -type projectors are used ($4f, 5f, 6f$; the dominant channel), with reference energies $\varepsilon_f = (-0.7649, 2.0045, 6.30)$ Ry. The extra projectors are placed at sufficiently separated energies to avoid near-linear dependence, following Ref. [19].

APPENDIX C: NUMERICAL COMPLEXITY

For the systems studied here, the entangled basis comprises $N_{\text{ch}} = 20$ ionic channels for CeO_2 (Ce^{4+}) and $N_{\text{ch}} = 280$ for

$\text{Ce}_2(\text{SO}_4)_3$ (Ce^{3+}). The full entangled matrix dimension is larger than N_{ch} because each channel is expanded over the PAW projector functions (27 in our setup), giving 540 (Ce^{4+}) and 7560 (Ce^{3+}), when the photoelectron is considered part of the ion. The PAW overlap and Hamiltonian in plane-wave space, S^{PW} and H^{PW} , act on each channel's PW coefficients; the plane-wave dimension per channel is $N_{\text{PW}} \approx 1.58 \times 10^5$ for CeO_2 and 2.38×10^5 for $\text{Ce}_2(\text{SO}_4)_3$, and these operators are applied in operator form fast fourier transform and PAW (FFT/PAW) rather than stored as dense matrices. The ionic/multiplet Hamiltonian blocks stored from the multiplet stage have sizes 540 (Ce^{4+}) and 7560 (Ce^{3+}), with $\sim 2.6 \times 10^4$ and $\sim 1.0 \times 10^6$ nonzero elements, respectively. Typical production values are $N_{\text{L}} \sim 200$ (Lanczos) and $N_{\text{C}} = 2000$ (KPM) for the spectra in Fig. 3.

APPENDIX D: PRISTINE XSPECTRA-LIKE BASELINE

To provide a direct reference to a pristine xspectra-like independent-particle spectrum, we performed a control calculation where all coefficients of the ionic Hamiltonian were set to zero. This removes any interference between channels that differ by the core-hole quantum numbers, leaving a single-particle-like response. The resulting Ce^{4+} $N_{4,5}$ spectrum is shown in Fig. 4. After applying the same phenomenological broadening used for the Ce^{4+} high- Q curve (FWHM 1.5 eV below 120 eV and 3 eV above), it consists of one strong $4f$ peak (the crystal field being not resolved due to the peak broadening). If instead of switching off all ionic dynamics, we keep the $4d$ spin-orbit term, we obtain the expected ~ 3.3 eV $N_{4,5}$ splitting while still retaining a single-particle-like photoelectron picture.

-
- [1] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.* **140**, A1133 (1965).
 - [2] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
 - [3] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Challenges for density functional theory, *Chem. Rev.* **112**, 289 (2012).
 - [4] F. Aryasetiawan and O. Gunnarsson, The GW method, *Rep. Prog. Phys.* **61**, 237 (1998).
 - [5] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated Fermion systems and the limit of infinite dimensions, *Rev. Mod. Phys.* **68**, 13 (1996).
 - [6] B. T. Thole, G. van der Laan, J. C. Fuggle, G. A. Sawatzky, R. C. Karnatak, and J.-M. Esteve, $3d$ x-ray-absorption lines and the $3d^9 4f^{n+1}$ multiplets of the lanthanides, *Phys. Rev. B* **32**, 5107 (1985).
 - [7] F. de Groot and A. Kotani, *Core Level Spectroscopy of Solids* (CRC Press, Boca Raton, FL, 2008).
 - [8] A. Mirone, M. Sacchi, and S. Gota, Ligand-field atomic-multiplet calculations for arbitrary symmetry, *Phys. Rev. B* **61**, 13540 (2000).
 - [9] M. W. Haverkort, M. Zwierzycki, and O. K. Andersen, Multiplet ligand-field theory using Wannier orbitals, *Phys. Rev. B* **85**, 165113 (2012).
 - [10] A. Longo, R. Wernert, A. Iadecola, C. J. Sahle, L. Stievano, L. Croguennec, D. Carlier, and A. Mirone, An original empirical method for simulating V $L_{2,3}$ edges: The example of KVPO_4F and KVPO_4 cathode materials, *J. Phys. Chem. C* **126**, 19782 (2022).
 - [11] A. A. Mostofi, J. R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, An updated version of wannier90: A tool for obtaining maximally-localised Wannier functions, *Comput. Phys. Commun.* **185**, 2309 (2014).
 - [12] 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 (2011).
 - [13] D. Prendergast and G. Galli, X-ray absorption spectra of water from first principles calculations, *Phys. Rev. Lett.* **96**, 215502 (2006).
 - [14] C. Vorwerk, C. Cocchi, and C. Draxl, Addressing electron-hole correlation in core excitations of solids: An all-electron many-body approach, *Phys. Rev. B* **95**, 155121 (2017).
 - [15] K. Gilmore, J. Vinson, E. Shirley, D. Prendergast, C. Pemmaraju, J. Kas, F. Vila, and J. Rehr, Efficient

- implementation of core-excitation Bethe–Salpeter equation calculations, *Comput. Phys. Commun.* **197**, 109 (2015).
- [16] G. Onida, L. Reining, and A. Rubio, Electronic excitations: Density-functional versus many-body Green’s-function approaches, *Rev. Mod. Phys.* **74**, 601 (2002).
- [17] P. Cudazzo and L. Reining, Correlation satellites in optical and loss spectra, *Phys. Rev. Res.* **2**, 012032(R) (2020).
- [18] J. Authier and P.-F. Loos, Dynamical kernels for optical excitations, *J. Chem. Phys.* **153**, 184105 (2020).
- [19] O. Bunău and M. Calandra, Projector augmented wave calculation of x-ray absorption spectra at the $L_{2,3}$ edges, *Phys. Rev. B* **87**, 205105 (2013).
- [20] M. Taillefumier, D. Cabaret, A.-M. Flank, and F. Mauri, X-ray absorption near-edge structure calculations with the pseudopotentials: Application to the K edge in diamond and α -quartz, *Phys. Rev. B* **66**, 195107 (2002).
- [21] C. Gougousis, M. Calandra, A. P. Seitsonen, and F. Mauri, First-principles calculations of x-ray absorption in a scheme based on ultrasoft pseudopotentials: From α -quartz to high- T_c compounds, *Phys. Rev. B* **80**, 075102 (2009).
- [22] P. Giannozzi Jr *et al.*, Advanced capabilities for materials modelling with QUANTUM ESPRESSO, *J. Phys.: Condens. Matter* **29**, 465901 (2017).
- [23] P. Giannozzi *et al.*, QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [24] P. E. Blöchl, Projector augmented-wave method, *Phys. Rev. B* **50**, 17953 (1994).
- [25] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, *Phys. Rev. B* **59**, 1758 (1999).
- [26] G. C. Baldwin and G. S. Klaiber, Photo-fission in heavy elements, *Phys. Rev.* **71**, 3 (1947).
- [27] J. S. Levinger and H. A. Bethe, Dipole transitions in the nuclear photo-effect, *Phys. Rev.* **78**, 115 (1950).
- [28] A. P. Yutsis, I. B. Levinson, and V. V. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum* (Academy of Sciences of the Lithuanian SSR, Vilnius, 1962).
- [29] R. A. Gordon, G. T. Seidler, T. T. Fister, M. W. Haverkort, G. A. Sawatzky, A. Tanaka, and T. K. Sham, High multipole transitions in NIXS: Valence and hybridization in 4f systems, *Europhys. Lett.* **81**, 26004 (2008).
- [30] A. Longo, A. Mirone, E. De Clermont Gallerande, C. J. Sahle, M. P. Casaleto, L. Amidani, S. A. Theofanidis, and F. Giannici, Oxygen vacancy clusters in bulk cerium oxide and the impact of gold atoms, *Cell Rep. Phys. Sci.* **4**, 101699 (2023).
- [31] S. K. Das, A. Longo, E. Bianchi, C. V. Bordenca, C. J. Sahle, M. P. Casaleto, A. Mirone, and F. Giannici, Deciphering the Ce^{3+} to Ce^{4+} evolution: Insight from x-ray Raman scattering spectroscopy at $Ce N_{4,5}$ edges, *ChemPhysChem* **26**, e202400742 (2025).
- [32] A. Weiße, G. Wellein, A. Alvermann, and H. Fehske, The kernel polynomial method, *Rev. Mod. Phys.* **78**, 275 (2006).
- [33] E. Dagotto, *Correlated Electrons in High-Temperature Superconductors* (Springer, Berlin, 1994).
- [34] The spin of on-site (ionic) electrons is part of $|a_i\rangle$; The spin of the *photoelectron* is carried by the channel index i in the PW factor, as discussed below.
- [35] Xspectrumlet project, GitLab (2025), <https://gitlab.com/xspectrumlet/xspectrumlet>.
- [36] D. Vanderbilt, Soft self-consistent pseudopotentials in a generalized eigenvalue formalism, *Phys. Rev. B* **41**, 7892 (1990).
- [37] A. Pieper, M. Kreutzer, A. Alvermann, M. Galgon, H. Fehske, G. Hager, B. Lang, and G. Wellein, High-performance implementation of Chebyshev filter diagonalization for interior eigenvalue computations, *J. Comput. Phys.* **325**, 226 (2016).
- [38] S. Sen Gupta, J. A. Bradley, M. W. Haverkort, G. T. Seidler, A. Tanaka, and G. A. Sawatzky, Coexistence of bound and virtual-bound states in shallow-core to valence x-ray spectroscopies, *Phys. Rev. B* **84**, 075134 (2011).